Digital Libraries

Usability of OPACs and retrieval engines*

Week 9   KAN Min-Yen

* Mostly based on Hearst's Chapter in
Modern Information Retrieval

What is HCI?

o For digital libraries, it is the means of expediting the information seeking process for a human user

"When an interactive system is well-designed, the interface almost disappears, enabling users to concentrate on their work, exploration and pleasure" --- Shneiderman 97

Tenets of HCI

- Offer informative feedback
- Reduce working memory load
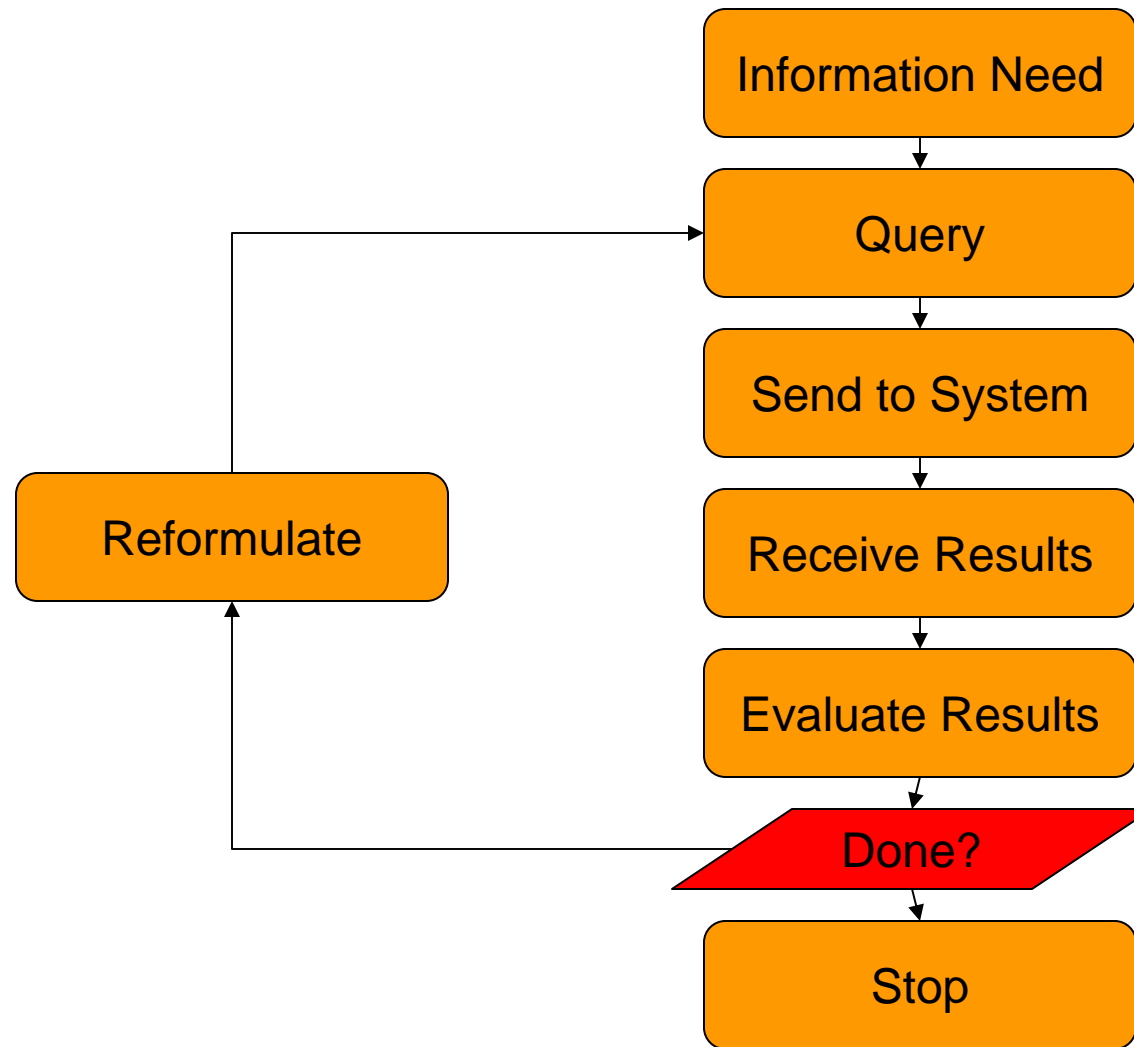- Provide alternative interfaces for novice and expert users

- Methods:
  - Color and Highlighting
  - Pan / Zoom
  - Focus + Context
  - Overview + Details
  - Magic Lens
  - Linking
  - Animation

Evaluating HCI

○ Ergonomics (human factors)

  o time to learn

  o speed of performance

  o rate of errors

  o retention over time

  o subjective satisfaction

ISP Overview

Information Need

↓

Query

↓

Send to System

↓

Receive Results

↓

Evaluate Results

↓

Done?

↓

Stop

Reformulate

Outline

- Searching
  - Query formulation
  - Displaying results

- Browsing
  - Categories and Hierarchies

- Integrating Frameworks

Query specification

- **Old: originated from command line interfaces**
  - Suit the system, not the user

    e.g. "FIND TW Mt St. Helens AND DATE 1981"

- **Then: translated for users on OPACs**
  - Subject: [   ]
  - Date: ⊙ 1980 – 1985   ◯ 1986-1990

- **Now: graphical means for query specification**

VQuery (Jones 98)

Filter / Flow model

○ Users can select from the set of attributes and get an appropriate filter widget

- (type-in for interest areas, sliders for cost, and buttons for scholarships)

○ The widget is placed on the screen with flow lines showing ANDs (sequential flow) and ORs (parallel flows).

○ Water flow dynamically indicates relevant # of items

Against a controlled vocabulary

- Lists
  - Seen on last slide from Filter / Flow
  - Only good if limited number of entries
- Partial Fill-in
  - Show possible completions of query terms if under a certain number (~ 5)
- Re-writing of form fill-in queries
  - Who is the leader of Sudan? →
    Who is the head of state of X (Sudan)?

Tiles (Anick et al. 90)

- Use table to represent conjunction and disjunction
  - Conjunction (AND): rows
  - Disjunction: columns
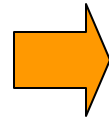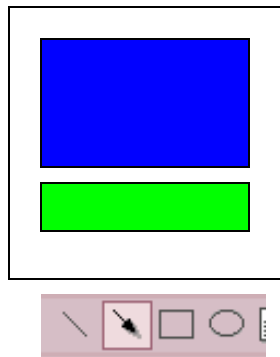- Use activation to see query preview from index

Copy	Backup 313	Saveset 104	From	Tape	Under	V 5.0 344
	Backup Saveset 15			Scratch tape 3		Version 5.0 840
						Version 5

Apply Changes   Display Titles

Query by Example

Query:

• QBIC System
(Flickner et al. 1995)
http://www.hermitagemuseum.org/fcgi-bin/db2www/qbicLayout.mac/qbic?selLang=English

1) **Palace in Athens**

Klenze, Leo von
1835

2) **Allegory of Carfhage**

Stefano,
Francesco di (
Pesellino)
Circa 1448

3) **Allegory of Rome**

Stefano, Francesco
di (Il Pesellino)
Circa 1448

4) **Boats at Saintes-Maries**

Gogh, Vincent
van   1888

5) **A Game of Bowls**

Matisse, Henri
1908

6) **Lake**

Hodler,
Ferdinand
1910-1912s

7) **The Top of Etna from the East. View from San**

8) **View of Lake Maggiore**

Faceted Metadata

Faceted objects give rise to easy methods for summarizing the data

- The Flamenco Project
  http://flamenco.sims.berkeley.edu/cgi-bin/flamenco/arts/Flamenco?username=default

- Facet Map
  http://facetmap.com

Flamenco Image Search

[   ] [Search]

Media

aquatint (2025)
basketry (44)
book (666)
ceramic (1008)
costume (660)
decorative box (163)
domestic object (176)

drawing (2624)
drypoint (2143)
etching (9507)
furnishing (127)
glass (651)
more...

Nature

animal material (515)
birds (1437)
bodies of water (3604)
creatures (801)
fish (219)
flowers (1220)
geological formations (2122)

heavens (2353)
hoofed mammals (2480)
invertebrates and
arthropods (330)
mammals (2116)
plant material (788)
more...

Location

Africa (463)
Asia (1325)
Australia (21)
Central America (134)
Europe (23331)

Middle East (78)
North America (11111)
Oceania (111)
Roman Empire (4)
South America (453)

Places and Spaces

bridges (592)
building parts (3088)
buildings (2393)
dwellings (1709)

lawn (20)
open spaces (1732)
roads (1480)
workplaces (753)

People

aristocrats (974)
children (2501)
men (7372)

occupations (715)
women (5906)

Date

1 - 1000 A.D. (138)
12th century (3)
13th century (1)
14th century (3)
15th century (76)
16th century (1225)
17th century (3058)

18th century (2287)
19th century (7552)
20th - 21th century (18)
20th century (14295)
21st century (12)
more...

Shapes, Colors, and Materials

colors (5861)
decorations (1441)
fabrics (345)
metal (273)

paper (457)
shapes (2752)
visual framing (5911)

Themes

Geographic queries

- Geospatial data makes the 2 & 3-D visualization a good metaphor

Geospatial queries

1. Point-in-polygon
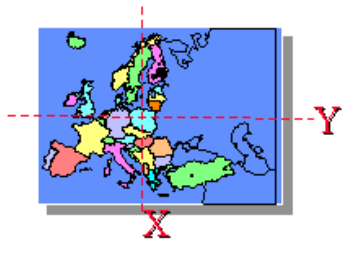
2. Region

3. Distance and Buffer Zone

4. Path

5. Multimedia
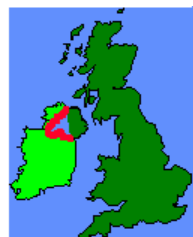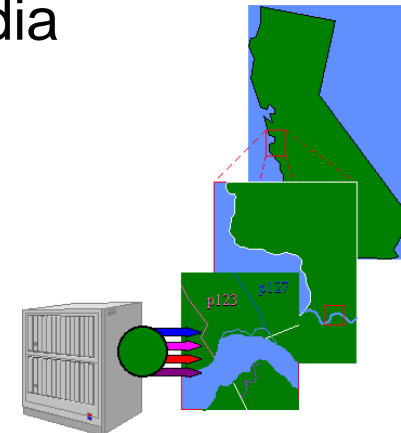
From Larson's typology
Used by permission.

Results display and integration

Keyword in Context (KWIC)

Differentiate
Keyword in Context
versus
Query Word in Context

- QWIC: query based

- KWIC: any keyword
(can be from metadata)

Tilebars (Hearst 95)

Topic_1: **"drought"**
Topic_2: **"ration"**
was submitted to the UC Berkeley Digital Library Database where **187** matching records records were foun
shown in the tilebars display below.

State Drought Water Bank
report

California Water Plan Update, Volume I
bulletin

California Water Plan Update Volume 2
bulletin

California's Continuing Drought 1987-91
report

Dublin San Ramon Services District Urban Water Manager
plan

New Updated Edition Urban Drought Guidebook
report

Regulatory Impact Assessment of The Final Water Quality
report

Urban Water Management Plan
plan

Irvine Ranch Water District 1995 Urban Water Managemen
plan

City of Santa Cruz Urban Water Management Plan
plan

California Water Plan Update Executive Summary
bulletin

Bay-Delta Oversight Council Draft: Briefing Paper on Stat

- Each row represents a topic (conjunction)

- The darkness of each tile in a row represent the frequency of occurrence of an item

Infocrystal (Spoerri 95)

○ Explosion of the Venn Diagram
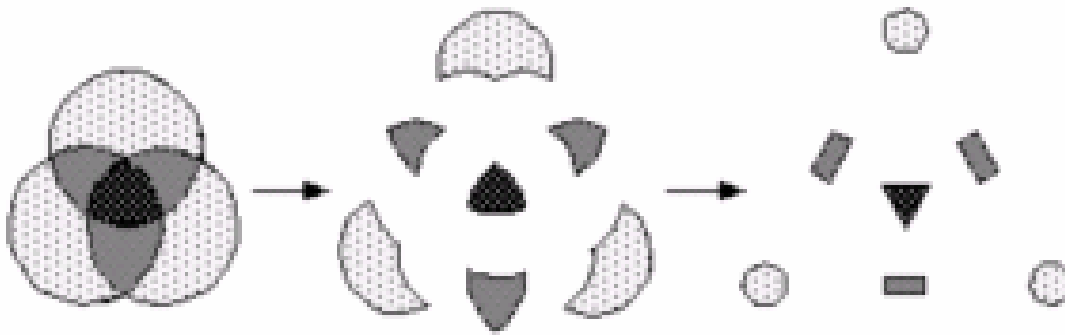
○ Uses shape and color to
  model to organize
  results
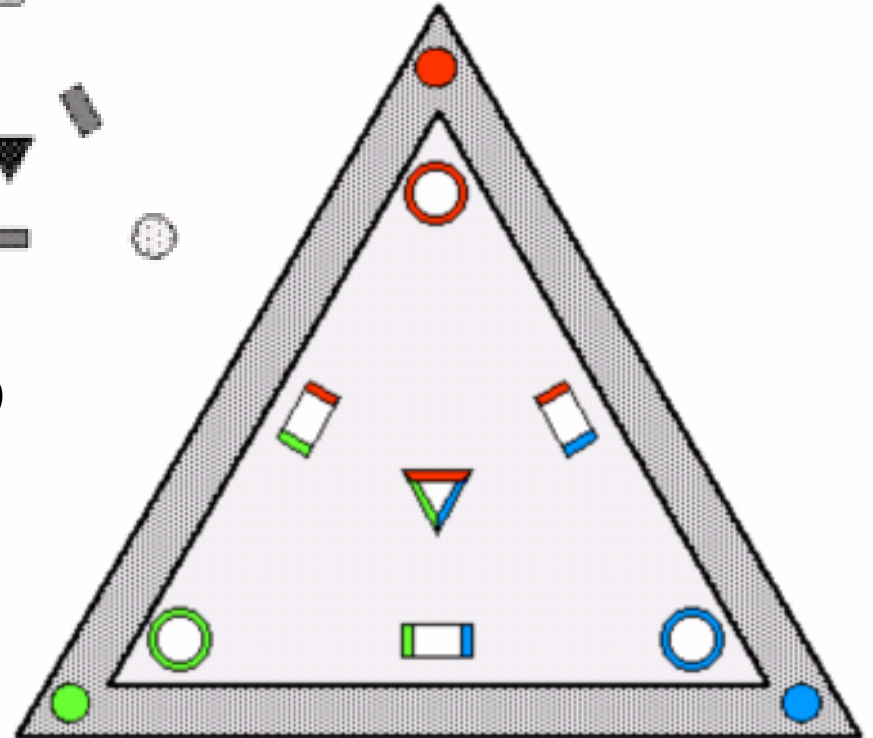
More on Infocrystal

- Queries can build on others to form nested queries
- Query terms can be weighted between 1 (default) and -1 (logical not)
- The icons are placed closer to the center of the crystal as they are more relevant

5 Oct 2
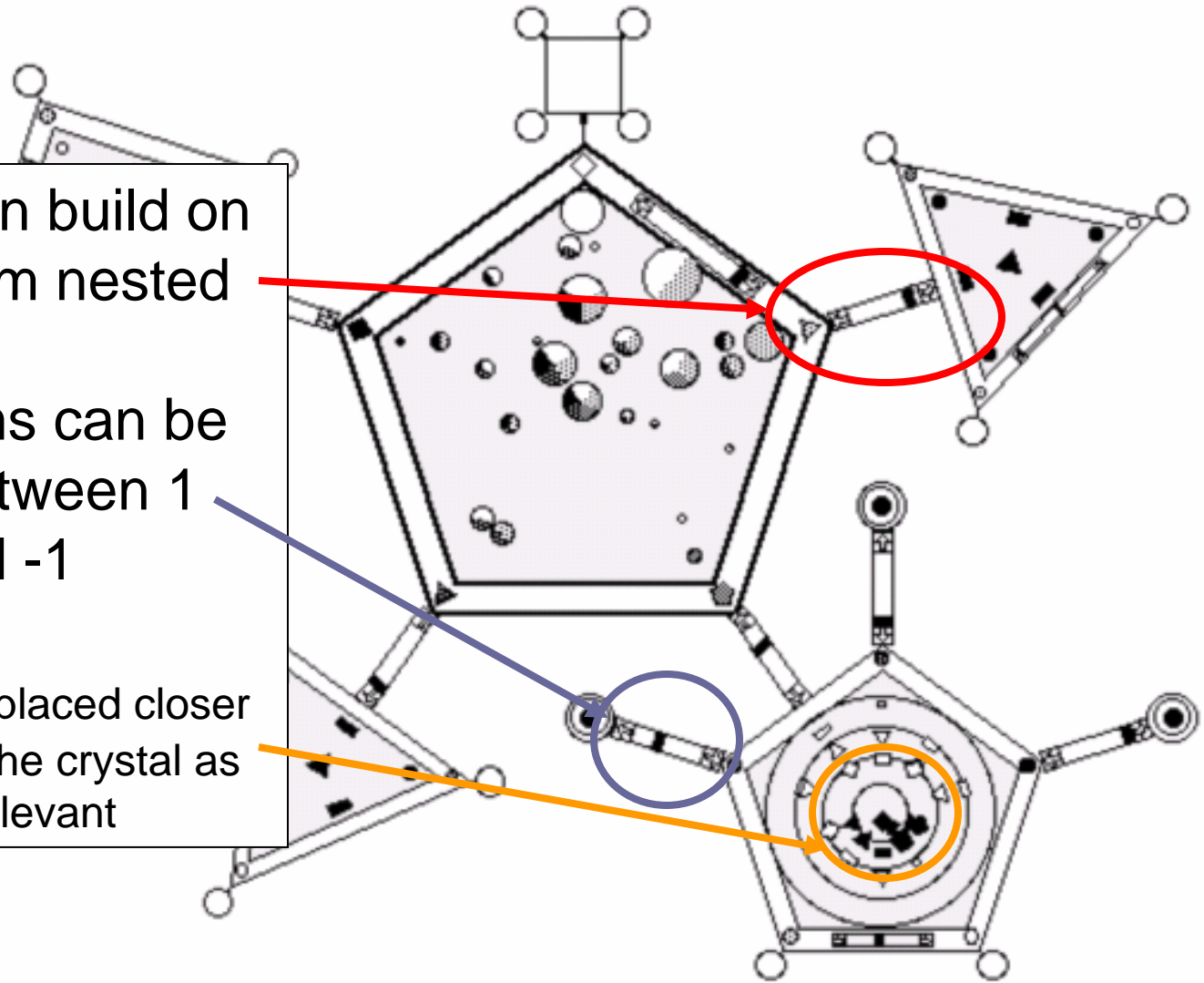
Superbook

- Searching a software technical manual
- Use a expandable table of contents to show results
- Shows hits within each level as query preview
- Closely modeled by browser in Windows Explorer

- **12** An Interactive Environment
  - **10** Preface
    - **1** How to Beat the Lottery
    - **1** Tutorial Introduction to
    - **7** Graphical Methods in S
      - **2** Loading in Data
      - ...
      - **3** Building Plots
      - **2** Specialized Plots
    - **2** Advanced Use of S
  - **2** Appendix I

Cha-cha

- Brings this paradigm to website searching

Tablelens (InXight)

○ http://www.tablelens.com

○ Focus + context to give fisheye distortion to table rows and values
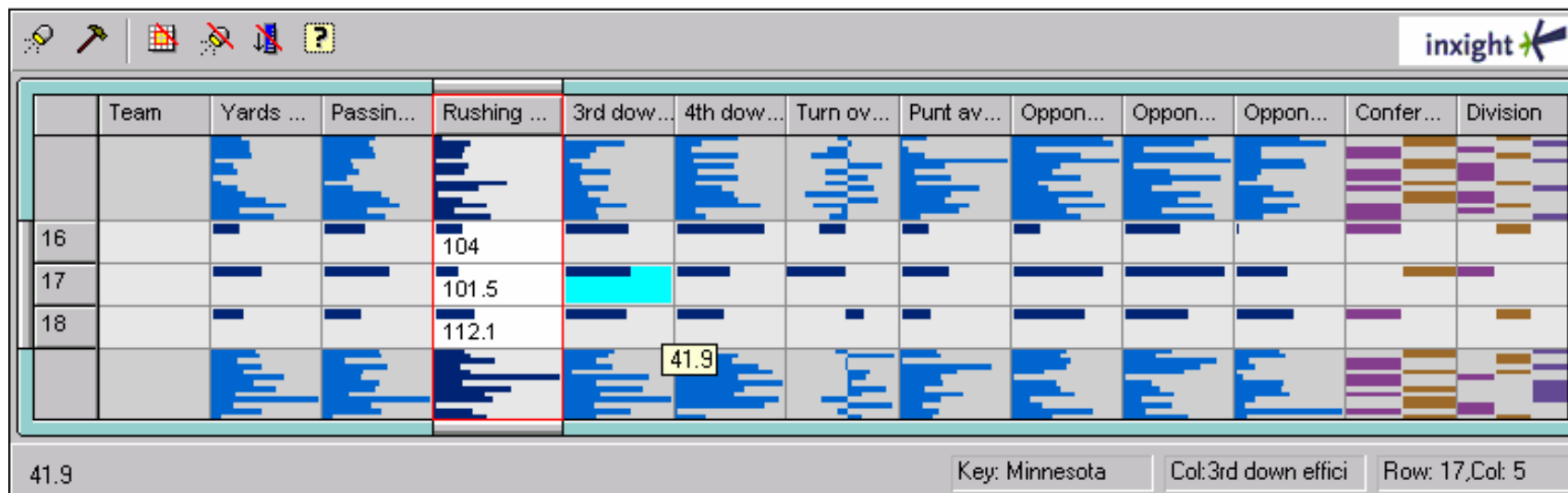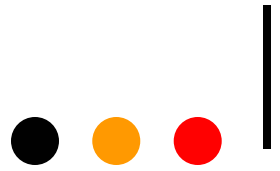
Cartographic Representations

CS 5244: HCI in Digital Libraries
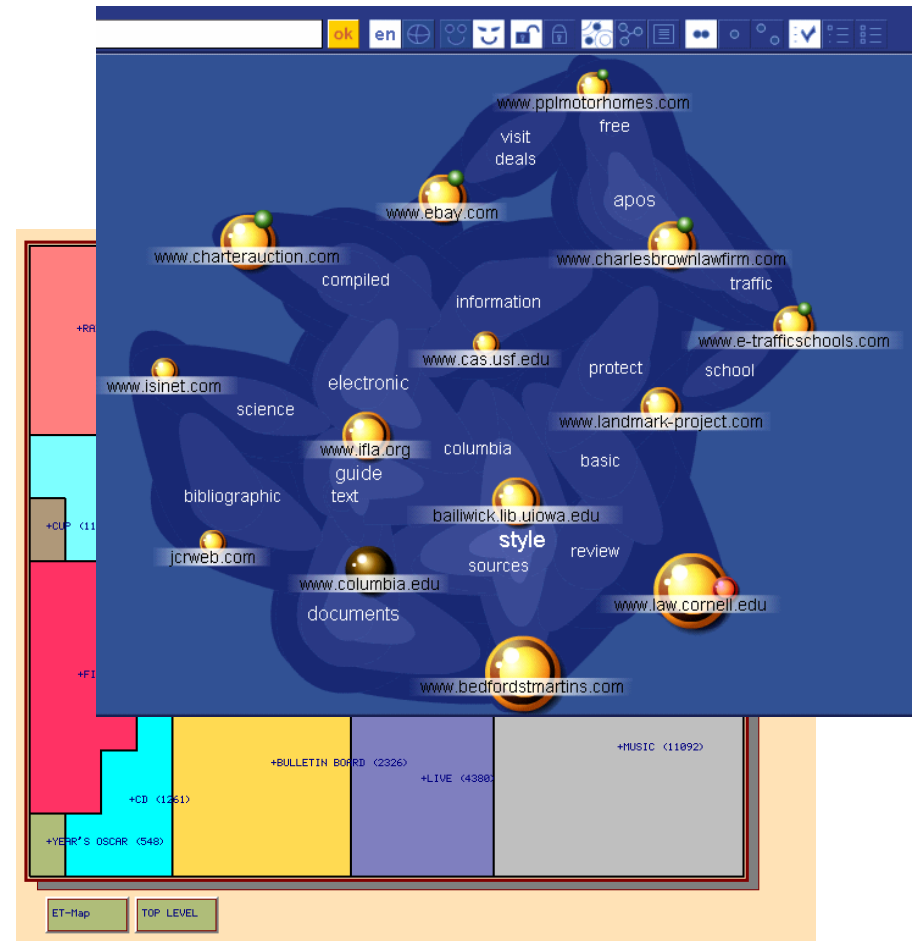
Scatter / Gather

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

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

Gradual refinement by selection of clusters:

Gather: select cluster(s)
        (relevance feedback)
Scatter: throw them into
        bins using automatic
        clustering
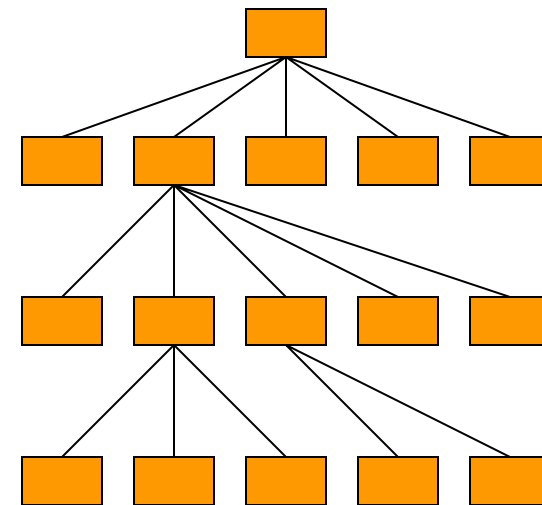
Scatter / Gather

Second cluster: stars (of Broadway)   Third cluster: stars (as heavenly objects)

○ 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

○ 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 lea

○ Williams, Hank
○ Crosby, Bing
○ Campbell, Glen
○ Belafonte, Harry
○ Shore, Dinah
○ Denver, John
○ Hampton, Lionel

☐ Cluster 1 Size: 12   black white nuclear hole reaction helium neutron gra

○ stellar evolution
○ gravitational collapse
○ black hole
○ main sequence
○ carbon cycle
○ mass−luminosity relation

☐ Cluster 2 Size: 49   galaxy type distance stellar variable spectral interstell

○ star
○ extragalactic systems
○ Galaxy, The
○ interstellar matter
○ cluster, star
○ population, stellar

☐ Cluster 3 Size: 29   constellation northern hemisphere sky locate dipper

○ constellation (astronomy)
○ Auriga
○ Big Dipper
○ Cassiopeia
○ Cygnus
○ Taurus

☐ Cluster 4 Size: 7   fraunhofer designate map joseph frown fur wollaston e

○ Fraunhofer lines
○ Fraunhofer, Joseph von

Visualizing Hierarchies

○ Fisheye zoom
(Furnas 84)

○ One implementation:
Inxight's StarTree

● http://www.inxight.com/
VizServerDemos/demo/
nasa/index.html
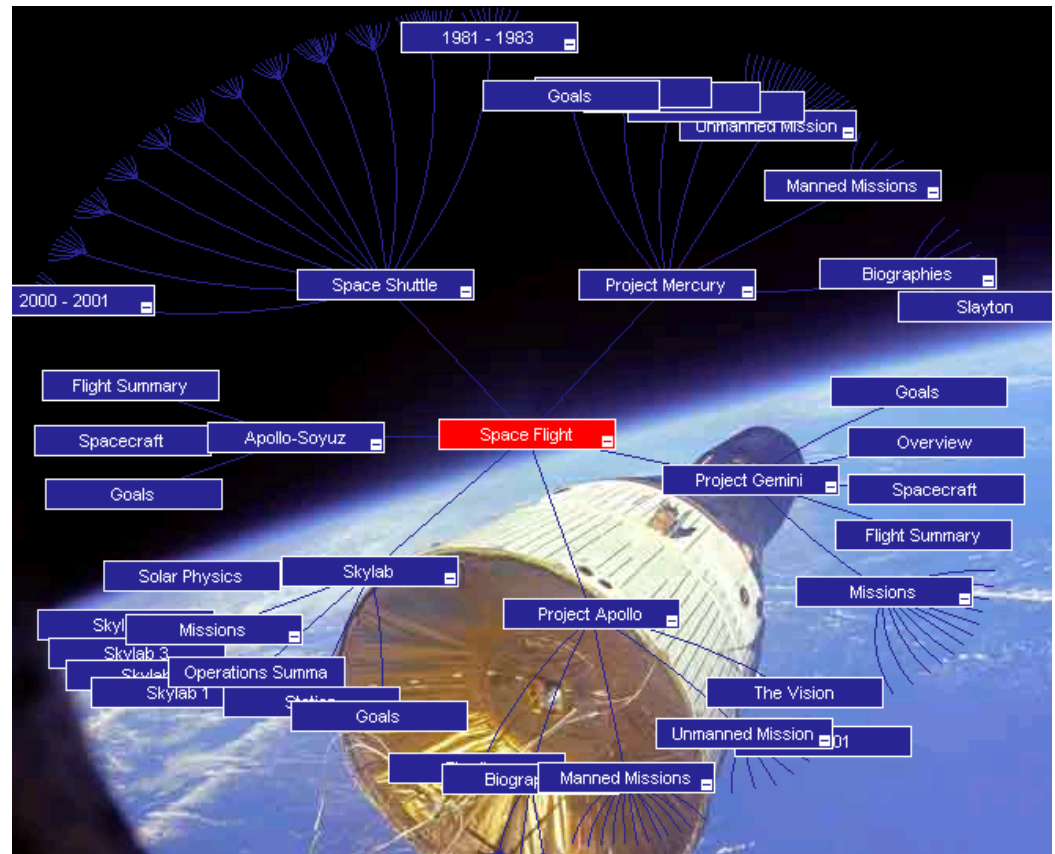
Summarization atop IR

- Like Superbook, but trades focus for extractive summary (Kan et al. 02)
  - **Relevant**: information for summary
  - **Irrelevant**: too broad or not on topic, for broader queries
  - **Intricate**: too detailed, for follow-ups

Query: Hypertension
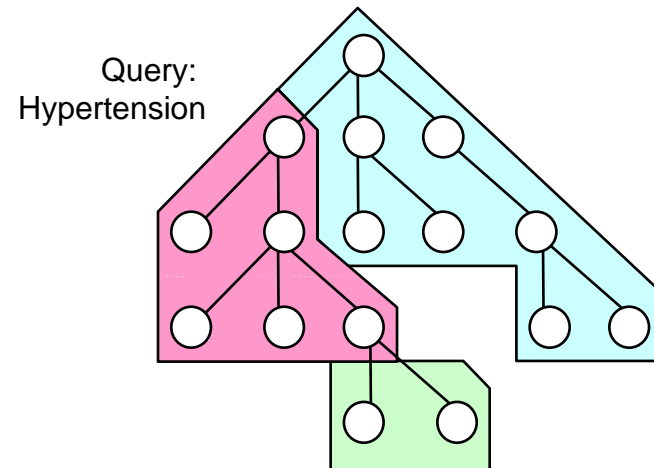
Merging Document Topic Trees

○ Norm for a particular type of document

  (e.g. travel leaflets, Univ. department descriptions)

  ● Create by aligning topics in trees by similarity

joined node
at level 1
(e.g. disease)

doc tree
1 (yellow)

doc tree 2
(blue)

newly
joined
node at
level 2
(e.g.
treatment)

doc tree
3 (purple)

joining nodes at level 2

diseas
e node

treatment node

newly joined
node at level
3 (e.g.
surgery)

joining nodes at level 3
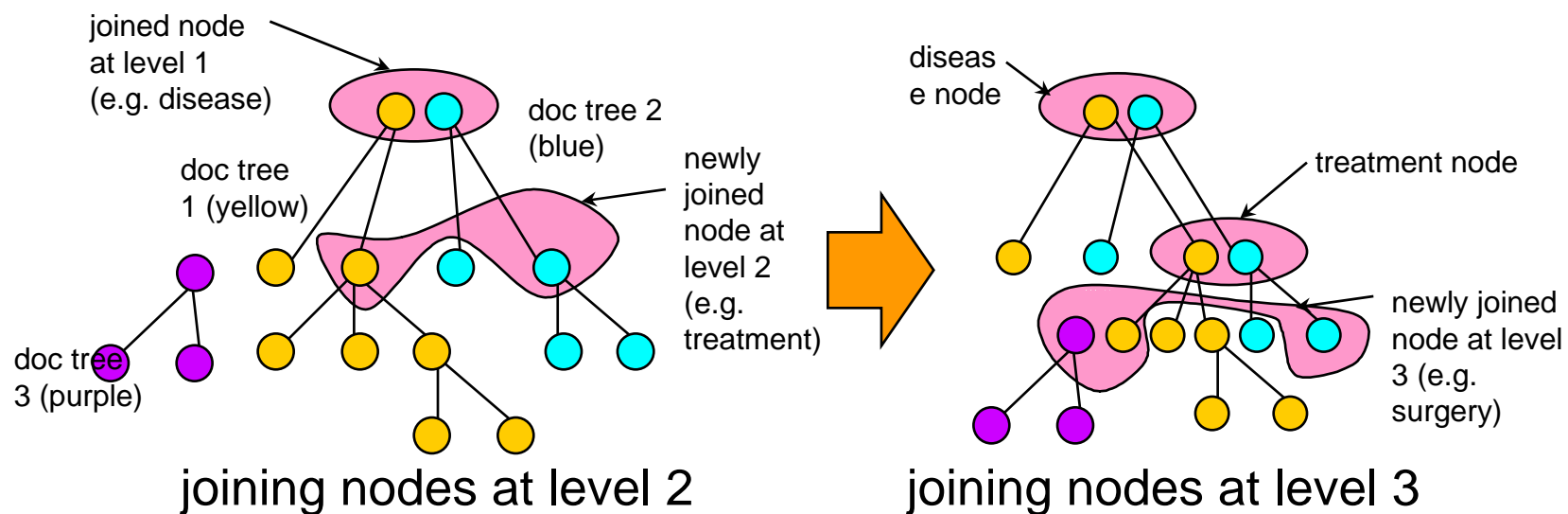
Flamenco Image Search

Refine your search further within these categories:

Media (group results)
costume (4), domestic object (1), drawing (136), lithograph (1), metalwork (1), print (1)

Location: all > Asia > India (group results)
Assam, Nagaland (1), Company School (6), Deccan (Maharatha) School (1), Deccan School? (1), Deccan/Hyderabad (1), Gujarat (3), Jain School (1), Kashmir School (15), Kashmir School (Afghan?) (1), Kashmir School? (1), more...

Date (group results)
1 - 1000 A.D. (1), 13th century (1), 15th century (1), 16th century (2), 17th century (12), 18th century (49), 19th century (53), 20th century (3), date ranges spanning multiple centuries (9), date unknown (17)

Themes (group results)
military (30), mortality (2), music, writing, and sport (47), nautical (4), religion (42)

Objects
clothing (15), containers (4), food (6), fuel (1), furnishings (11), jewelry and riches (3), lighting (1), timepieces (1), vehicles (3), wares (1)

Nature (group results)
animal material (6), birds (18), bodies of water (14), creatures (6), fish (2), flowers (9), geological formations (17), heavens (3),

These terms define your current search. Click the ☒ to remove a term.

Location: Asia > India ☒

start a new search

[Search] ⦿ all items ○ within current results

148 items (grouped by objects)

view ungrouped items

clothing 15

A Lady at a Shrine
Anonymous
19th century

A Man Greets a P...
Anonymous
circa 1775

A Young Hill Raj...
Eden
circa 1836 - 1841

Divan Garah Singh
Anonymous
circa 1750 - 1780

Durga
Anonymous
circa 1800

Durga Killing a ...
Anonymous
circa 1770 - 1775

Ganesh and an At...
Anonymous
late 18th century

Hat
no artist
early 20th century

Additional notes on faceted interface

○ **Multiple interfaces possible**

- Columnar style
- Drop down list style
- Indented ancestors

You've selected these headings:

Red Wines (171)   French (55)   Bargains under $20 (237)
White Wines (149)   German (6)   Top shelf (over $100) (11)
Bubbly (40)   Italian (67)
                                        Set your own Price:
                      ...land (2)
                      ...ropean (8)   from [   ] to [   ]
                      ...se (19)
                      ...merican   [Set]

You've selected these headings:

> Any Varietal   Limit Varietal to: [▼]

> The World   Limit Region to:

> $0 through $1000   Limit Price to:

                      or select value: [

You've selected these headings:
> The World > **Portuguese**

Browse Varietal   **Browse Price**
Dessert/Fortified Wines   Bargains under $20 (3), Top shelf (over $100) (2)
(19)
                             Set your own Price: from [   ] to [   ]
                             [Set]

You've narrowed your search down to 19 results. Here are the top 10.

Delaforce 1977 Vintage Port
Delaforce 1994 Vintage Port
Delaforce 1995 Quinta Da Corte
Croft 1983 Quinta Da Roeda
Cockburn Fine Ruby
Cockburn Fine Tawny
Cockburn 10 year-old Tawny
Cockburn 20 year old Tawny
Cockburn NV Special Reserve
Cockburn 1995 Anno Late Bottled Port

Features
- Gives partial results
- Gives multiple filters at once
- A selection drops non-matching items from results
- Allows users to define most useful criteria
- In UI, only show fields that address audience

Multiviews via Magic Lenses

○ More than two views of the same space

○ Lenses show and hide information that isn't pertinent

○ Has been applied in query focusing as well

○ Demo:
http://www.memorialhall.mass.edu/activities/media.jsp?itemid=15414&img=0
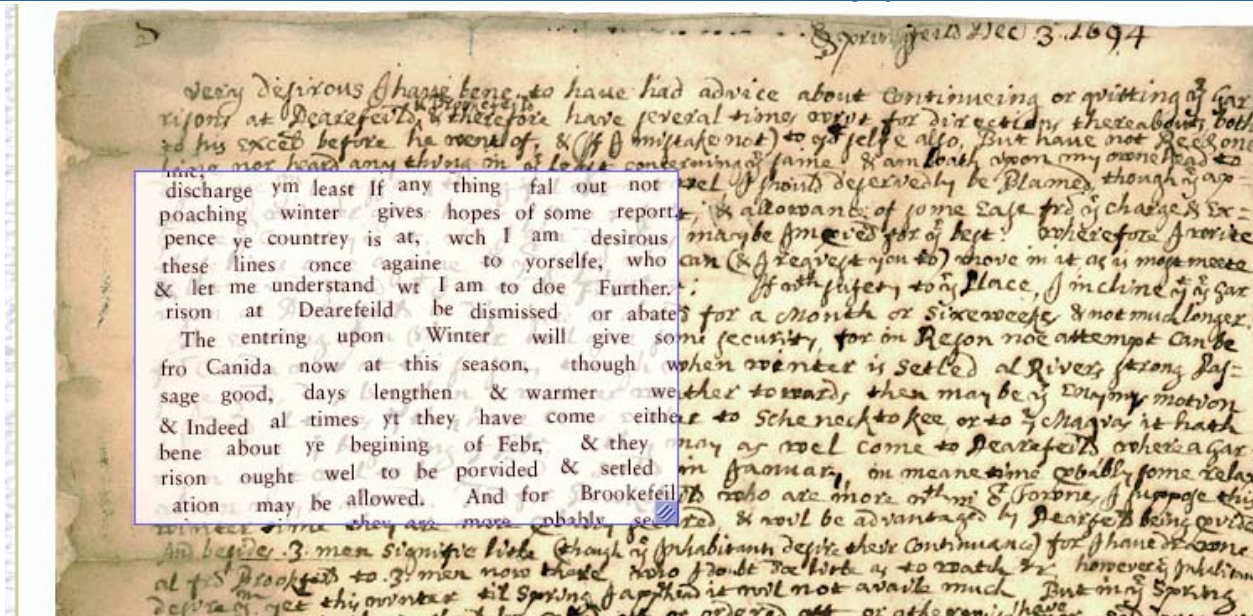
Bibliometric citations

An overview of the author co-citation map (1977-2001), consisting of 380 authors in the ARIST data set with 9 or more citations. The map is dominated by the largest specialty of citation indexing. No strong concentration of other specialties are found, which implies the diversity of the domain.
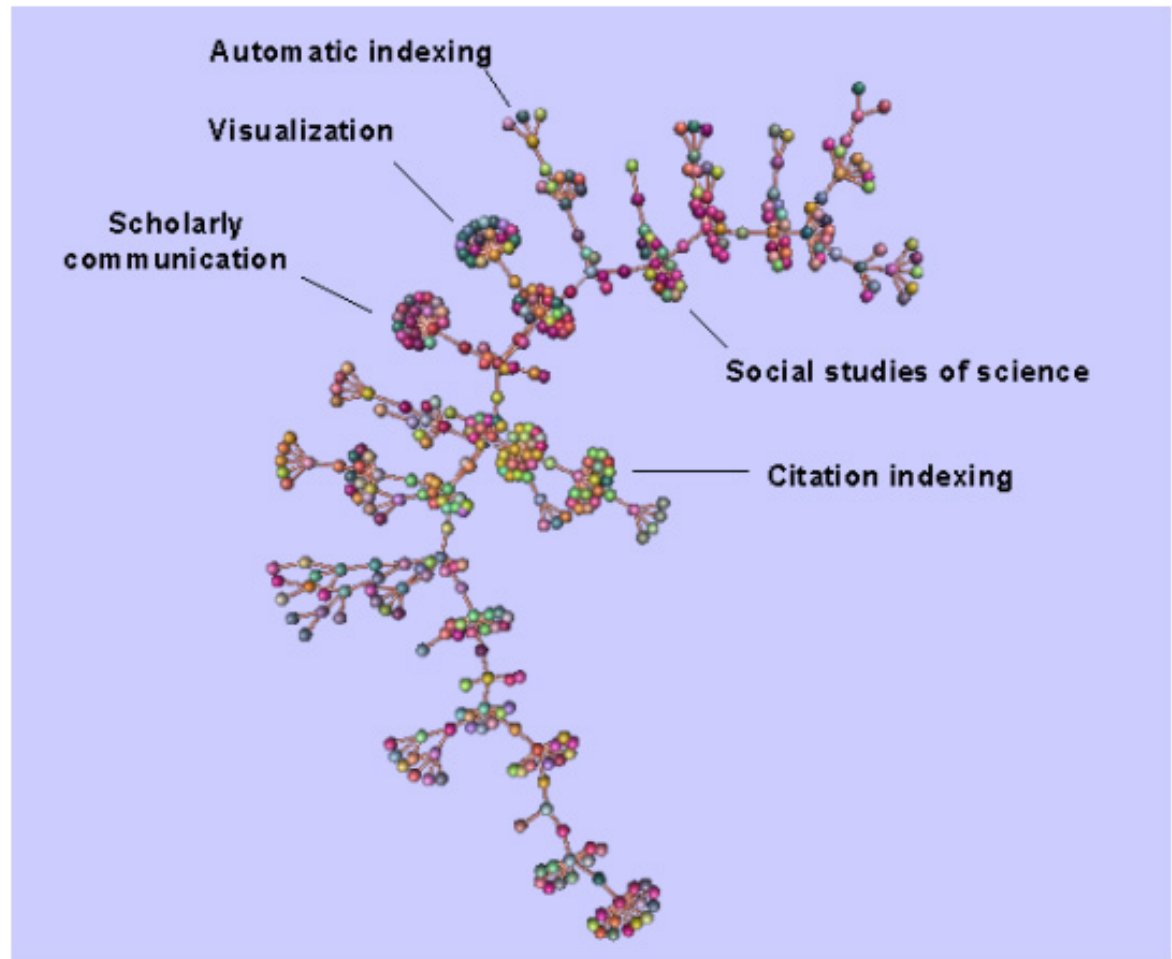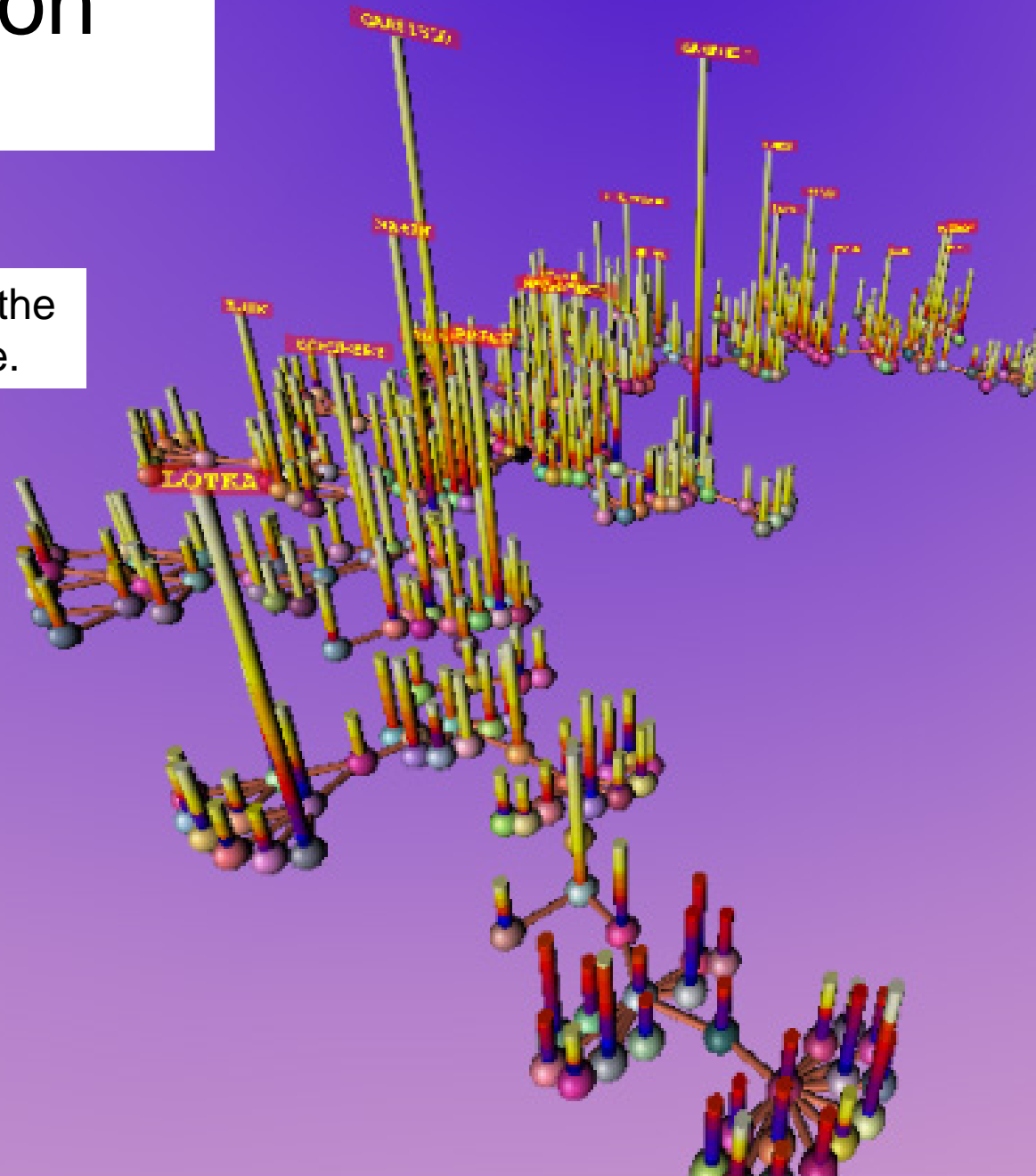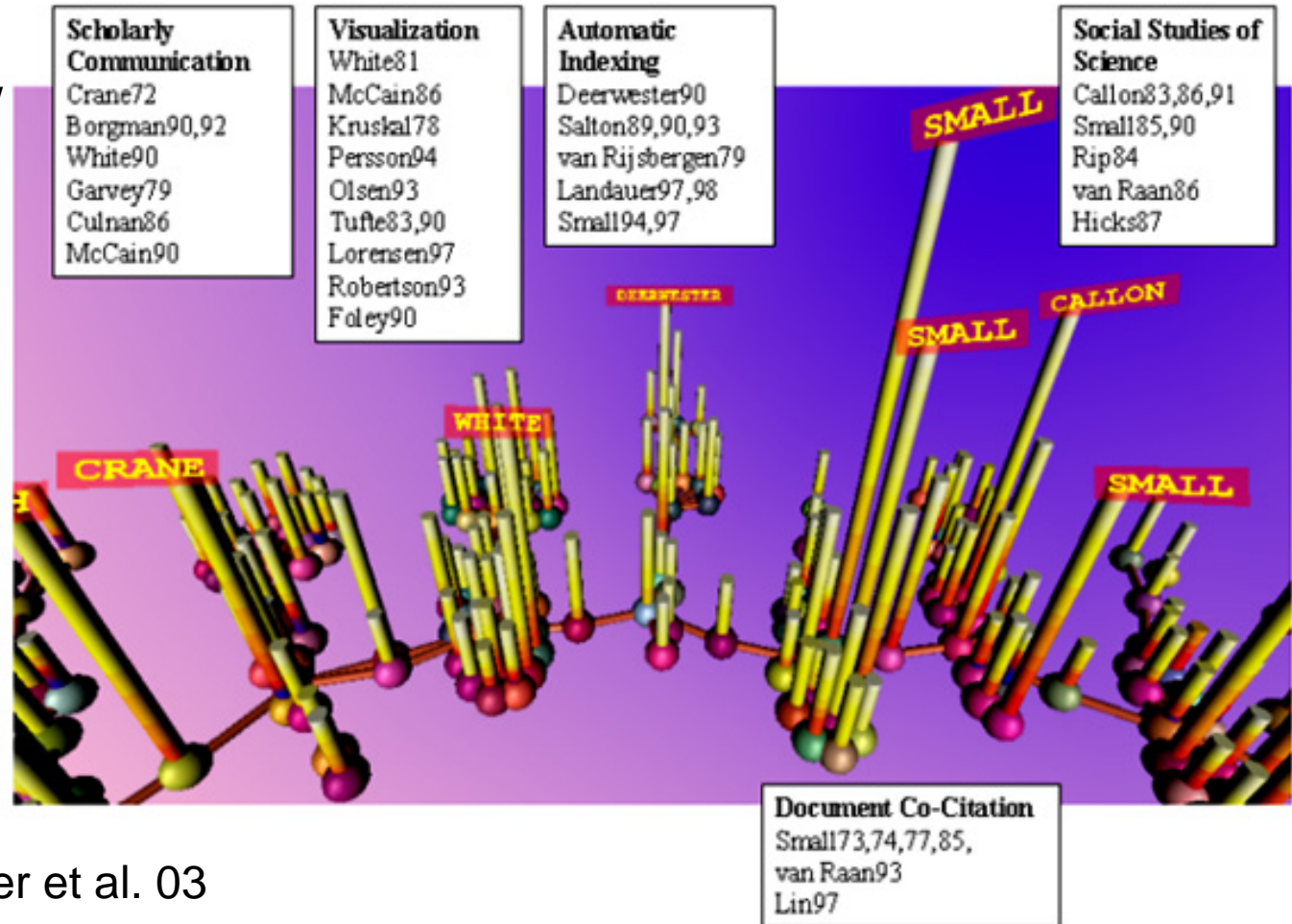
-- Börner et al. 03

With citation rates

A landscape view of the DCA map at distance.

-- Börner et al. 03

Zoom in on the artery

A close-up view of a few clusters along the main artery in the DCA map. The height of a bar represents the number of citations to a publication. Labels indicate articles in clusters, for example, Small73 for an article of Small in 1973. Multiple publications within the same year are not distinguished at this level. For example, Small73 includes all Small's publications in 1973.

-- Börner et al. 03

Scholarly Communication
Crane72
Borgman90,92
White90
Garvey79
Culnan86
McCain90

Visualization
White81
McCain86
Kruskal78
Persson94
Olsen93
Tufte83,90
Lorensen97
Robertson93
Foley90

Automatic Indexing
Deerwester90
Salton89,90,93
van Rijsbergen79
Landauer97,98
Small94,97

Social Studies of Science
Callon83,86,91
Small85,90
Rip84
van Raan86
Hicks87

Document Co-Citation
Small73,74,77,85,
van Raan93
Lin97

Relevance feedback

- Koenenmann and Belkin (96) checked whether relevance feedback helped users

- Control: no RF
- Opaque: saw new documents
- Transparent: can see modifications
- Penetrable: can see modifications and change them
  - Best one

- Possible conclusion: RF can help users create better queries as a form of online training

Query History

- Relevance feedback as one form

- Pros:
  - Saved searches can help tune a generic system
  - Decide whether to use advance or naïve interface for person
  - Start off where one left off

- Cons:
  - Need to track the user
  - What happens if multiple purposes or used for guests?
    - (e.g. Buying gifts for a friend)

Conclusions

○ New systems show advanced text and graphics / animated displays but many are not yet well-evaluated

○ Tighter integration in query and display

○ Simultaneous browsing and searching support at all levels

○ Screen real estate proportions in frameworks dictate useful alternatives

To think about

- What types of information do these various UI need to calculate at run-time and what types can be pre-computed?

- How do these UI support the tenets we mentioned at the beginning?

- Do you feel that HTML / WWW has enhanced UI design or deterred its creativity?

- For images and audio:
  - When is query by example a useful technique and when is it inferior to a search using metadata?