Usability of OPACs and retrieval engines*

Week 9

KAN Min-Yen

* Mostly based on Hearst’s Chapter in *Modern Information Retrieval*
What is HCI?

- 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)
  - time to learn
  - speed of performance
  - rate of errors
  - retention over time
  - subjective satisfaction
ISP Overview

Information Need → Query → Send to System → Receive Results → Evaluate Results → Reformulate → Done? → Stop
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: 

- Now: graphical means for query specification
VQuery (Jones 98)

- lion: 25
- tiger: 16
- jungle: 10
- Tarzan: 2
- elephant: 12
- Jane: 12
- cheetah: 1

Collections:
- Comp Sci Tech. Reports
- MedDoc Collection
- Computists’ Communiqué
- FAQ Archive
- HCI Bibliography
- Indigenous Peoples
- Oxford Text Archive
- Project Gutenberg Collection
- TidBITS

Search for any documents in "Oxford Text Archive" containing either lion and tiger; or jungle; but not Tarzan

Submit Query to NZDL
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

<table>
<thead>
<tr>
<th>Copy</th>
<th>Backup</th>
<th>Saveset</th>
<th>From</th>
<th>Tape</th>
<th>Under</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>313</td>
<td>104</td>
<td></td>
<td></td>
<td>V 5.0 344</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Scratch tape 3</td>
<td></td>
</tr>
<tr>
<td>Backup Saveset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Version 5.0 840</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td>Version 5</td>
</tr>
</tbody>
</table>

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
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

<table>
<thead>
<tr>
<th>Media</th>
<th>Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>aquatint (2025)</td>
<td>animal material (515)</td>
</tr>
<tr>
<td>basketry (44)</td>
<td>heavens (2363)</td>
</tr>
<tr>
<td>book (666)</td>
<td>hoofed mammals (2480)</td>
</tr>
<tr>
<td>ceramic (1008)</td>
<td>invertebrates and arthropods (330)</td>
</tr>
<tr>
<td>costume (660)</td>
<td>mammals (2110)</td>
</tr>
<tr>
<td>decorative box (163)</td>
<td>plant material (783)</td>
</tr>
<tr>
<td>domestic object (176)</td>
<td>more...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location</th>
<th>Places and Spaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa (463)</td>
<td>bridges (592)</td>
</tr>
<tr>
<td>Asia (1325)</td>
<td>lawn (20)</td>
</tr>
<tr>
<td>Australia (21)</td>
<td>open spaces (1732)</td>
</tr>
<tr>
<td>Central America (134)</td>
<td>roads (1480)</td>
</tr>
<tr>
<td>Europe (23331)</td>
<td>workplaces (753)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>People</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 1000 A.D. (138)</td>
<td>aristocrats (974)</td>
</tr>
<tr>
<td>12th century (3)</td>
<td>children (2501)</td>
</tr>
<tr>
<td>13th century (1)</td>
<td>men (7372)</td>
</tr>
<tr>
<td>14th century (3)</td>
<td>occupations (715)</td>
</tr>
<tr>
<td>15th century (76)</td>
<td>women (5906)</td>
</tr>
<tr>
<td>16th century (1226)</td>
<td>more...</td>
</tr>
<tr>
<td>17th century (3068)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Themes</th>
<th>Shapes, Colors, and Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>drawing (2624)</td>
<td>colors (5861)</td>
</tr>
<tr>
<td>drypoint (2143)</td>
<td>decorations (1441)</td>
</tr>
<tr>
<td>etching (9607)</td>
<td>fabrics (346)</td>
</tr>
<tr>
<td>furnishing (127)</td>
<td>metal (273)</td>
</tr>
<tr>
<td>glass (651)</td>
<td>paper (467)</td>
</tr>
<tr>
<td>more...</td>
<td>shapes (2762)</td>
</tr>
<tr>
<td></td>
<td>visual framing (5911)</td>
</tr>
</tbody>
</table>
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)

- Each row represents a topic (conjunction)
- The darkness of each tile in a row represents 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
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
Brings this paradigm to website searching
Tablelens (InXight)

- [http://www.tablelens.com](http://www.tablelens.com)
- Focus + context to give fisheye distortion to table rows and values
Cartographic Representations
Scatter / Gather

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)**

- Musial, Stan
- Bench, Johnny
- Carew, Rod
- Robertson, Oscar
- Bovieux, Jean
- Casper, Billy
- Chinese checkers
- Best, George
- Beamon, Bob

**Third cluster: stars (as heavenly objects)**

- Burstyn, Ellen
- Stanwyck, Barbara
- Berle, Milton
- Bankhead, Tallulah
- Murphy, Eddie
- Walsh, Raoul
- Martin, Mary
- Zukor, Adolph
- Crosby, Bill

---

**Cluster 1 Size: 12**
- black
- white
- nuclear
- hole
- reaction
- helium
- neutron
- gravitational
- collapse
- black
- hole
- main
- sequence
- carbon
- cycle
- mass–luminosity
- relation

**Cluster 2 Size: 47**
- role
- stage
- broadway
- comedy
- performance
- stellar
- evolution
- 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

- Fraunhofer lines
- Fraunhofer, Joseph von
Visualizing Hierarchies

- Fisheye zoom
  (Furnas 84)
- One implementation: Inxight’s StarTree

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

joining nodes at level 2

joining nodes at level 3
Multiple Hierarchies: Faceted Queries

Refine your search further within these categories:

- **Media** (group results)
  - costume (4), domestic object (7),
  - drawing (136), lithograph (1), metalwork (1),
  - print (1)

- **Location** (group results)
  - all > Asia > India (group results)
    - Assam, Nagaland (1), Company School (6),
    - Deccan (Maharashtra) School (1), Deccan
      School (1), Deccan/Hyderabad (1),
    - Gujarar (3), Jain School (1), Kashmir
      School (16), Kashmir School (Afghan?) (1),
    - Kashmir School (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 (41), jewelry and
    riches (3), lighting (1), furniture/interiors (1),
    vehicles (3), wares (1)

- **Nature** (group results)
  - animal material (6), birds (18), bodies of
    water (14), creatures (6), fish (2), flowers (3),
    geological formations (11), heavens (3)

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

- Location: Asia > India

148 items (grouped by objects)
Additional notes on faceted interface

- Multiple interfaces possible
  - Columnar style
  - Drop down list style
  - Indented ancestors

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
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?