Information Search

Information Search

- **Introduction**
- **Searching in Textual Documents and Database Querying**
- **Multimedia Document Searches**
- **Advanced Filtering and Search Interfaces**

Information Search (cont.)

- **Information search should be a joyous experience, but it takes a skilled searcher with robust tools to perform an effective search**

- **Information retrieval and database management have evolved into:**

  - information gathering

  - Seeking

  - filtering

  - collaborative filtering

  - sensemaking

  - visual analytics

- **Now the concentration of CS scientists are on**

  - data mining from data warehouses and data marts

  - knowledge networks or semantic webs

- **All the above is complicated by the increased volume of material to search**

Search terminology

- ***Task objects*** (such as movies for rent) are stored in structured relational databases, textual document libraries, or multimedia document libraries

- A ***structured relational database*** consists of *relations* and a *schema* to describe the relations

- ***Relations*** have **items** (usually called tuples or records), and each item has multiple attributes (often called fields), which each have attribute values

- ***A textual document library*** consists of a set of collections (typically up to a few hundred collections per library) plus some descriptive attributes or metadata about the library (for example, name, location, owner)

Search terminology (cont.)

- *Task actions* are decomposed into *browsing* or *searching*
- Here are some examples of task actions:
    - *Specific fact finding (known-item search)*
        - Find the e-mail address of the President of the United States.
    - *Extended fact finding*
        - What other books are by the author of "Jurassic Park"?
    - *Exploration of availability*
        - Is there new work on voice recognition in the ACM digital library?
    - *Open-ended browsing and problem analysis*
        - Is there promising new research on breast cancer that might help my patient?

Search terminology (cont.)

- Once users have clarified their information needs, the first step towards satisfying those needs **is *deciding where to search***

- Supplemental ***finding aids*** can help users to clarify and pursue their information needs, e.g. table of contents or indexes

- Additional ***preview and overview surrogates*** for items and collections can be created to facilitate browsing

Searching in Textual Documents and Database Querying

- **Expert users can use SQL:**

  SELECT DOCUMENT#

  FROM JOURNAL-DB

  WHERE (DATE >= 2004 AND DATE <= 2008)

  AND (LANGUAGE = ENGLISH OR FRENCH)

  AND (PUBLISHER = ASIST OR HFES OR ACM)
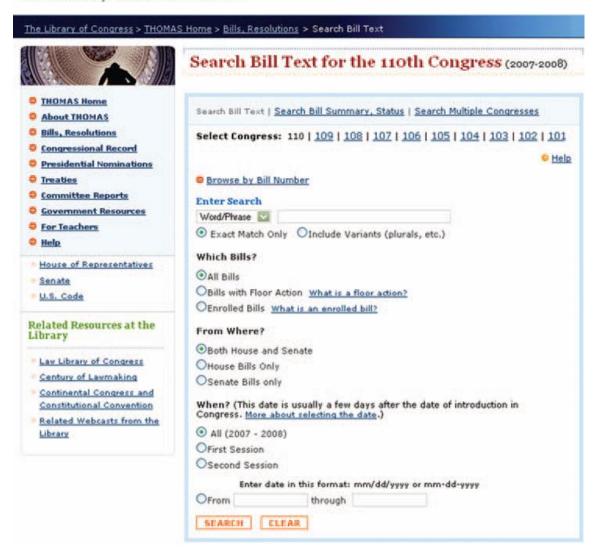
- **SQL has powerful features, but it requires training**

  - While SQL is a standard, form fill-in queries have simplified query formulation

Searching in Textual Documents and Database Querying (cont.)

- **Other methods include:**
  - Natural language queries
  - Form fill-in
  - Query by example (QBE)
- **Evidence shows that users perform better and have higher satisfaction when they can view and control the search**

Searching in Textual Documents and Database Querying (cont.)

Five-phase framework to clarify user interfaces for textual search

1. **Formulation:** expressing the search
2. **Initiation of action:** launching the search
3. **Review of results:** reading messages and outcomes
4. **Refinement:** formulating the next step
5. **Use:** compiling or disseminating results

Five-phase framework to clarify user interfaces for textual search (cont.)

1. **Formulation**
   - Provide access to the appropriate sources in libraries and collections.
   - Use *fields* for limiting the source: structured fields such as year, media, or language; and text fields such as titles or abstracts of documents.
   - Recognize *phrases* to allow entry of names, such as George Washington or Environmental Protection Agency, and concepts, such as abortion rights reform or gallium arsenide.
   - Permit *variants* to allow relaxation of search constraints, such as case sensitivity, stemming, partial matches, phonetic variations, abbreviations, or synonyms from a thesaurus.
   - Control the size of the result set.

2. **Initiation of action**
   - Include *explicit actions* initiated by buttons with consistent labels (such as "Search"), locations, sizes, and colors.
   - Include *implicit actions* initiated by changes to a parameter of the formulation phase that immediately produce new sets of search results.

3. **Review of results**
   - Present explanatory messages.
   - View an overview of the results and previews of items.
   - Manipulate visualizations.
   - Adjust the size of the result set and which fields are displayed.
   - Change the sequencing (alphabetical, chronological, relevance ranked, and so on).
   - Explore clustering (by attribute value, topics, and so on).
   - Examine selected items.

4. **Refinement**
   - Use meaningful messages to guide users in progressive refinement; for example, if the two words in a phrase are not found near each other, offer easy selection of individual words or variants.
   - Make changing of search parameters convenient.
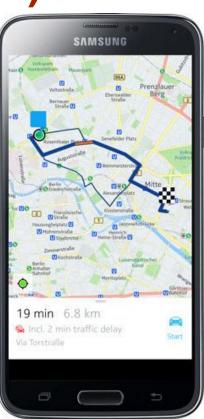   - Explore relevance feedback.

5. **Use**
   - Allow queries, parameter settings, and results to be saved and annotated, sent by e-mail, or used as input to other programs (such as visualization or statistical tools).

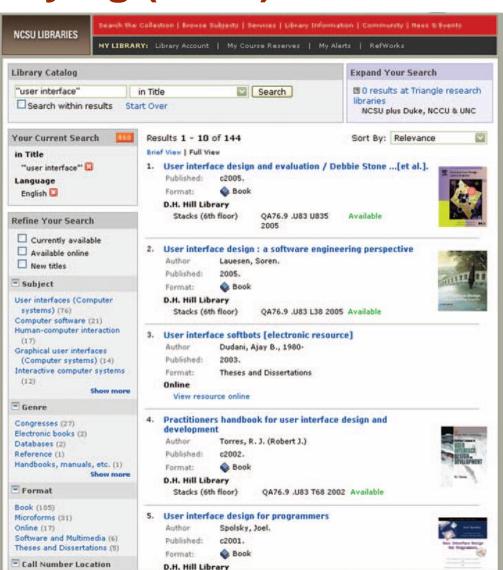# Searching in Textual Documents and Database Querying (cont.)

As users press keys on the keypad (left figure), the data are shown and a search is implicitly initiated to display the list of information that match the series of keys pressed. On the right figure, blue wedges at the edge of the screen hint at the locations of off-screen results on a map

Searching in Textual Documents and Database Querying (cont.)

A search for "user interface" returns144 results grouped into 10 pages. The menu at the upper right allows users to sort results by relevance or by date, while on the left a summary of the results organized by Subject, Genre, or Format provides an overview of the results and facilitates further refinement of the search.

Multimedia Document Searches

- **Image search**
- **Map search**
- **Design or diagram search**
- **Sound search**
- **Video search**
- **Animation search**

Multimedia document searches (cont.)

- **Image Search:**
  - **Finding photos with images such as the Statue of Liberty is a challenge**
    - Query-by-Image-Content (QBIC) is difficult
    - Search by profile (shape of lady), distinctive features (torch), colors (green copper)
  - **Use simple drawing tools to build templates or profiles to search with**
  - **More success is attainable by searching restricted collections**
    - Search a vase collection
    - Find a vase with a long neck by drawing a profile of it
  - **Critical searches such as fingerprint matching requires a minimum of 20 distinct features**
  - **For small collections of personal photos effective browsing and lightweight annotation are important**

Multimedia document searches (cont.)

- **Map Search**
  - **On-line maps are plentiful**
  - **Search by latitude/longitude is the structured-database solution**
  - **Today's maps are allow utilizing structured aspects and multiple layers**
    - **City, state, and site searches**
    - **Flight information searches**
    - **Weather information searches**
    - **Mapquest, Google Maps, etc.**
  - **Mobile devices can allow "here" as a point of reference**

Multimedia document searches (cont.)

- **Design/Diagram Searches**
  - Some computer-assisted design packages support search of designs
  - Allows searches of diagrams, blueprints, newspapers, etc., e.g. search for a red circle in a blue square or a piston in an engine
  - Document-structure recognition for searching newspapers
- **Sound Search**
  - MIR supports audio input
  - Search for phone conversations may be possible in future on speaker independent basis
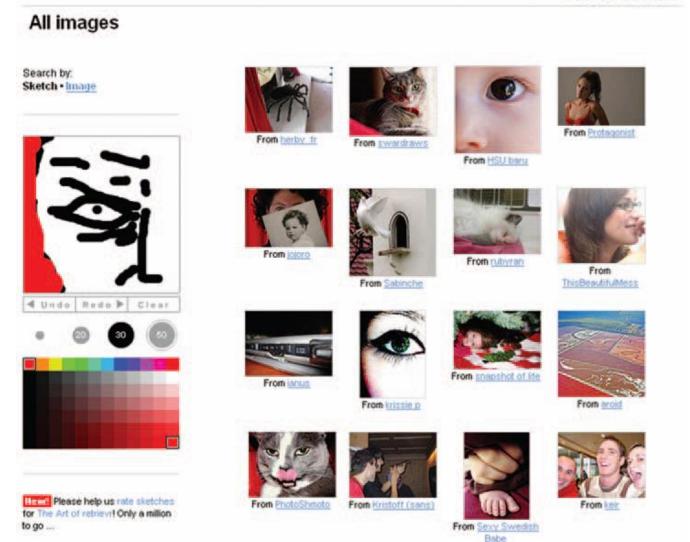- **Video Search**
  - Provide an overview
  - Segmentation into scenes and frames
  - Support multiple search methods
  - Infomedia project
- **Animation Search**
  - Prevalence increased with the popularity of Flash
  - Possible to search for specific animations like a spinning globe
  - Search for moving text on a black background

Multimedia Document Searches (cont.)

Advanced Filtering and Search Interfaces

- **Filtering with complex Boolean queries**
- **Automatic filtering**
- **Dynamic queries**
- **Faceted metadata search**
- **Query by example**
- **Implicit search**
- **Collaborative filtering**
- **Multilingual searches**
- **Visual field specification**

Advanced filtering and search interfaces (cont.)

For advanced uses there are alternatives to form fill-in query interfaces:

- **Filtering with complex Boolean queries**
  - **Problem with informal English, e.g. use of 'and' and 'or'**
  - **Venn diagrams, decision tables, and metaphor of water flowing have not worked for complex queries**
- **Dynamic Queries - Adjusting sliders, buttons, etc and getting immediate feedback**
  - **"Direct manipulation" queries**
  - **Use sliders and other related controls to adjust the query**
  - **Get immediate (less than 100 msec) feedback with data**
  - **Dynamic HomeFinder and Blue Nile**
  - **Hard to update fast with large databases**
- **Query previews present an overview to give users information and the distribution of data and thereby eliminate undesired items**
- **Faceted metadata search**
  - **Integrates category browsing with keyword searching**

Advanced filtering and search interfaces (cont.)

- **Collaborative Filtering**
  - Groups of users combine evaluations to help in finding items in a large database
  - User "votes" and his info is used for rating the item of interest, e.g. a user rating six restaurants highly is given a list of restaurants also rated highly by those who agree the six are good
- **Multilingual searches**
  - Current systems provide rudimentary translation searches
  - Prototypes of systems with specific dictionaries and more sophisticated translation
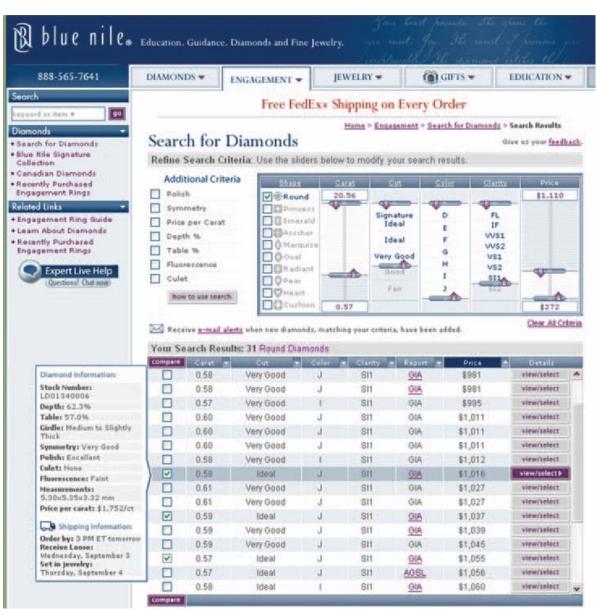- **Visual searches**
  - Specialized visual representations of the possible values, e.g. dates on a calendar or seats on a plane
  - On a map the location may be more important than the name
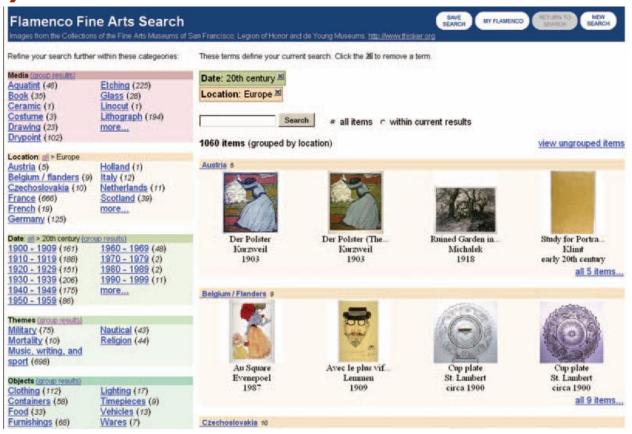 - Implicit initiation and immediate feedback

Advanced Filtering &Search Interfaces (cont.)

using dynamic queries narrows down the results of searches. Here, the double-sided sliders were adjusted to show only lower-priced diamonds with very good cut and high carat ratings.

Advanced Filtering and Search Interfaces (cont.)

an example of a faceted metadata search. Facets include Media, Location, Date, Themes, and so on. Here, two attribute values are selected (Date = 20th century and Location = Europe) with results grouped by location. The image previews are updated immediately as constraints are added or removed (another example of implicit query initiation). Clicking on a group heading such as "Belgium/Flanders" refines the query into that category, while clicking on "All" dates relaxes the date constraint.

Advanced Filtering and Search Interfaces (cont.)

users can review all waterproof binoculars in the catalog of products and browse the items in the list, grouped by manufacturer.

Each box corresponds to a pair of binoculars, and the size of the box is proportional to its price.

Green boxes are best-sellers.

Users can filter the results using the dynamic query sliders on the right. Here all the binoculars with less than three user reviews have been filtered out, leaving only 61 binoculars to consider.