Lecture 13

Chapter 13 Information Search and Visualization

Information Search and Visualization

• Who earns > $50,000 among the residents of Eugene, Oregon?

Stages of Action in Human-Computer Interaction
Introduction

• Information activities:
  – Information gathering
    • Knowing where to look and availability
  – Searching versus Browsing
    • A know-item-search versus making sense and discovering
  – Filtering
  – Information evaluation
    • Is this what I want?
  – Information analysis and interpretation
    • Summarizing information
    • Comparing information

• Information activities are on-going, iterative tasks
  – Interruption and resumption
  – Trace of the information gathering tasks
  – Archiving and annotating

Introduction

• Problem: Huge volumes of computer-stored data available:
  – Databases
    • Textual document libraries
      – Structured, relational databases
      – Relations have records
      – Records have fields, and fields have values
      – Set of items (10 to 100,000)
    • Multimedia document libraries
      – Contains images, sound, video, animations, etc
  – Websites
    – Contains network of websites with network of web pages
    – Contains rich, linked, information resources
  – Digital archives are more loosely organized
    – Directories contain metadata
  – Websites and Databases: Data mining
  – Data warehouses and data marts
  – Knowledge networks or semantic webs

Introduction

BUT searching and discovering is difficult:

• Traditional interfaces have been difficult for novice users
  – Command Languages
    • Complex commands
    • Boolsen operation
    • Unintuitive concept
    • EXAMPLE: SQL query language to relational databases

• Traditional interfaces have been inadequate for expert users
  – Difficulty in repeating searches across multiple databases
  – Weak methods for discovering where to narrow broad searches
  – Poor integration with other tools
Introduction

• Solution: Developing more powerful search and visualization methods, integration of technology with task
  - Searching in Textual Documents and Database Querying (Chapter 14.2)
    - From form in HTML instead of SQL query language
    - Customizable search options and displays using control panels
    - Google uses statistical frequency of co-occurrence of words to determine meaning
  - Multimedia Document Searches (Chapter 14.3)
    - Pattern recognition for picture searching
  - Advanced Filtering and Search Interfaces (Chapter 14.4)
    - Designers are just learning how to present large amounts of data in orderly and user-controlled ways (Chapter 14.5)
    - Information Visualization

Searching in textual documents and database querying (Chap 14.2)

• Traditional information finding resources
  - Finding aides
    - Table of contents, Indexes, Description introductions, Subject classification, Key-Word-In-Context (KWIC)
    - Preview and overview surrogates
  - Searching in structured relational database systems well established task using SQL command language
    - Users write queries that specify matches on attribute levels
      - Example of SQL command
        - SELECT DOCUMENT#
        - FROM JOURNAL-DB
        - WHERE (Date >= and Date=< 1998)
        - and (Language = English or French)
        - and (publisher = ASIST or HFES or ACM).
    - SQL has powerful features, but it requires 2 to 20 hours training
      - While SQL is a standard form-fillin queries have simplified query formulation
      - Finding a way not to overwhelm novice users is a challenge

Searching in textual documents and database querying

• New searching and querying interfaces
  - WWW search engines
    - Google, Yahoo, etc.
    - Natural language integration into text searching
    - Google uses statistical frequency of co-occurrence of words to determine meaning
    - World Wide Web search engines have greatly improved their performance by using statistical data and the information on the site’s hyperlink structure
  - WWW to Database interfaces
    - Form fills in HTML instead of SQL query language
    - Customizable search options and displays using control panels
  - 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

- Ethical problems

Searching in textual documents and database querying

- Searching & Querying User Interfaces: Basic tasks
  - Overview
    - Gain an overview of the entire collection
    - Adjoining detail view
    - The overview might contain a movable field-of-view box to control the contents of the detail view
    - Allowing zoom factors of 3 to 30
    - Fish-eye view
  - Zoom
    - Zoom in on items of interest
    - Allows a more detailed view
    - Need to maintain context
    - Particularly important for small displays
  - Filter
    - Filter out uninteresting items
    - Allows user to reduce size of search

Searching in textual documents and database querying

- User Interfaces: Basic tasks (cont.)
  - Details-on-Demand
    - Select an item or group and get details when needed
    - Useful to pinpoint a good item
    - Usually click on an item and review details in a separate or pop-up window
  - Relate
    - View relationships among items
    - Use human perceptual ability – proximity, containment, connected line, color coding
    - Example: Set director’s name, and view all movies with that director
  - History
    - Keep a history to allow undo, replay, and progressive refinement
    - Allows a mistake to be undone, or a series of steps to be replayed
  - Extract
    - Extract the items or data
    - Save to file, print, or drag to another application
Searching in textual documents and database querying

- **Example:** ZFIN database
  - WWW Genetics database for zebrafish
  - Used by international research scientists
  - Developed at UO by S. Douglas (CS) and Monte Westerfield (Neuroscience Institute), 1994-2005

<http://zfin.org>
- Search for gene "cox"
- Search for mutant "cyclops"

Multimedia document searches (Chapter 14.3)

- Searches for databases and textual documents are good, but multimedia searches are in a primitive stage
- Current multimedia searches require descriptive documents or metadata searches
- Search by date, text captions, or media is possible
- Useful to have computers perform some filtering
- New systems will incorporate powerful annotation and indexing, with better search algorithms and browsing

Multimedia document searches (Chapter 14.3)

- **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 (Chapter 14.3)

• 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
    • Example: www.mapquest.com
  – Mobile devices can allow “here” as a point of reference

Multimedia document searches (Chapter 14.3)

• Design/Diagram Searches
  – Some computer-assisted design packages support search of designs
  – Alters 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 for 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

Advanced filtering and search interfaces (Chap 14.4)

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
  – Flameco
Interactive Graphics

Advanced filtering and search interfaces (Chap 14.4)

- Collaborative Filtering
  - Groups of users combine evaluations to help in finding items in a large database
  - User “votes” and her/his info is used for rating the item of interest
  - E.g. a user rating sex 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

Summary

Problem: Huge volumes of computer-stored data available
- Databases
  - Structured relational Databases
  - Multimedia document libraries
- Websites
  - Websites and Databases: Data mining

BUT searching and discovering is difficult:
- Traditional interfaces have been difficult for novice users
- Traditional interfaces have been inadequate for expert users
- Difficult in reporting and live access random databases
- Weak methods for discovering where to narrow broad searches
- Poor integration with other tools

Solution: Developing more powerful search and visualization methods, integration of technology with task
- Searching in Textual Documents and Database Querying (Chapter 14.2)
- Multimedia Document Searches (Chapter 14.3)
- Advanced Filtering and Search Interfaces (Chapter 14.4)
- Designers are just learning how to present large amounts of data in orderly and user-controlled ways

"Information Visualization"