Information Search and Visualization

Chapter 14

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
  - Searching versus browsing
  - Filtering
  - Information evaluation
  - How do I know?
  - Information analysis and interpretation
- 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
    - Structured relational Databases
      - contains relations and a schema to describe the relations
      - relations have records
      - records have fields, and fields have values: Textual document libraries
    - Multimedia document libraries
      - Contains images, sound, video, animations, etc
      - Digital archives are more loosely organized
  - Websites
    - Contains network of websites with network of web pages
    - Gigantic information resource
    - Contains text, audio, video, graphics, programs
  - 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
  - Limited operators
  - Unusually concepts
  - 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 title in HTML, instead of SQL, query language
    - Controllable search options and displays using control panels
    - Design using statistical frequency of occurrence of words to determine meaning
  - Multimedia Document Searches (Chapter 14.3)
    - Audio recognition for picture searching
    - Advanced Filtering and Search Interfaces (Chapter 14.4)
  - Designers are just learning how to present large amounts of data in order and user-controlled ways (Chapter 14.5)
  - Information visualization

Searching in textual documents and database querying (Chap 14.2)

- Traditional information finding resources
  - Finding aids
    - 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
    - WHERE title = 'title
    - AND author = 'author
    - AND language = 'language
  - SQL has powerful features, but it requires 2 to 30 hours training
  - While SQL is a standard form-Bit 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
  - Word web sites, search engines have greatly improved search performance by using additional clues and the information on the site of hypertext document

- WWW to Database interfaces
  - From title in HTML, instead of SQL, query language
  - Controllable 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

- **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
  - Zoom
    - Zoom in on items of interest
    - Need to maintain context
    - Particularly important for small displays
  - Filter
    - Filter out uninteresting items
    - Allows user to reduce size of search

- **Details-on-Demand**
  - Select an item or group and get details when needed
  - Useful to preview a good item
  - Usually click on an item and review details in a separate or pop-up window

- **Rota**
  - View relationships among items
  - Use human perceptual ability – proximity, containment, connected line, color coding
  - Example: Set directors 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

- **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
  - Allows searches of diagrams, blueprints, newspapers, etc.
  - E.g. search for a red circle in a blue square or a paton in an engine
  - Document structure recognition for searching newspapers

- Sound Search
  - MRI supports audio input
  - Search for phone conversations may be possible in future on speaker independent basis

- Video Search
  - Provides an overview
  - Segmentation into scenes and frames
  - Support multiple search methods
  - Info media 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 Home/Info and Blue File
  - 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 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

Information visualization (Chapter 14.5)

- Information visualization Definition
  - Use of interactive visual representations of abstract data to amplify cognition
  - Categorical variables and the discovery of patterns, trends, clusters, outliers, and gaps
  - Innovative ways of visualizing the data
- Compare to Scientific visualization
  - Continuous variables, volumes and surfaces

3D Histogram

Who earns > $50,000?
Tree Map Visualization

How a Tree Map Works

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
  - Limited to language
- Traditional interfaces have been inadequate for expert users
  - Difficult to express queries across multiple relations
- Weak methods for discovering where to navigate 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 14.3)
- Multimedia Document Retrieval (Chapter 14.4)
- Web Information Retrieval (Chapter 14.5)
- Designing user interfaces that present large amounts of data in orderly and user-controlled ways
  - "Information Visualization"