Lecture 17

Chapter 14 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
  – 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
      • Database on a server and a schema to describe the relations
      • Relations have records
      • Records have fields, and fields have values
      • Set of items (10 to 100,000)
    – Structured relational databases
      • Relations and fields have values
      • Set of items (10 to 100,000)
  – Multimedia document libraries
    – Digital and text are more loosely organized
    – Directly usable information
  – Websites
    • Contains network of websites with network of web pages
    • Contains information resources
    • Contains text, audio, graphics, video, 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
    • Boolean operators
    • 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 free-text HTML instead of SQL query language
    • Controllable word selection and displays using control panels
    – Design uses statistical frequency of occurrence of words to determine meaning
  – Multimedia Document Searches (Chapter 14.3)
    • Web recognition to 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

• 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
      • 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
    – Finding a way not to overwhelm novice users is a challenge

• New searching and querying interfaces
  – WWW search engines
    • Google, Yahoo, etc.
    • Natural language integration and text searching
    – Design uses statistical frequency of occurrence of words to determine meaning
    – WWW, Web search engines have greatly improved search performance by using statistical usage and the information in the site's hyperlink structure
  – WWW to Database interfaces
    • From free-text 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

• 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
    • Fisheye view
  – Zoom
    • Zoom in on items of interest
    • Allows a more detailed view
    • 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
  – 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 allowing 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

• **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
  - Example: search for a red circle in a blue square or a piston in an engine

• **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 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 fillin query interfaces:

• Filtering with complex boolean queries
  - Problem with informal English, e.g. use of “and” and “or”
  - Flow diagrams, decision tables, and metaphor of water flowing have not worked
• 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/or 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 sex are good.

- **Multilingual searches**
  - Current systems provide rudimentary translation searches.
  - Proposals 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

http://www.hivegroup.com/
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 to repeat queries across multiple 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 just learning how to present large amounts of data in orderly and user-controlled ways (Chapter 14.5)
  - "Information Visualization"