Ad-hoc IR in text-oriented IS

- The context (L1)
- Querying and matching (L2, L3)
- How to evaluate results (L5)
- How it all works internally (L4, L10 - algs)
- Better organization and visualization of search results (L7)
- Techniques for improving search effectiveness: RF, QE, HCI (L8); UM (L9)
- What else?
A Broader view of the course

- Classic Information Retrieval
- Improving Classic Information Retrieval
- What else beyond Classic IR?
  - Multimedia IR (L6)
  - Alternative ways of Information Access: Hypertext, Information Visualization (L6)
  - User Modeling for IR (L6, L9)
  - Web IR (L11)

“Alternative Techniques”

- Natural Language Processing
- Citation Processing
- Multimedia IR
- Hypertext: Links and browsing
- Information Visualization
- Dynamic queries
- Adaptive hypermedia / navigation support
- Information Filtering (L9)
- Implicit Queries (recommender systems) (L9)
Citation processing

- Citations as extra information about a document
  - Co-citation as a similarity measure
  - Link between documents citing same source
- Citations as a measure of importance
- Citation-based navigation
- Examples: WebOfScience, CiteSeer
- Problem: no citation standard

What’s in a citation?

- Authors
- Title
- Source data
  - Journal: Title, publisher
  - Book: Title, editor, publisher
  - Conference: Title, date, location
- Location in the source
Citation processing

- Is it only relevant to scientific papers?
- Where else similar technologies can be used?

Hypertext: Definition

- Shneiderman [SK89]:
  - "a database that has active cross-references and allows the reader to "jump" to other parts of the database as desired".
- A hypertext is a database.
- The typical user action is a jump between parts of the database.
- Apart from pieces of information, called *nodes*, the database contains links between related nodes.
Hypertext

Hypertext: Anatomy

- **Nodes**
  - scrolling (WWW “pages”)
  - fixed size (HyperCard)
  - several pages
- **Anchors** *(hot world, hot spot)*
- **Links**
  - Anchor to node
  - Source anchor to destination anchor *(Intermedia)*
Hypermedia

- J. McDaid [McDaid-91]:
  - Hypermedia is ... an extension of Nelson's earlier coinage, "hypertext" (for non-sequential writing), hypermedia implies linking and navigation through material stored in many media: text, graphics, sound, music, video, etc.

History of Hypertext: Founders

1945: Vannevar Bush proposes *Memex* in his article "As We May Think".
1965: Ted Nelson introduces *Xanadu* and coins the term hypertext.
1967: Andries van Dam develops the *Hypertext Editing System* at Brown University, the first working hypertext.
1968: Doug Engelbart gives a demo of *NLS*, a part of the Augment project, started in 1962.
History of Hypertext: Promoters

1975: A team at CMU, headed by Robertson, develops the ZOG system, later KMS.
1978: A team at MIT, headed by Andrew Lippman, develops the Aspen Movie Map
1985: Janet Walker develops the Symbolics Document Examiner, the first hypertext system used by "real" customers.
1985: Several other hypertext systems - NoteCards (Xerox), Intermedia (Brown University) ...

History of Hypertext: Real Word

1986: OWL introduces Guide for the Macintosh, the first widely available hypertext system, based on the Unix Guide (Peter Brown, University of Kent)
1987: Apple delivers HyperCard free with every Macintosh
1987: The ACM organizes the first Conference on Hypertext
1990: The World Wide Web delivers Hypertext to millions
**Link Behavior**

- **Jump** (goto) to another node
  - node shown in the same window
  - node shown in different window (NoteCards)
- Pop-up window with definition (Guide, HyperCard)
- Stretchtext

**Another Node**

Now jump back!
Link Behavior

- **Jump** (goto) to another node
  - node shown in the same window
  - node shown in different window (NoteCards)

- Pop-up window with definition (Guide, HyperCard)

- **Stretchtext**

  Pop-up Window activates when you click the mouse button and disappears when you release the button
Link Behavior

- **Jump** (goto) to another node
  - node shown in the same window
  - node shown in different window (NoteCards)
- **Pop-up window** with definition (Guide, HyperCard)
- **Stretchtext** - expansion of text (Guide)
  - additional text is inserted when you click!
  - you may also **collapse** the expansion back

Hypertext Navigation

- **Goals**
  - Understanding “what’s there”
  - Find some information
  - Learn something
- **Problems** ⇒ **Navigation aids**
  - Where am I?
  - Where I can go from here?
  - Where I should go?
Navigation aids

- Metaphors (museum, travel)
- Getting the whole picture
- Going back
- Avoiding loops
- Giving more information to decide
- Using structure (sequence, tree)
- Adaptive navigation support

Getting the whole picture...

... without getting lost
- Guided tours
- Landmarks
- Index pages
- Maps
  - Local map
  - Global map
- Fisheye views and hyperbolic trees
Going back

- "Back" button
- History list
  - partial: path only (Netscape)
  - full: all visited nodes (IE)
- History tree
- Bookmarks

Where I should (should not) go?

- Loops and inefficient navigation
  - Unique anchors
  - Bread crumbs
  - Highlighting visited links
- More information to decide
  - Typed links
  - Sneak preview
- Adaptive navigation support
Typed Links

- Verbal typing:
  - subheading, prefix
- Spatial typing:
  - location, grouping
- Typed by visual cues
  - Font: size, style, type
  - Icon/bullet
  - Color of font, background, icon/bullet

Sneak Previews

More at http://www8.org/w8-papers/4b-links/visual/visual.html
Using the structure

- Sequence (HyperCard)
  - forward
  - backward
  - home
- Tree
  - top
  - ancestor
  - descendants
  - siblings
- Table of contents

Navigation aids: summary

- backtracking
- bread crumbs
- highlighting links
- unique anchors
- sneak preview
- typed links
- history list
- history tree
- landmarks
- bookmarks
- maps
- indexes
- fish-eye views
- guided tours
Hypertext Course

- http://wwwis.win.tue.nl/2L690/
- Your Student id is your pitt e-mail

From IR to Hypertext

- Browsing offers an alternative and powerful way of information access
- What can we do if there are no links between documents and browsing is not possible?
- Similarity-based navigation
  - Document similarity
  - Access similarity
Information Visualization

“Transformation of the symbolic into the geometric”
(McCormick et al., 1987)

“... finding the artificial memory that best supports our natural means of perception.”
(Bertin, 1983)

The depiction of information using spatial or graphical representations, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system (Hearst 03).

Symbolic Data Processing

<table>
<thead>
<tr>
<th>City</th>
<th>Latitude</th>
<th>Longitude</th>
<th>City</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apalachicola, FL</td>
<td>29.73</td>
<td>85.03</td>
<td>Naples, FL</td>
<td>26.13</td>
<td>81.80</td>
</tr>
<tr>
<td>Daytona Beach, FL</td>
<td>29.18</td>
<td>81.05</td>
<td>Orlando, FL</td>
<td>28.43</td>
<td>81.32</td>
</tr>
<tr>
<td>Fort Myers, FL</td>
<td>26.58</td>
<td>81.87</td>
<td>Saint Peters, FL</td>
<td>27.92</td>
<td>82.68</td>
</tr>
<tr>
<td>Ft Lauderdale, FL</td>
<td>26.07</td>
<td>80.15</td>
<td>Sarasota, FL</td>
<td>27.40</td>
<td>82.55</td>
</tr>
<tr>
<td>Jacksonville, FL</td>
<td>30.23</td>
<td>81.68</td>
<td>Tallahassee, FL</td>
<td>30.38</td>
<td>84.37</td>
</tr>
<tr>
<td>Key West, FL</td>
<td>24.55</td>
<td>81.75</td>
<td>WPalm Beach, FL</td>
<td>26.68</td>
<td>80.12</td>
</tr>
</tbody>
</table>

- Two pairs of closest cities
- Two most distant
- Most centered – shortest combined distance to all others

Adopted from M. Spring
Symbolic Processing vs Visualization

- Two pairs of closest cities
- Two most distant
- Most centered – shortest combined distance to all others

Why Information Visualization?

- Problem:
  - HUGE volume of data: How to comprehend them?
  - How to locate relevant data?

- Solution
  - Take better advantage of human perceptual system: convert information into a graphical representation.

- Challenge
  - How to convert abstract information into useful graphical form that can provide an added value?

Adopted from M. Hearst
Information Visualization: an Early Success Story

Mystery: what is causing a cholera epidemic in London in 1854?

Adopted from M. Hearst

The Power of Visualization

Illustration of John Snow’s deduction that a cholera epidemic was caused by a bad water pump, circa 1854.

Horizontal lines indicate location of deaths.


Adopted from M. Hearst
From Visual Explanations by Edward Tufte,
Graphics Press, 1997

Adopted from M. Hearst

Images from yahoo.com

Adopted from M. Hearst
The Power of Visualization

1. Start out going Southwest on ELLSWORTH AVE Towards BROADWAY by turning right.
2. Turn RIGHT onto BROADWAY.
3. Turn RIGHT onto QUINCY ST.
4. Turn LEFT onto CAMBRIDGE ST.
5. Turn SLIGHT RIGHT onto MASSACHUSETTS AVE.
6. Turn RIGHT onto RUSSELL ST.

Image from mapquest.com

Adopted from M. Hearst

The Power of Visualization

Line drawing tool by Maneesh Agrawala http://graphics.stanford.edu/~maneesh/

Adopted from M. Hearst
Which Information Visualization?

- InfoViz is a mixed set of techniques!
  - grouped by their “non-textual” nature
- Different InfoViz techniques support different aspects of information access
  - Better interfaces for query formulation
  - Graphical interfaces for the search process
  - Better presentation of search results
  - InfoViz as an alternative to search, an alternative approach for information access

Where we will learn InfoVis?

- Improving search process (L8)
  - Better interfaces for query formulation
  - Graphical interfaces for the search process
- Presentation (L7)
  - Better presentation of search results
- Alternative access techniques (L6)
  - InfoViz as an alternative to search, an alternative approach for information access
Information visualization for Information Access

- Information visualization can provide an alternative to search and be used as a different information access paradigm.
- Information visualization:
  - Similar to browsing: finding documents by navigation and manipulation.
  - Uses more expressive 2D and 3D representation.
  - Allows to see “the whole picture”.

Information visualization for Information Access

- Geospatial visualization (real maps)
- Information space visualization (information maps)
- Visualization of trees and hierarchies
- Visualization of linked data, hypertext and the Web
- Interfaces for exploration of formatted data (tables, temporary data, …)
- Adaptive Information Visualization
Information Space Visualization (information maps)

- Tabular (semi-spatial 2D)
  - Self-Organized Maps (SOM) also known as Kohonen maps
- Spatial: 2D or 3D
  - Force-based visualization
  - Multidimensional Data Scaling (MDS)
  - FastMap algorithm
- Information Landscapes
- Linked graphs
  - PathFinder algorithm

Tabular “whole picture”

[Diagram showing a complex tabular visualization with connected nodes and labels]
Graphical “whole picture”
Visualization of “Linked” Data

- Visualization of trees and hierarchies
  - Hyperbolic Browser
  - Treemap
- Visualization of general linked data, hypertext and the Web
  - Global maps
    - Graph Drawing problems
    - Generalized Fisheye Views
  - Local area maps
    - The Brain

Hyperbolic Trees

TreeMap

The TreeMap was introduced in 1991 by Johnson & Shneiderman

Idea:
– Show a hierarchy as a 2D layout
– Fill up the space with rectangles representing objects
– Size on screen indicates relative size of underlying objects.

TreeMap History

– TreeViz (Mac, Johnson, 1992)
– NBA-Tree (Sun, Turo, 1993)
– Winsurfer (Teittinen, 1996)
– Diskmapper (Windows, Micrologic)
– Treemap97 (Windows, UMd)
Early TreeMap: File System

Early TreeMap Problems

- Too disorderly
  - What does adjacency mean?
  - Aspect ratios uncontrolled leads to lots of skinny boxes that clutter
- Color not used appropriately
  - In fact, is meaningless here
- Wrong application
  - Don’t need all this to just see the largest files in the OS
Modern TreeMaps

- Applications (SmartMoney and Honeycomb)
  - HistoryWired: http://historywired.si.edu/index.html
  - Peets Coffee Selector
    http://www.peets.com/selector_coffee/coffee_selector.asp

- Tools
  - HoneyComb: http://www.hivegroup.com/
  - U of Maryland API
  - SourceForge API

A Good Use of TreeMaps and Interactivity: Map of the Market

www.smartmoney.com/marketmap

Adopted from M. Hearst
Treemaps in Peets site

Analysis vs. Communication

- MarketMap’s use of TreeMaps allows for sophisticated analysis
- Peets' use of TreeMaps is more for presentation and communication
- This is a key contrast
Successful Application of TreeMaps

- Think more about the use
  - Break into meaningful groups
  - Fix these into a useful aspect ratio

- Use visual properties properly
  - Use color to distinguish meaningfully
    - Use only two colors:
      - Can then distinguish one thing from another
      - When exact numbers aren’t very important

- Provide excellent interactivity
  - Access to the real data
  - Makes it into a useful tool

Adopted from M. Hearst

Exploring Formatted Data

- Each information item has a number of formatted fields with various data
- The kind of data can help in choosing proper visualization (numeric, temporary, geospatial, qualitative)
- Tables (XEROX PARC): TableLens
- Starfield and Dynamic Queries (UMD)
  - HomeFinder
  - FilmFinder
  - Lifelines
Table Lens

- Super Spreadsheets
  - Combines overview + details in an integrated view
  - Focus + Context allows for compressed representation
  - Sorting multiple columns allows patterns to emerge
  - Represents nominal data in a way that allows patterns to appear


Dynamic Queries: HomeFinder

- Attribute filter with immediate visual feedback
- “Starfield” geospatial presentation
Dynamic Queries: FilmFinder

Ahberg & Shneiderman. Color plate 1. The FilmFinder.

Dynamic Queries: FilmFinder

Ahberg & Shneiderman. Color plate 2. Categories have been selected, the displayed is zoomed.
Dynamic Queries

- Query is issued using GUI controls
- Query response is visible and visualized immediately
- Query can be dynamically modified
- Attributes can be explored
- There is tight coupling between displays and controls
- Examples: FilmFinder, LifeLines...

LifeLines: Mastering the Temporal Dimension

- LifeLines: Visualizing Personal Histories. B. Milash, C. Plaisant, A. Rose (University of Maryland). CHI 96
Dynamic queries: where else?

- Name 3 possible application areas

More Dynamic Queries: Influence/Attribute Explorer

- Visualization for Functional Design, Bob Spense, Lisa Tweedie, Huw Dawkes, Hua Su, InfoVis 95
Browsing InfoSpace

- Can we support the user who is browsing?
- No query: need to know information need or user model
  - How to deduce that? ⇔ User modeling
- No “search results”: need to guide the user to good pages
  - How to guide? ⇔ Adaptive navigation support and adaptive visualization

Knowledge Sea: Adaptive IV