L697: Information Visualization
(Spring 2001)

Project 4
Final Project

- Concentrates on interaction.
- Uses data mining (project 2) and layout algorithms (project 3) to visualize meaningful information.

Differs from previous projects in that you will
- Work with a faculty/staff ‘client’.
- Create an interactive visualization that might be used beyond this course.
- Write a scientific paper.
Final project options

We got 6 projects that will be presented by
1. Title of the project
2. Main contact person
3. Data you have (data format, amount & sample subset)
4. Intended users (researcher, lay persons)
5. Visualization goal (what do to visualize)
6. Desired interactivity (zoom, filter, ...)
   + Vision of how the visualization should look & interact.

   + Similar systems
   + Katy’s vision of the interface
TransPAC Network traffic visualization

2. John Hicks, jhicks@iu.edu
3. Top10 TransPAC users per five minute period. http://radioflyer.uits.iupui.edu/top10
5. Visualize most frequent users, collaborators, types of traffic.

“Little mining - lots of parsing – animated or cleverly color coded map”

K. Borner
L697: Information Visualization
MIDS Internet Weather Report animation frames

(Source: MIDS, http://www.mids.org)
Abilene weather map

(Source: http://hydra.uits.iu.edu/~abilene/traffic/)
TrasPAC sketch

- Type of traffic
- Time of day

Line color = type of traffic or time of day
Line thickness = amount of traffic

Manual acquisition of geographic locations
Cognitive Science’s papers & board of reviewers

2. Rob Goldstone, Psychology, rgoldsto@indiana.edu
3. Reviewers plus their areas of specialization
4. Web browsers, people interested in the journal
5. Visualize CogSci papers, show how well different areas of CogSci are covered by reviewers. See similarity among reviewers. Match up articles with reviewers.

Optional: Filter out keywords. Search paper, reviewer, see their areas of expertise. See all reviewers with expertise in particular area.

“Highly visible – LSA intensive – map”
JAIR Space (Foltz, 1999)

(Source: http://www.infoarch.ai.mit.edu/jair/)
Antarcti.ca's Visual Net

Use IE to explore map.net.

K. Borner
How to visualize information on

how many documents a certain cluster contains

AND

semantic relationships among documents in each cluster?
Semantic TreeMap

*combines treemap and semantic layout*

Document by term matrix -> LSA -> sim matrix -> clustering -> cluster hierarchy -> select best partition, preserve number of documents in each cluster -> visualize as treemap + apply spring algorithms for each area in treemap.

**Result:**
Treemap that visualizes number of documents in each topic area. If you click in area you get to see semantic relationships between the documents in each area.

**Optional:**
Zooming – preserves context.
Semantic Treemap

Results in area expansion and semantic layout becomes visible.
CogSci sketch

Map of papers or reviewers

- Papers
- Reviewers

Search

Reviewer-keywords, paper-full_text -> LSA-> SemanticTreemap
Need to download articles from
http://www.elsevier.nl/gej-ng/10/15/15/show/toc.htm
Using Spring Embedder and Hyperbolic Visualization Algorithms to Display Term Relationships from a Concept Discovery System

2. Javed Mostafa, jm@indiana.edu
3. Six thousand titles and abstracts of medical documents will be converted to weighted vectors using both tf*idf and LSI; the weighted vectors will be the data set used by the Hyperbolic and Spring embedder algorithms.
4. Mainly researchers in health informatics area.
5. To identify implicit relationships between drugs and diseases based on content of a large medical corpus.
6. **Zooming** - User would be able to zoom into individual document nodes to identify and select neighbors in the hyperbolic representation; users will also be able to zoom into term nodes in the final spring embedder representation to identify nearest neighbors. **Filter** - Users will be able to select a subset of documents from the hyperbolic representation for application of spring embedder.

“Medical - discovery - map”
Vx Insight, Sandia Lab

Is a tool for discovering relationships within very large databases. It reveals the implicit structure of the data.

Show
Movies/Insight_demo.exe

(Source: <http://www.cs.sandia.gov/projects/VxInsight.html>
Self-organized Feature Maps

Kohonen, WEBSOM document map of over million documents from over 80 Usenet newsgroups produced by the Neural Networks Research Centre at the Helsinki University of Technology, Finland.

(Source: http://websom.hut.fi/websom)
Diseases & Drugs sketch

Map of diseases and/or drugs

Documents-keywords -> LSA-> SemanticTreemap
Interactive tree interface to Roget's thesaurus categories

2. John Old, UITS, jold@indiana.edu

3. Data Set: Roget's Thesaurus Hierarchy
   Class - Rsubclass - Lsubclass - Category

4. Researchers, but hopefully it would be transparent enough for a lay person to also use.

5. Roget's Thesaurus upper level hierarchy enabling selection of a Category. Each leaf node would be a Category name with a URL to a database which returns the full Category (words) as a web page.

6. Clicking on a leaf node should pass a parameter to the following URL
   http://ella.lib.indiana.edu/~jold/scripts/categoryxref.cfm?inputstring=1 where "1" is the category name).
A Spring Display of Words and Senses derived from the Roget Thesaurus

2. John Old, UITS, jold@indiana.edu

3. ASCII data set of words and sense numbers; Ternary relation separated by PIPE ( | ) e.g. node1|POS|node2

4. Researchers, but hopefully it would be transparent enough for a lay person to also use.

5. Display a set of related words returned from a query. It may consist of two node types, Sense and Word, or just Word nodes connected if they share a Sense.

6. No Zoom needed, but a word node should be able to invoke a new query to display a new set. POS is included so that Sense nodes can be colored by Part-of-Speech (noun, verb, adjective=red, green, blue etc).

“Very special — search enhancing - hyperbolic tree”
SiteBrain by TheBrain (Harlan Hugh, 1998)

Information modeling technology to the Web.
Screenshot is mapping the Web site for the graduate school of architecture and planning at Columbia University.

(Source: http://www.thebrain.com/products/sitebrain/)
3D Typographic Maps - Thinkmap

Tool for displaying complex information from database. Illustrates interrelationship of items in a data set. Strong visual interaction, 3D moving type.

Demo at http://www.plumbdesign.com/thesaurus/

K. Borner
Roget sketch

Hyperbolic tree layout

Search above

Related terms get color coded and move towards middle of screen.

1,000 leaf nodes
Bookmark visualization

2. Katy Borner, katy@indiana.edu
3. Howard’s Bookmark list of 5000 entries + full text of associated web pages
4. (SLIS?) Web users
5. Semantically organize web links in 3-D space, preserve bookmark hierarchy. Label clusters.
6. Click on node retrieves web page.

“Semantic link layout for SLIS’s Memory Palace
http://vw.indiana.edu/
Bookmark sketch

Full text of links -> LSA -> Semantic Treemap

Map of labeled link clusters

Click on link dot displays Web page here.

Search

HCI

Search for HCI highlights all HCI related documents
Newsgroup Votes

2. John Paolillo, paolillo@indiana.edu

3. Newgroups messages since 1989 that pertain to the comp.hierarchy. The final votes (contained in the last message of each file) need to be extracted and entered into a large matrix of yes/no/abstain votes (participant by newsgroup). In addition, the date of the creation of each newsgroup will be retained in a separate file. 25 MB raw, it will be substantially smaller processed.

4. Newsgroup users, researchers

5. Visualize which people voted how for which newsgroup.

“Parsing intensive, semantic vote map”
Newsgroup sketch (I)

Correlation between newsgroup descriptions and the votes they receive.
LSA -> Spring -> color coding of nodes according to vote result
Grouping according to newsgroup similarity

Correlation between voters according to the newsgroups they voted for (organized by date of posting).
LSA -> Spring -> color coding of nodes according to type of newsgroup
Grouping according to similar voting behavior

Color code day of newsgroup creation …
Newsgroup sketch (II)

Y-axis: newsgroup hierarchy

X-axis: voters

No semantics
Guidelines for project write-up
(adopted from Shneiderman, http://otal.umd.edu/Olive/Class/)

Title, Authors, Addresses, Electronic Mail, and Date

Abstract: 100-150 words
Overview of visualization problem, data mining and visualization approach utilized, results, and discussion.

1. Introduction 1-2 printed page(s)
Overview of the visualization problem, possibly including:
• Personal encounters with the problem
• Current interface
• Review of other systems / IVs
• Discussion of extracts from relevant textbooks
• Relevant psychological or other theories
2. Visualization Goal 1-2 printed page(s)
discuss users, their tasks & available data

3. Data Mining/Parsing 2-3 printed page(s)
Present and discuss the data mining/parsing approach(es) you used

4. Visualization & Interaction 3-4 printed page(s)
Present and discuss the visualization algorithm(s) used. Which psychological features are met? Discuss the interaction possibilities – strengths & problems. Include image(s) of your visualization!

5. Discussion 1-2 printed page(s)
User feedback, complexity & scaling issues. Desirable modifications & extensions.
Acknowledgements  
*a few sentences*

Thanks to people who helped you. If you work in a team indicate who worked on which part.

References  
*5-20 references*

Links to all sources you used & to related work. Complete citations!

Appendices  
*2-50 pages*

possibly including:

A. Raw data
B. Compiled data
C. Statistics runs from computer
D. Screen plots
Final project timeline

03-08-2001  Presentation of Project Options
03-22-2001  (1 week later + Spring Break)
  Present your project sketch during lab
04-12-2001  (3 weeks later)
  Present your (partially functional) interface during lab
04-18-2001  (1 week later)
  Handin draft write-up for feedback
04-25-2001  (1 week later)
  Handin final write-up (ca. 10 pages)
04-26-2001
  Final project demo
Projects are graded according to

- The technical quality, including its reliability, ease-of-use, internal consistency, robustness, and performance, and
- The quality of the content, including the accuracy and completeness of information, the expressiveness and clarity in communication of ideas, and the appropriateness of the attribution(s) for the work of others.

As for project 4, I will especially look at the
- Appropriate selection of data set, data mining algorithm, and visualization & interaction approach that match your users and their tasks.
- Reliability, ease-of-use, internal consistency, robustness, and performance of the implementation.
- Quality of Write–up.