M259 Visualizing Information

Jan 14: DATA SOURCE

George Legrady, legrady@mat.ucsb.edu
Yoon Chung Han <hanyoonjung@gmail.com>

This Week:

TUES: Jan 14
Discuss Data Source

THUR: Jan 16
a) Karl Yerkes Presentation on database structure
b) Yoon will follow with SQL examples

2 Directions for Data Analysis:
1. The study of the data organization, its anomalies, history, etc.
2. The exploration of content (cultural information)

Data Visualization Focus Areas:

DATA SYSTEMS
- Data retrieval and analysis
- System anomaly analysis

CULTURE
- Cultural patterns:
- Historical Context: How has the library transformed itself

VISUALIZATIONS
- What are ways to visualize data (content focus)
- Invent new visualizations (exploration of the visual language)

ALGORITHMS
- Apply Algorithms: What are interesting algorithms by which to process data
- Create new algorithms
Data Source

- Patrons check out books, cds, dvds from the Seattle Public Library
- Each time someone checks out a movie, book, cd, data is received by the hour
- Appx 30000 per day; 10 million annual;
- Over 67 million datasets since September 2005

Rem Koolhaas Visionary Architecture

- Seattle Public Library main collection of books, government publications, periodicals, audio visual materials and the technology to access and distribute information online.
- The building is divided into eight horizontal layers, each varying in size to fit its function.
“Making Visible the Invisible” ArtWork

- Seattle Public Library artwork opened Sept. 7, 2005 extended to 2019
- Provides raw data for the study of cultural library patron practices over time
- Possibly the longest running public dynamic data visualization project

Project Concept

- Conceptualize the library as a “Data Exchange Center”
- Correlation is made between the flow of data (books, DVD, etc.) leaving the library and
- What patrons considers interesting information at any specific time
- This data is information that can be calculated mathematically and represented visually

Rem Koolhaas Concept

HTTP://WWW.SPL.COM/~MIXING.CHAMBER
Data Source
- Patrons check out books, cds, dvds from the Seattle Public Library
- Each time someone checks out a movie, book, cd, data is received by the hour
- Appx 30000 per day; 10 million annual;
- **Over 67 million datasets** since September 2005

Dewey Classification System

**Ten topics each subdivided into 100 subclasses:**

- 000 - Generalities
- 100 - Philosophy & Psychology
- 200 - Religion
- 300 - Social Science
- 400 - Language
- 500 - Natural Science & Mathematics
- 600 - Technology & Applied Sciences
- 700 - Arts
- 800 - Literature
- 900 - Geography & History

MySQL:
```sql
SELECT itemNumber, cout, collcode, itemtype, barcode, title, callNumber, deweyClass, subj FROM inraw WHERE year(cout) = 2007 AND month(cout) >= 1 AND month(cout) <= 4 LIMIT 50;
```

Daily Volume Activity (see at outlog/inlog)
Data is Multivariate

- Data is multivariate. Each transaction includes numeric, ordinal, interval scale (time, date), string, and other classification data.
- Include:
  - ItemType (bks, cds)
  - Collection date
  - Check-out/check-in hour/day
  - Title
  - Dewey Classification
  - Keywords
  - Unique IDs (barcode, etc.)

Daily Volume Activity (see at outlog/inlog)

Data is Multivariate

- Data search: Knowledge discovery
- Data analysis: What patterns emerge
- Data formulation through algorithmic processing
- Translation into visualization
- Correlation with External Data
- Publication
Review Course Links:

- Dewey Classification System
- Library itemtypes
- Seattle Library Item Search
- Legrady Project: Top 20 Dewey
- Previous Course Projects
- MySQL Queries

1st Assignment: A Compelling Data Query

- **State a Question**: Can be cultural, system based, compare data, etc. *Should be exploratory or compelling….*
- **Translate into a Query**:
- **Report Result**: (a Data printout to .csv file)
- **Processing Time**: How long the search
- **Comment & Analysis**: Report what you found through the data

Assignment is due Tues Jan 21

How to Proceed

- **Ask a general question**
- **Change syntax if necessary**
- **Review Results**
- **Ask question differently**
- **Review Results**
- **Go into more detail**

Begin with Inraw – by the week-end switch to more specific tables with MySQL innerjoin function
Data Analysis Directions

- Statistical
- Narrative
- Visual Explorations

Superhero Popularity, Domagoj Baricevic

Hot Topics 2005 – 2010, Patrick Rudolph

"Catcher in the Rye" – Anis Haron

Map of 1st checkout of New Item (System Exploration by Karl Yerkes)
What to map to show "Change Over time"?

Search for patterns: What exactly to look for?

External correlation (news events)? relevant?

Feedback: How does the visualization impact on circulation?

Look to the future: Technology changes every 3 years. How will the project live on for 10 years?

RJ Duran Frequency Pattern Tree Algorithm for the word “Water” in relation to other Words

• Over 75 million datasets
• Focus on the syntax visual language
• MySQL for data query
• Processing: Computational design

Integrate Your Expertise

**Computer Science:** Integrate complex algorithms to visualization

**Statistics:** Implement statistical probability problems to data analysis and visualization

**Sound/Signal processing:** Consider data as signal and explore translation between sonic, signal and visual patterns

**Social Science:** Identify cultural patterns, changes, transformations

**Geography:** Explore spatial mapping

**Cinematic/Literary:** Explore data pattern as narrative development
Data Processing Functions

- **Validation**: Ensuring that supplied data is "clean, correct and useful"
- **Sorting**: Arranging items in some sequence and/or in different sets
- **Summarization**: Reducing detail data to its main points
- **Aggregation**: Combining multiple pieces of data (from various sources)
- **Analysis**: Collection, organization, analysis, interpretation and presentation of data
- **Reporting**: List detail or summary data or computed information

Some References

- Atlas of Science, Katy Borner
- Graphics of Large Datasets, Unwin, Theus, Hofmann
- Visualizing Data, Ben Fry
- Robert Kosara, social analysis

Previous student projects:
http://vislab.mat.ucsb.edu/courses.html