Information Visualization for the Future Generation Catalogue

Charles-Antoine Julien
Assistant Professor
School of Information Studies
McGill University
Outline

• Visualization definitions
• Current Library Catalogues
• Visualization in Libraries
  • Management
  • Physical world metaphors
  • Topical Browsing
Research Interests

• Human-information interaction
  • Information Interface design and testing

• Library data is incredibly rich
  • Current online interfaces exploit a portion of it
    • Facets are a clear improvement

• How do collections populate topical organization structures?
  • LCSH, MeSH, PIM folder trees, etc.
  • Much of the structures are practically empty

• Developing new ways to interact with information using large topical structures and searching
  • Simplification or filtering
  • Integration of topical browsing and searching
Visualization Definitions

• Data Visualization
  • Broad term for visual representations of any data

• Scientific Visualization
  • always about physical objects
  • using data with an inherent spatial component
  • E.g., wind tunnel vector data or three-dimensional (3D) medical images

• Information Visualization
  • Typically involve abstract, non-spatial data
  • Usually do not have inherent geometries by which to map information
    • E.g., financial data or document collections
  • The visual metaphor design is an arbitrary choice which should be tested
    • E.g., Windows Desktop
Scientific Visualization
Visualization Sub-types NOT covered today

• Geographic Information Visualization (GIS)
  • Requires longitude/latitude
  • If not part of metadata then
    • produced through a conversion process
    • E.g., Cities’ geographical center and boundary coordinates are available online

• Numerical data analysis and presentation
  • What can mostly be done using Excel or equivalent tools
  • Traditional ‘chart’: bar, line, scatter plot, etc.

• Software visualization
  • Software Engineering
Benefits of IV

• Information visualization
  • helps users focus on information that matters the most to them.
  • helping users to see patterns,
  • make connections, and
  • draw conclusions from the data.

• Makes the data accessible to all users
  • not just those who possess advanced analytic skills.
Visualization Process (Ben Fry)

- Acquire data
- Parse
- Filter
- Mine
- Represent Visually
- Refine
- Interact
- (Refine)

≈ 80% of data mining process time
Data parsing/filtering

  - Over 90% of the information available is in unstructured and semi-structured forms
    - which is very difficult to search.
  - Text mining is a tool to help Information professionals
    - to find the relevant information and deliver to its users
  - Superficially, bibliographic records appear to be a homogenous dataset
  - Examination reveals that records change through time
  - There are methods to disaggregate such data
    - Analogous to those used in some scientific disciplines
Library Catalogues

Discovery
U of Washington

2008

2017
Latest
Generation
Catalogues

http://wakeipac.co.wake.nc.us/

Facets++
Cover art
Relevance
Spell check/closest match
Discover Layers
Cover Browsing (IV?)

Current Discovery Layer Interfaces

• Discovery Layers are now essentially equivalent across major vendors
• All have adopted faceted browsing/filtering, “relevance” ranking, etc.
  • Topical descriptors (e.g., LCSH or MeSH) are treated as independent
  • i.e., Relationships between topics are ignored.
• Essentially text based beyond cover art
  • Visualizations?
we believe that the OPAC is dead.

In the world we live in today, you should not encourage your users to start their search in a local library catalogue.

• our idea of ‘collection’ has expanded from the library and catalogue-bound holdings
  • to a concentric collection model
    • which situates discovery AND access within the entire biblio-blogosphere
    • which are merging through mass digitisation and open discovery.

• We will move away from an institutional catalogue and set of subscribed databases to
  • "managing our imprint on shared global discovery systems“ (Mark Dahl)

• Libraries manage/offer access to collections
  • Text/data mining and visualization aim to facilitate access to information
Visualization for Libraries
Why Information Visualization for Libraries?


• Reduce costs of eResources renewals
  • Understand what users are truly using
  • Eliminates (reduces) the need to manually access and download journal usage reports from the sites of individual publishers.

• Communicate data backed decisions

• Educational resources

• IV is going to happen (happening) to the general public
  • information professionals must be IV literate to serve the public

• Providing raw data and nothing else is just not enough
IV for Library Management
HARVARD LIBRARY EXPLORER

Browse the library collection by clicking this bar

Bar width denotes # of books in category
color denotes subject
opacity denotes "popularity" (circulation events/total books)

SEARCH

A-Z All the Books
0-1000

Key
- A - General Works
- B - Philosophy, Psychology, Religion
- C - Auxiliary Sciences of History
- D - History: General & Outside the Americas
- E - History: United States
- F - History: United States Local & America
- G - Geography, Anthropology, Recreation
- H - Social Sciences
- J - Political Science
- K - Law
- L - Education
- M - Music
- N - Fine Arts
- P - Language and Literature
- Q - Science
- R - Medicine
- S - Agriculture
- T - Technology
- U - Military Science
- V - Naval Science
- Z - Library Science & Information Resources

The graph above shows a running total of books published in the given subject by year, from 2000-2010.

Total number of circulation events (2002-2011): 232,0131
Total number of books held: 12,384,023
Ratio of books to circulation events .23

HOLDINGS
CIRCULATION

Circulation events include any time a book is checked out, put on course reserve, held, or lost.
FIGURE 7
Data Bubbles Representing Student Enrollment by Discipline, X-Axis Showing Expenditures & Y-Axis is Book Circulation Since Purchase for School of Humanities & Social Sciences and School of Sciences & Mathematics, 2013–2014

Topical Browsing and Exploration

Figure 1. Visualization of the Results for a Search for “Climate Change.”
Topical Browsing

Topical Browsing and Collection Searching

Physical World Metaphors

New imitates old
Tick Tock

James Patterson, Michael Ledwidge

Synopsis

NYC's #1 detective, Michael Bennett, has a huge problem—the Son of Sam, the Werewolf of Wisteria and the Mad Bomber are all back. The city has never been more terrified!

Tick—A killer's countdown begins

A rash of horrifying crimes tears through the city, throwing it into complete chaos and terrorizing everyone living there. Immediately, it becomes clear that they are not the work of an insane, but of a calculating, efficient, and deadly mastermind.

Tick—Michael Bennett is on the chase

The city calls on Detective Michael Bennett, pulling him away from a seaside retreat with his ten adopted children, his grandfather, and their beloved nanny, Mary Catherine. Not only does it tear apart their vacation, it leaves the entire family open to attack.
Reproducing the physical bookshelf


**Figure 6:** Selection of sub-disciplines in the faculty of computer science.

**Figure 7:** 3D shelf view with books and an expanded sorting option (left).

Fig. 13. The Book Pile visualization.

• a small number of the most popular search queries
  • accounts for a disproportionate amount of the overall queries.

• Why not produce visualizations that address some of these queries?
  • Bolster default result sets for some queries
IV Barriers in libraries (Chen 2017)

- Affordability of new technologies
- Lack of defined roles for staff
  - Lack of expertise and skills in information visualization
- Process of information visualization is both
  - technically complex and
  - highly challenging,
  - it requires high technical proficiency.
- Requires range of skills
  - math and statistics and
  - knowledge of data storage,
  - mining methods,
  - front-end design, and
  - development.
- Access and standardization of data
- Culture of educational institutions
Conclusion

• Visualizations can appealing, exciting, and memorable.
• The utility of visualizations depends on the specific user and
  • the specific task being carried out
  • Difficult to ‘guess’ which visualization might be useful a priori
• Data analysis/Visualization is often an ad-hoc service
• Creating visualizations is a technical and artistic craft that must be practiced.
Other sources

- Lynema, E., C Lown, D Woodbury. Virtual browse: designing user-oriented services for discovery of related resources. - Library Trends