Users and Tasks

Michel Beaudouin-Lafon suggested to
- “explain the kind of things one can discover/understand with information visualization” and
- “what it takes to generate such visualizations (in terms of quality of the metadata, for example).”

Tasks that might benefit from visualizations:
- New tools to access the DL, which could include visualization tools, e.g. in conjunction with the author pages, the co-authorship lists, etc.
- Supporting social navigation based on download statistics.
- Finding a new editor-in-chief for a journal.
- Evaluation of journal proposals (whether it's a timely proposal, whether there really is a field behind it, etc.).
- Proactive encouragement of new publications in a given area.
Overview

1. Visual Interfaces to Digital Libraries
2. Knowledge Domain Analysis and Visualizations
3. Cyberinfrastructure for InfoVis/KDVis Research
4. Managing Humanity’s Knowledge and Expertise


1) Visual Interfaces to Digital Libraries

Facing the Information Flood:
- Information available in electronic form doubles every 18 months.
- Human perception stays constant.
- Almost no development in online interfaces. Can’t pack more text.

Let's see how much our means of accessing information have changed using http://www.archive.org/.

8 years back in time

Yahoo Oct 17, 1996

Yahoo Oct 19, 2004

5 years back in time

Amazon Sept 02, 1999

Amazon Oct 19, 2004
Facing the Information Flood:

- Information available in electronic form doubles every 18 months.
- Human perception stays constant.

Opportunity & Challenge:

Shift user’s mental load from slow reading to faster perceptual processes such as visual pattern recognition.

Facilitated by:

- CPU speed & hard disk sizes have increased by two orders of magnitude.
- Bandwidth: Since the invention of the web browser, international IP bandwidth deployments have more than doubled each year.
- Monitor resolution has increased by a factor of 4 (800x600 -> 1600x1200).

2) Knowledge Domain Analysis and Visualization

To answer questions such as:

- What are the major research areas, experts, institutions, regions, nations, grants, publications, journals in xx research?
- Which areas are most insular?
- What are the main connections for each area?
- What is the relative speed of areas?
- Which areas are the most dynamic/static?
- What new research areas are evolving?
- Impact of xx research on other fields?
- How does funding influence the number and quality of publications?

Answers are needed by funding agencies, companies, and researchers.
User Groups

- **Students** can gain an overview of a particular knowledge domain, identify major research areas, experts, institutions, grants, publications, patents, citations, and journals as well as their interconnections, or see the influence of certain theories.
- **Researchers** can monitor and access research results, relevant funding opportunities, potential collaborators inside and outside the fields of inquiry, the dynamics (speed of growth, diversification) of scientific fields, and complementary capabilities.
- **Grant agencies/R&D managers** could use the maps to select reviewers or expert panels, to augment peer-review, to monitor (long-term) money flow and research developments, evaluate funding strategies for different programs, decisions on project durations, and funding patterns, but also to identify the impact of strategic and applied research funding programs.
- **Industry** can use the maps to access scientific results and knowledge carriers, to detect research frontiers, etc. Information on needed technologies could be incorporated into the maps, facilitating industry pulls for specific directions of research.
- **Data providers** benefit as the maps provide unique visual interfaces to digital libraries.
- Last but not least, the availability of dynamically evolving maps of science (as ubiquitous as daily weather forecast maps) would dramatically improve the communication of scientific results to the general public.

Process of Mapping Knowledge Domains

<table>
<thead>
<tr>
<th>DATA EXTRATION</th>
<th>UNIT OF ANALYSIS</th>
<th>MEASURES</th>
<th>LAYOUT (either one code does both similarity and coordinate maps)</th>
<th>DISPLAY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SEARCHES</strong></td>
<td><strong>COMMON</strong></td>
<td><strong>COUNT FREQUENCIES</strong></td>
<td><strong>SCALAR (just by unit matrix)</strong></td>
<td><strong>SIMILARITY</strong></td>
</tr>
<tr>
<td>IR, INSPEC</td>
<td>OCSIS</td>
<td>Articles by keyword</td>
<td>Direct citation</td>
<td>Dimensionality Reduction</td>
</tr>
<tr>
<td>BioMedScan</td>
<td>Document</td>
<td>Authors cited</td>
<td>Co-citation</td>
<td>By proximity of eigenvalues</td>
</tr>
<tr>
<td>Medical</td>
<td>Author</td>
<td>Co-citation</td>
<td>Combined linkage</td>
<td>Factor Analysis (FA) and</td>
</tr>
<tr>
<td>Research index</td>
<td>Term</td>
<td>co-citation</td>
<td>Co-clustering</td>
<td>Principal Components Analysis (PCA)</td>
</tr>
<tr>
<td>Patents, etc.</td>
<td><strong>BROADENING</strong></td>
<td><strong>THRESHOLDS</strong></td>
<td><strong>SCALAR</strong></td>
<td>Multi-dimensional scaling (MDS)</td>
</tr>
<tr>
<td>By citation</td>
<td><strong>INTERACTION</strong></td>
<td><strong>VECTOR</strong> (just by attribute matrix)</td>
<td><strong>VECTOR</strong></td>
<td>LSA, Topic</td>
</tr>
<tr>
<td>By term</td>
<td><strong>ANALYSIS</strong></td>
<td><strong>VECTORS</strong></td>
<td><strong>VECTORS</strong></td>
<td>PubMed wrappers (PMWR)</td>
</tr>
<tr>
<td><strong>CORRELATION</strong> (if desired)</td>
<td><strong>RELATION</strong></td>
<td><strong>RELATION</strong></td>
<td><strong>RELATION</strong></td>
<td>Self-organizing maps (SOM)</td>
</tr>
<tr>
<td>Pearson's R or any of above</td>
<td><strong>TRANSFORMATION</strong></td>
<td><strong>TRANSFORMATION</strong></td>
<td><strong>TRANSFORMATION</strong></td>
<td>include SOM ET-maps, etc.</td>
</tr>
<tr>
<td><strong>SCALAR</strong></td>
<td></td>
<td><strong>TRANSFORMATION</strong></td>
<td><strong>TRANSFORMATION</strong></td>
<td>Analyze</td>
</tr>
<tr>
<td><strong>TRANSFORMATION</strong></td>
<td></td>
<td><strong>TRANSFORMATION</strong></td>
<td><strong>TRANSFORMATION</strong></td>
<td>Analyze</td>
</tr>
</tbody>
</table>

Indicator-Assisted Evaluation and Funding of Research

Visualizing the influence of grants on the number and citation counts of research papers (Boyack & Börner, 2003)

Mapping Topic Bursts

(Mane & Börner, 2004)

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001. 
Mapping Medline Papers, Genes, and Proteins Related to Melanoma Research

(Boyack, Mane & Börner, 2004)

Mapping the Evolution of Co-Authorship Networks
Won 1st price at the IEEE InfoVis Contest
(Ke, Visvanath & Börner, 2004)
“As Figure 1 shows, there is some danger of *Cognitive Science* becoming too dominated by psychology. In the journal’s recent past, we have had strong representation from many mainstays of cognitive science including learning, neuroscience, problem solving, language, reasoning, computational modeling, and representation. However, the presence of philosophy, anthropology, artificial intelligence, and machine learning seems sparser than is warranted by their historical influence on cognitive science. Monitoring the diversity of the journal and field is critical if we wish to cultivate future developments of general principles that govern intelligent systems in all of their guises.”

3) Cyberinfrastructure for InfoVis and KDVis Research

**IVC DB Data Sets** ([http://iv.slis.indiana.edu/db](http://iv.slis.indiana.edu/db))

### Papers and Patents

<table>
<thead>
<tr>
<th>Database</th>
<th>Number of Entries</th>
<th>Years Covered</th>
<th>Size (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NLM</strong></td>
<td>11,491,477</td>
<td>1983-2001</td>
<td>135 MB</td>
</tr>
<tr>
<td><strong>PNAS</strong></td>
<td>16,160</td>
<td>1997-2012</td>
<td>53 MB</td>
</tr>
</tbody>
</table>

### Grant Awards

<table>
<thead>
<tr>
<th>Grant Agency</th>
<th>Number of Entries</th>
<th>Years Covered</th>
<th>Size (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>National Science Foundation (NSF)</strong></td>
<td>101,722</td>
<td>1995-2002</td>
<td>40 MB</td>
</tr>
<tr>
<td><strong>National Institute of Health (NIH)</strong></td>
<td>1,503,622</td>
<td>1973-1980 and 1994-2012</td>
<td>2.2 GB</td>
</tr>
</tbody>
</table>

### Funding Opportunities

<table>
<thead>
<tr>
<th>Community of Science (COS)</th>
<th>Number of Entries</th>
<th>Years Covered</th>
<th>Size (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>36,154 (2,000 new entries per month)</td>
<td>2001-present</td>
<td>60 MB</td>
</tr>
</tbody>
</table>

4) How to Manage Humanity’s Knowledge and Expertise

Given the steadily increasing flood of information, how can we keep track and make use of what we collectively know?

- Shift user’s mental load from slow reading to faster perceptual processes such as visual pattern recognition.
- Aim for reusability of data and methods/approaches/algorithms and reproducibility of results. \(\Rightarrow\) Interrelate data, code, results, authors.
- Use usage log data to support social navigation and to create novel reputation systems. \(\Rightarrow\) … & usage data. Basically, a new infrastructure to keep track of knowledge.
- Give people global knowledge of the structure and evolution of scientific knowledge. \(\Rightarrow\) Global maps of science
- Provide access to knowledge and expertise. \(\Rightarrow\) … & expertise

Interrelate Data, Code, Papers, Authors & Usage Data

Data-code-computing cyberinfrastructures that interrelate data, code, results, authors, and usage data

- Enable data/algorithm/result comparison at data/code/data level.
- Facilitate new types of searches, e.g., retrieve all users that worked with data set x, retrieve all papers that used algorithm y.
- Support algorithm comparison and re-use, e.g., the re-application of an algorithm sequence reported in a paper to a different data set.
- Do provide bridges between algorithm developers and users.
- Could provide a great testbed application for novel ways to store, preserve, integrate, correlate, access, analyze, map or interact with data.
- Are of interest to diverse communities.

Given the steadily increasing flood of information, how can we keep track and make use of what we collectively know?

- Shift user’s mental load from slow reading to faster perceptual processes such as visual pattern recognition.

- Aim for reusability of data and methods/approaches/algorithms and reproducibility of results. \( \rightarrow \) Interrelate data, code, results, authors.
- Use usage log data to support social navigation and to create novel reputation systems. \( \rightarrow \) … & usage data. **Basically, a new infrastructure to keep track of knowledge.**

- Give people global knowledge of the structure and evolution of scientific knowledge. \( \rightarrow \) Global maps of science
- Provide access to knowledge and expertise. \( \rightarrow \) … & expertise
Acknowledgements & References

Support comes from the School of Library and Information Science, Indiana University's High Performance Network Applications Program, a Pervasive Technology Lab Fellowship, an Academic Equipment Grant by SUN Microsystems, NIA, and an SBC (formerly Ameritech) Fellow Grant. This material is based upon work supported by the National Science Foundation under Grant No. DUE-0333623 and IIS-0238261.