Mapping Scholarly Communication Patterns

Dr. Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

Talk in Dept. of Telecommunications, IUB
Dec 06, 2007

NIH Funding Patterns

Advantages for Funding Agencies
- Supports monitoring of (long-term) money flow and research developments, evaluation of funding strategies for different programs, decisions on project durations, funding patterns.
- Staff resources can be used for scientific program development, to identify areas for future development, and the stimulation of new research areas.

Advantages for Researchers
- Easy access to research results, relevant funding programs and their success rates, potential collaborators, competitors, related projects/publications (research push).
- More time for research and teaching.

Advantages for Industry
- Fast and easy access to major results, experts, etc.
- Can influence the direction of research by entering information on needed technologies (industry-pull).

Advantages for Publishers
- Unique interface to their data.
- Publicly funded development of databases and their interlinkage.

For Society
- Dramatically improved access to scientific knowledge and expertise.

Opportunities for Mapping Science
Science map applications: Identifying core competency
Kevin W. Boyack & Richard Klavans, unpublished work.

Funding patterns of the US Department of Energy (DOE)

Science map applications: Identifying core competency
Kevin W. Boyack & Richard Klavans, unpublished work.

Funding Patterns of the National Science Foundation (NSF)
Science map applications: Identifying core competency
Kevin W. Boyack & Richard Klavans, unpublished work.

Funding Patterns of the National Institutes of Health (NIH)

Places & Spaces: Mapping Science exhibit, see also http://scimaps.org.
"Places & Spaces: Mapping Science"
on display at the American Museum of Science and Energy, Oak Ridge, TN,
Nanotechnology

This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes place in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the biological and medical sciences, at the lower left.
Science Puzzle Map for Kids by Fileve Palmer, Julie Smith, Elisha Hardy and Katy Börner, Indiana University, 2006.
(Base map taken from Illuminated Diagram display by Kevin Boyack, Richard Klavans, and W. Bradford Paley.)
Scholarly Knowledge Management Tools

Tools we developed for ourselves and our clients

Information Visualization Laboratory  Taxonomy Visualization/
Management System  Validation System
Information Visualization Laboratory Management System

https://ivl.slis.indiana.edu
Challenges - Data Collection & Integration

Figure 1: The interoperability and cross-linkage problem. Many but not all of today’s scholarly datasets, e.g., papers, patents, grants, are stored and made available so that ‘vertical’ citation linkages can be traversed. There are very few instances in which datasets of different origin and/or type are ‘horizontally’ interlinked.

Scholarly Database: Web Interface

Search across publications, patents, grants.
Download records and/or (evolving) co-author, paper-citation networks.

Register for free access at https://sdb.slis.indiana.edu.
Datasets available via the Scholarly Database (* future feature)

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Records</th>
<th>Years Covered</th>
<th>Updated</th>
<th>Restricted Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline</td>
<td>13,149,741</td>
<td>1965-2005</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>PhysRev</td>
<td>398,005</td>
<td>1893-2006</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>PNAS</td>
<td>16,167</td>
<td>1997-2002</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>USPTO</td>
<td>3,179,930</td>
<td>1976-2004</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>NSF</td>
<td>174,835</td>
<td>1985-2003</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>NIH</td>
<td>1,043,804</td>
<td>1972-2002</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>18,021,560</strong></td>
<td><strong>1893-2006</strong></td>
<td><strong>4</strong></td>
<td><strong>3</strong></td>
</tr>
</tbody>
</table>

Aim for comprehensive time, geospatial, and topic coverage.
SEI: Network Workbench: A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research. NSF IIS-0513650 award (Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert (Senior Personnel), $1,120,926) Sept. 05 - Aug. 08. http://nwb.slis.indiana.edu

CAREER: Visualizing Knowledge Domains. NSF IIS-0238261 award (Katy Börner, $451,000) Sept. 03 - Aug. 08. http://iv.slis.indiana.edu

US Patents
Requirements:
- Requires the design & implementation of ‘software glue’ that can interlink datasets and algorithms written in different languages using different data formats.
- The smaller the glue or ‘CI Shell’, the more likely it can be maintained.
- Dataset and algorithm ‘plugins’ are provided by application holders/community.
- Applications resemble custom ‘fillings’.

Cyberinfrastructure Shell (CIShell)

CIShell is an ‘empty shell’ that supports:
- Easy integration of new datasets and algorithms by algorithm developers and
- Easy usage of algorithms by algorithm users.

Its plug-and-play architecture supports the integration and utilization of diverse:
- Datasets, e.g., stored in files, databases, streaming data.
- Algorithms, e.g., data processing, analysis, modeling, visualization.
- Interfaces, e.g., remote services, scripting engines, peer-to-peer clients.
- Services, e.g., workflow support, scheduler.
Hence, it can be used for custom UI/Toolkit development.
CIShell – Needs of Algorithm Developers & Users

Developers

Users

CIShell Wizards  
CIShell  
IVC Interface

NFB Interface

CIShell – Deployment

Data-Algorithm Repositories

Peer-to-Peer

Server-Client

Stand Alone

CIShell applications can be deployed as distributed data and algorithm repositories, stand alone applications, peer-to-peer architectures, and server-client architectures.
References

- Börner, Katy, Mapping All of Science: How to Collect, Organize and Make Sense of Mankind’s Scholarly Knowledge and Expertise. Accepted for Environment and Planning B, Special Issue on Mapping Humanity’s Knowledge and Expertise in the Digital Domain.
- Holloway, Todd, Božičević, Miran and Börner, Katy. Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors. Accepted for Complexity. Also available as cs.IR/0512085.