Overview

1. **Data mining and visualization research** that aims to increase our scientific understanding of the structure and dynamics of science and technology.

2. **Novel approaches and services** that improve information access, researcher networking, and research management.

3. **Data services and plug-and-play macroscope tools** that commoditize data mining and visualization.
Find your way

Find collaborators, friends

Identify trends

Terra bytes of data

Descriptive & Predictive Models

Plug-and-Play Macrosopes
## Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th></th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, all of science.</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts in 20-years of PNAS</td>
<td>113 Years of Physics Research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping a states intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td>Base knowledge from which one grant draws.</td>
<td>Knowledge flows in Chemistry research</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network</td>
<td>NIH’s core competency</td>
</tr>
</tbody>
</table>
Mapping Indiana’s Intellectual Space

Identify
- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia

---

Mapping the Evolution of Co-Authorship Networks

Mapping the Evolution of Co-Authorship Networks

Legend
Nodes = Authors
Node size = Number of papers published
Node color = Number of citations
Edges = Co-authorship relations
Edge color = Year of first co-authorship

Mapping Transdisciplinary Tobacco Use Research Centers Publications

Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.
Zoss & Börner, forthcoming.
Supported by NIH/NCI Contract HHSN261200800812
Mapping Topic Bursts

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.


References


http://www.pnas.org/content/vol101/suppl_1/


http://scimaps.org/atlas

Places & Spaces: Mapping Science Exhibit (http://scimaps.org)

After eight years, there now exist 80 out of 100 maps.

Mapping Science Exhibit at MEDIA X, Wallenberg Hall, Stanford University, 2009

Eric Fischer. 2012. Language Communities of Twitter.
Facebook Draws a Map of the Connected World. Dec 14, 2010

Science Maps in “Expedition Zukunft” science train visiting 62 cities in 7 months 12 coaches, 300 m long Opening was on April 23rd, 2009 by German Chancellor Merkel

http://www.expedition-zukunft.de
Illuminated Diagram Display
on display at the Smithsonian in DC.
http://scimaps.org/exhibit_info/#ID
About
This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what’s important by using projectors or screens as smart spotlights, promoting the research impact of individuals, giving a “grand tour” of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.

Elinor Ostrom - Nobel Prize in Economic Sciences 2009
Born: 7 August 1933, New York, NY, USA
Affiliation at the time of the award: Indiana University, Bloomington, IN, USA; Arizona State University, Tempe, AZ, USA
Prize motivation: "for her analysis of economic governance, especially the commons"
Field: Economic governance
Contribution: Challenged the conventional wisdom by demonstrating how local property can be successfully managed by local commons without any regulation by central authorities or privatization.

Interact
Select any location on the Geographic Map location (by breathing your finger over an area on the lectern's touch screen) and topics studied in that area will highlight on the Science Map: the brighter a topic glow, the more papers on that topic originated in the selected area. Conversely, Toucheing a scientific area on the Science Map Illustrates places on the Geographic Map where that topic is studied. People and topic buttons support the exploration of publication output by selected Noble laureates and particular lines of research using MEDLINE data from 2000-2009.
Overview

1. Data + data mining and visualization research that aims to increase our scientific understanding of the structure and dynamics of science and technology.

2. Novel approaches and services that improve information access, researcher networking, and research management.

3. Data services and plug-and-play macroscope tools that commoditize data mining and visualization.

Scholarly Database at IU
Scholarly Database at Indiana University
http://sdb.wiki.cns.iu.edu

Supports federated search of 25 million publication, patent, grant records.
Results can be downloaded as data dump and (evolving) co-author, paper-citation networks.

Register for free access at http://sdb.cns.iu.edu
Since March 2009:
Users can download networks:
- Co-author
- Co-investigator
- Co-inventor
- Patent citation
and tables for
burst analysis in NWB.

VIVO International Researcher
Network
VIVO: A Semantic Approach to Creating a National Network of Researchers (http://vivoweb.org)

- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.
- Simplify reporting tasks, e.g., generate biosketch, department report.

Funded by $12 million NIH award.

**Temporal Analysis (When)** Temporal visualizations of the number of papers/funding award at the institution, school, department, and people level

**Topical Analysis (What)** Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)
Network Analysis (With Whom?) Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects?

Geospatial Analysis (Where) Where is what science performed by whom? Science is global and needs to be studied globally.
VIVO On-The-Go

Overview, Interactivity, Details on Demand come to commonly used devices and environments

Develop VIVO Visualizations
See also Visualization in VIVO Workshop on Aug 24, 2011
http://wiki.cns.iu.edu/display/PRES/VIVO+Presentation

VIVO Presentation
A Added by Chin Hwee Kang, last edited by Christian Tan on Aug 24, 2011 (view change)

August, 2011 Workshop

Material

- Java 1.5 or higher - A programming language and computing platform for developing cross OS softwares.
- Science of Science tool (SoS2) - An desktop application for information analysis and visualization.
- GeOLA - An interactive visualization tool for networks and contains systems, dynamic and hierarchical orcahs.
- VIVO August 2011 workshop data.zip - Hands on workshop data package

Slides

- Tutorial Slides presented at the VIVO Conference 2011
- Pre-Questionnaire and Post-Questionnaire

Demo Links

- Map of Science Visualization (dev link)
- Temporal Graph Visualization (dev link)
- National Researcher Networking Visualization
- Word Cloud Visualization (dev link)
Mapping Sustainability Research
The geographic map at state level.
The geographic map at city level.

Search result for “corn”
Icons have same size but represent different #records
Click on one icon to display all records of one type. Here publications in the state of Florida.

Detailed information on demand via original source site for exploration and study.
The science map at 13 top-level scientific disciplines level.

The science map at 554 sub-disciplines level.
Thermal tolerant avicelase from Acidothermus cellulolyticus

Abstract

The invention provides a thermal tolerant (thermostable) cellulase, AvrIII, that is a member of the glycoside hydrolase (GH) family. AvrIII was isolated and characterized from Acidothermus cellulolyticus and, like many cellulases, the disclosed polypeptide and/or its derivatives may be useful for the conversion of biomass into biofuels and chemicals.
Overview

1. Data + data mining and visualization research that aims to increase our scientific understanding of the structure and dynamics of science and technology.

2. Novel approaches and services that improve information access, researcher networking, and research management.

3. Data services and plug-and-play macroscope tools that commoditize data mining and visualization.


Video and paper are at http://www.scivee.tv/node/27704
**Needs-Driven Workflow Design** using a modular data acquisition/analysis/modeling/visualization pipeline as well as modular visualization layers.

---

**OSGi & CIShell**

- CIShell ([http://cishell.org](http://cishell.org)) is an open source software specification for the integration and utilization of datasets, algorithms, and tools.
- It extends the Open Services Gateway Initiative (OSGi) ([http://osgi.org](http://osgi.org)), a standardized, component oriented, computing environment for networked services widely used in industry since more than 10 years.
- Specifically, CIShell provides “sockets” into which existing and new datasets, algorithms, and tools can be plugged using a wizard-driven process.
About the Cyberinfrastructure Shell
The Cyberinfrastructure Shell (CIShell) is an open-source, community-driven platform for the integration and utilization of datasets, algorithms, tools, and computing resources. Algorithms integration support is built in for Java and most other programming languages. Being Java-based, it will run on almost all platforms. The software and specification is released under an Apache 2.0 License.

CIShell is the basis of Network Workbench, TexTrend, Sci2, and the upcoming Epic tool.

CIShell supports remote execution of algorithms. A standard web service definition is in development that will allow pools of algorithms to transparently be used in a peer-to-peer, client-server, or web front-end fashion.

CIShell Features
A framework for easy integration of new and existing algorithms written in any programming language
Using CIShell, an algorithm writer can fully concentrate on creating their own algorithm in whatever language they are comfortable with. Simple tools are provided to then take their algorithm and
The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.

In February 2009, the tool provides more 169 plugins that support the preprocessing, analysis, modeling, and visualization of networks.

More than 50 of these plugins can be applied or were specifically designed for S&T studies.

It has been downloaded more than 110,000 times since December 2006.


Computational Proteomics

What relationships exist between protein targets of all drugs and all disease-gene products in the human protein–protein interaction network?


Figure 2 Drug-target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Cylcles and rectangles correspond to drug and target proteins, respectively. A link is placed between a drug node and a target node if the protein is a known target of that drug. The size of the drug (protein) node is proportional to the number of targets that the drug has (the number of drugs targeting the protein). Color codes are given in the legend. Drug codes (circles) are colored according to their Anatomical Therapeutic Chemical Classification, and the target proteins (rectangular boxes) are colored according to their cellular component obtained from the Gene Ontology database.
Computational Economics

Does the type of product that a country exports matter for subsequent economic performance?


--

Computational Social Science

Studying large scale social networks such as Wikipedia


http://gigapan.com/gigapans/4277
Computational Epidemics
Forecasting (and preventing the effects of) the next pandemic.


Sci2 Tool – “Open Code for S&T Assessment”
OSGi/CIShell powered tool with NWB plugins and many new scientometrics and visualizations plugins.

The Sci² Tool is used by the

- National Science Foundation,
- National Institutes of Health,
- US Department of Agriculture, and
- National Oceanic and Atmospheric Administration

Tool registrations come from 73 countries and professions such as

- professor
- researcher
- assistant
- student
- program
- manager
- librarian
- fellow
- scientist
- director
- policy
- senior
- phd
- analyst
A number of other projects recently adopted OSGi and/or CIShell:

- **Cytoscape** ([http://cytoscape.org](http://cytoscape.org)) Led by Trey Ideker at the University of California, San Diego is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).

- **MAEviz** ([https://wiki.ncsa.uiuc.edu/display/MAE/Home](https://wiki.ncsa.uiuc.edu/display/MAE/Home)) Managed by Jong Lee at NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.

- **Taverna Workbench** ([http://taverna.org.uk](http://taverna.org.uk)) Developed by the myGrid team led by Carol Goble at the University of Manchester, U.K. is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.

- **TEXTrend** ([http://textrend.org](http://textrend.org)) Led by George Kampis at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.

- **DynaNets** ([http://www.dynanets.org](http://www.dynanets.org)) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.


As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.
All papers, maps, tools, talks, press are linked from [http://cns.iu.edu](http://cns.iu.edu)

CNS Facebook: [http://www.facebook.com/cnscenter](http://www.facebook.com/cnscenter)
Mapping Science Exhibit Facebook: [http://www.facebook.com/mappingscience](http://www.facebook.com/mappingscience)