

InfoVis Cyberinfrastructure



Katy Börner

School of Library and Information Science

INDIANA UNIVERSITY
BLOOMINGTON

katy@indiana.edu

SLIS Colloquium, November 19th, 2004

Information Visualization CyberInfrastructure

The InfoVis Cyberinfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse data sets.

DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated.
<http://iv.slis.indiana.edu/db>

COMPUTING RESOURCES

The InfoVis Cyberinfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1200 servers with 12 900MI tz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 4GB memory serves as the web front end for the database servers.
<http://iv.slis.indiana.edu/cr>

SOFTWARE

An open source IV² framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feels for the interface and even entire toolkits can be easily "plugged in" or "unplugged".
<http://iv.slis.indiana.edu/sw>

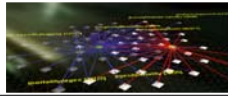
LEARNING MODULES

A set of associated learning modules aims to equip learners with a practical skill set by providing code and advice to quickly modify and run different algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations.
<http://iv.slis.indiana.edu/lm>

InfoVis Lab, School of Library and Information Science, Indiana University (2004).
For more information, contact Katy Börner at katy@indiana.edu.

This material is based upon work supported by the National Science Foundation under Grant No. IRI-0229261 and IRI-0313623.

Photo: Andrew G. Gardner/Corbis/Outlines



Motivation

IVC Database

- Provide access to major scholarly databases.

IVC Software Framework

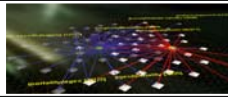
- Support developers and programmers in the comparison and distribution of new algorithms.
- Interconnect algorithm developers and users. What algorithms do users need/want?

IVC Learning Modules

- Support (non-programmer) users in the utilization of advanced InfoVis algorithms.
- Provide a unique resource for InfoVis education.

Support InfoVis & 'Knowledge Domain Visualization' research.

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



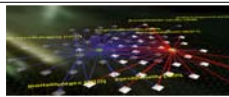
Publications about the Infrastructure

- Börner, Katy and Zhou, Yuezheng. (2001) A Software Repository for Education and Research in Information Visualization. Information Visualisation Conference, London, England, July 25-27, pp. 257-262.
- Baumgartner, Jason and Börner, Katy (2002). Towards an XML Toolkit for a Software Repository Supporting Information Visualization Education. IEEE Information Visualization Conference, Boston, MA, 2002. Interactive Poster.
- Baumgartner, Jason, Börner, Katy, Deckard, Nathan J., Sheth, Nihar. (2003). An XML Toolkit for an Information Visualization Software Repository. Poster Compendium, IEEE Information Visualization Conference, pp. 72-73.
- Penumarthy, Shashikant, Börner, Katy and Herr, Bruce. Information Visualization Cyberinfrastructure Software Framework. Submitted to [Information Visualization](#).

Moral:

Do not do infrastructure development if you need/want scholarly publications.

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



Grants

- Center of Excellence for Computational Diagnostics. 21st Century Grant (Susanne Ragg, David Clemmer, Sven Rahmann, and Ilka Ott, Terry Vik, R Clement McDonald, Nunroe Pecoock, Zina Ben Miled & Katy Börner, \$1,994,951) Sept. 04 - Aug. 06.
- Sun Center of Excellence in Knowledge Management and Discovery, SUN Microsystems (Stephanie Burks, Katy Börner, Zina Ben-Miled), March 2004.
- Outstanding Junior Faculty Award. (Principal Investigator, \$14,000), 2004.
- Data-Code-Computing Infrastructure for Data Mining, Modeling, and Visualization Research and Education. [Pervasive Technology Labs Fellowship](#) (Principal Investigator, \$48,750) Sept. 2003 - Aug. 2004.
- CAREER: Visualizing Knowledge Domains. NSF IIS-0238261 award (Principal Investigator, \$400,000) Sept. 2003-Aug. 2008.
- Information Visualization Learning Modules. SBC (formerly Ameritech) Fellow Grant (Principal Investigator, \$15,000) May 2003-June 2004.
- Development of a Spatial-Experimental Laboratory for Research and Policy Analysis Related to Complex Systems. NSF/BCS-0215738 Major Research Instrumentation Grant. This project benefits multiple departments at IU. (PIs are Elinor Ostrom, Jerome Busemeyer, Tom Evans, Robert Huckfeldt & James Walker \$847,874) Aug. 2002-July 2007.

Moral:

A good infrastructure (development) attracts grant funding.

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004

Information Visualization CyberInfrastructure

The InfoVis Cyberinfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse data sets.

DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated.
<http://iv.slis.indiana.edu/db>

COMPUTING RESOURCES

The InfoVis Cyberinfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1300 servers with 12 9000MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 4GB memory serves as the web front-end for the database servers.
<http://iv.slis.indiana.edu/cr>

SOFTWARE

An open source IV² framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feels for the interface and even entire toolkits can be easily "plugged in" or "unplugged".
<http://iv.slis.indiana.edu/sw>

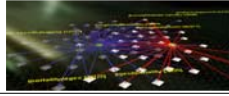
LEARNING MODULES

A set of associated learning modules aims to equip learners with a practical skill set by providing code and advice to quickly modify and run different algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations.
<http://iv.slis.indiana.edu/lm>

InfoVis Lab, School of Library and Information Science, Indiana University (2006).
For more information, contact Katy Börner at kborner@slis.indiana.edu

This material is based upon work supported by the National Science Foundation under Grant No. IRI-0229521 and IRI-0313623.

Photo: Photo by Corbis Outlines 2004



IVC Database – The Team

Design and Implementation

Jay Askren

Saiful Bahari

Andrew Bangert

Christopher Friend

Stephanie Gato

Todd Holloway (Lead)

Ruchi Kapoor

Ketan Mane

Lalitha Visvanath

Qian Wang



Jose Montalvo

Elijah Wright

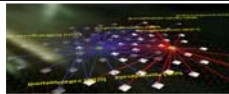
Graphic Design

Caroline Courtney

Project Start

September 2003

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



IVC Database - System Overview

Oracle/Apache/Tomcat/Java—Well understood and reliable tools

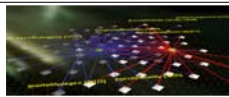
Oracle DB

- Several terabytes of data
- Relational design
- Allow for more collections to be added

Search Engine

- Search on abstract, author, title, journal, date published, and more
- User login for both IU and non-IU users
- User histories
- Administration of data and user accounts
- Compressed downloading of results
- Term-by-document and co-author matrices of results

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



IVC Database - Data Sets

(<http://iv.slis.indiana.edu/db>)

Papers and Patents



Medline
 Number of Entries: 11,693,477
 Years covered: 1963-2002
 Size: 135 MB (gunzipped)



Proceedings of the National Academy of Science (PNAS)
 Number of Entries: 16,169
 Years covered: 1997-2002
 Size: 583 MB



United States Patent and Trademark Office (Patents)
 Number of Entries: 2,582,647
 Years covered: 1976-2003
 Size: 350 MB

Grant Awards



National Science Foundation (NSF)
 Number of Entries: 181,132
 Years covered: 1985-2002
 Size: 400 MB



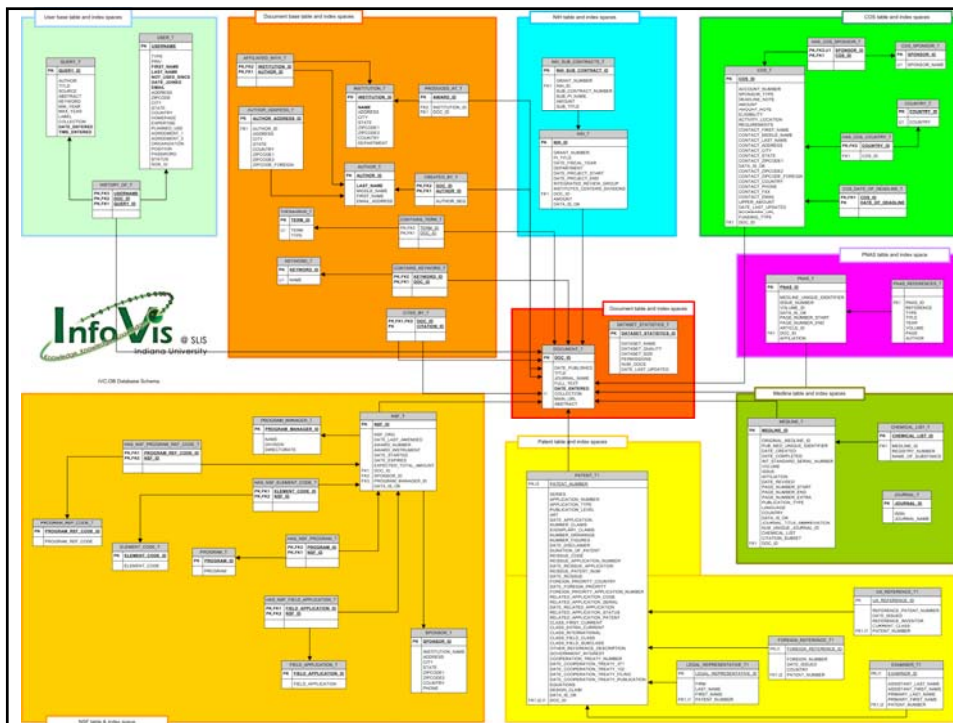
National Institute of Health (NIH)
 Number of Entries: 1,003,521
 Years covered: 1972-1992 and 1994-2002
 Size: 2.3 GB

Funding Opportunities



Community of Science (COS)
 Number of Entries: 38,154 (5,000 new entries per month)
 Years covered: 2001-present
 Size: 60 MB

Katy Börner: InfoVis Cyberinfrastructure.



Information Visualization CyberInfrastructure

The InfoVis CyberInfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse data sets.

DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated.

<http://iv.slis.indiana.edu/db>



COMPUTING RESOURCES

The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1200 servers with 12 900MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 6GB memory serves as the web front-end for the database servers.

<http://iv.slis.indiana.edu/cr>



SOFTWARE

An open source IV framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feels for the interface and entire toolkits can be easily "plugged in" or "unplugged".

<http://iv.slis.indiana.edu/sw>

IV aims to equip learners with a practical skill set by providing code and advice to quickly modify and run different algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations.

<http://iv.slis.indiana.edu/lm>



InfoVis Lab, School of Library and Information Science, Indiana University (2006).
For more information, contact Katy Bowler at kbowler@slis.indiana.edu

This material is based upon work supported by the National Science Foundation under Grant No. IRI-0238261 and DUE-0338623.



Photo: Photo by Gordon Gammie, 2004. <http://www.flickr.com/photos/gammie/>

Information Visualization CyberInfrastructure

The InfoVis CyberInfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse data sets.

DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated.

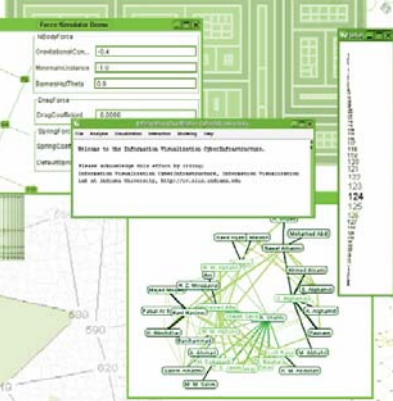
<http://iv.slis.indiana.edu/db>



COMPUTING RESOURCES

The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1200 servers with 12 900MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 6GB memory serves as the web front-end for the database servers.

<http://iv.slis.indiana.edu/cr>



SOFTWARE

An open source IV framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feels for the interface and entire toolkits can be easily "plugged in" or "unplugged".

<http://iv.slis.indiana.edu/sw>

LEARNING MODULES

A set of associated learning modules aims to equip learners with a practical skill set by providing code and advice to quickly modify and run different algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations.

<http://iv.slis.indiana.edu/lm>

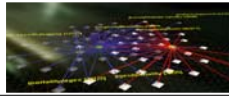


InfoVis Lab, School of Library and Information Science, Indiana University (2006).
For more information, contact Katy Bowler at kbowler@slis.indiana.edu

This material is based upon work supported by the National Science Foundation under Grant No. IRI-0238261 and DUE-0338623.



Photo: Photo by Gordon Gammie, 2004. <http://www.flickr.com/photos/gammie/>



IVC Software Framework – The Team

Master Minds/Programmers

Jason Baumgartner, SLIS
 Nathan James Deckard, CS
 Nihar Sheth, Informatics
Bruce William Herr, CS
Shashikant Penumarthy, SLIS



Graphic Design

Caroline Courtney, Fine Art

Project Start

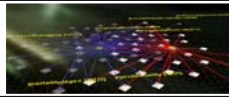
2001

Algorithm Development and Integration

Vivek Agrawal, Summer Intern
 Renee LeBeau, SLIS
 Josh Bonner, CS
 Todd Holloway, CS
 Jeegar Maru, CS
 Laura Northrup, CS
 Sriram Raghuraman, Informatics
 Nihar Sanghvi, Informatics
 Hardik Sheth, Informatics
 Sidharth Thakur, CS
 Ning Yu, SLIS
 Yuezheng Zhou, CS

Students taking K. R. Subramanian's (UNC Charlotte) InfoVis class integrated diverse algorithms into the IVC.

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



IVC Software Framework – Algorithms

(<http://iv.slis.indiana.edu/sw>)

Software

- [XML Toolkit](#)
- [Preprocessing](#)
- [Data Mining](#)
- [Layout Algorithms](#)

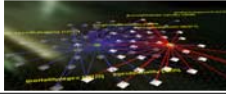
- [Social Visualizations](#)
- [Interaction Algorithms](#)
- [Other Resources](#)

This page provides pointers to commonly used data analysis and visualization algorithms. An XML Toolkit was implemented to facilitate the efficient visualization of diverse data sets as well as an easy comparison of visualizations generated by different algorithms. The toolkit provides a unified architecture in which algorithms can be easily incorporated. Many software packages are available in Java and hence can be run on any platform that supports Java 1.4.

Most software packages come with

- Algorithm Description
- Pros & Cons
- Sample Applications
- Implementation Details
- Usage Hints
- References
- Acknowledgements

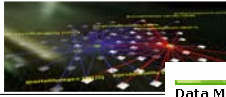
Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



Preprocessing

- Parsers & Converters
- Stop Word Removal
- Porter Stemming Algorithm
- NICE stemmer

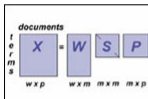
Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



Data Mining

TF x IDF

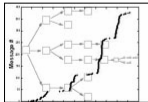
Vector Space Model
Developed by Gerard Salton
Soon to be in the [XML Toolkit](#)



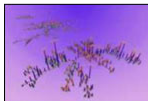
Latent Semantic Analysis
Developed by Tom K. Landauer and Susan Dumais
[Code in XML Toolkit](#)
[Original code by Michael Berry](#)



Topics Model
Developed by Tom Griffith & Mark Steyvers
Soon to be available via the [XML Toolkit](#)

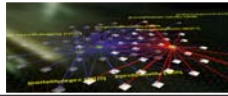


Burst Detection
Developed by Jon Kleinberg
Soon to be available via the [XML Toolkit](#)



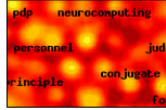
Pathfinder Network Scaling
Developed by Roger Schwanefeldt
KNOT Tools for Pathfinder Network Analysis are available via [Interlink Inc.](#)

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



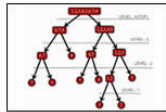
Multidimensional Scaling

Developed by Roger N. Shepard
Fast non-linear MDS algorithm by [Matthew Chalmers](#) and [Alistair Morrison](#) will soon to be available in the [XML Toolkit](#)



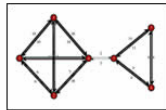
Self Organizing Maps

Developed by Teuvo Kohonen,
[Original code](#) from the [WEBSOM research group](#)



Clustering: Ward's Algorithm

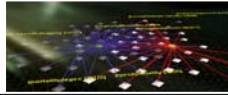
Developed by Ward
[Code in XML Toolkit](#)



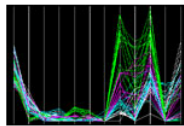
Clustering: Betweenness Centrality

Developed by Ulrik Brandes
[Code in XML Toolkit](#)

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004

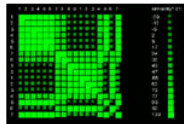


Layout Algorithms



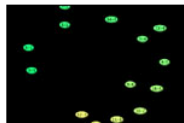
Parallel Coordinates

Developed by A. Inselberg
Soon to be available via in [XML Toolkit](#)



SimVis

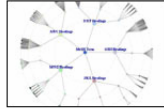
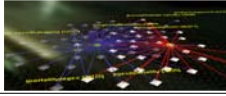
Developed by Yuezheng Zhou
[Code in XML Toolkit](#)



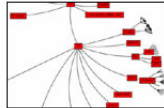
Spring Embedding Algorithm

Developed by Eades
[Code in XML Toolkit](#)

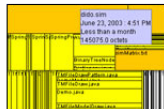
Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



Radial Tree
Implemented by Jason Baumgartner and Nihar Sheth
[Code in XML Toolkit](#)



Hyperbolic Tree
[Code in XML Toolkit](#)

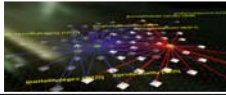


Treemap
Development led by Catherine Plaisant and Ben Shneiderman
[Code in XML Toolkit](#)
[Treemap code](#) available via HCIL@UMD



GRIDL
Designed and developed by Anne Rose, David Feldman, and Ben Shneiderman, with software improvements by Harry Hochheiser
[Code in XML Toolkit](#)

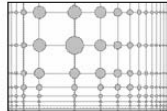
Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



Interaction Algorithms



Fisheye Table
Developed by Ben Bederson
[Code in XML Toolkit](#)
[Original code](#) available via HCIL@UMD

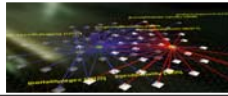


Distortion
Code available from UC Berkeley's User Interface Research Group's [prefuse](http://prefuse.sourceforge.net) (<http://prefuse.sourceforge.net>)



Zooming Pan
Code available from UC Berkeley's User Interface Research Group's [prefuse](http://prefuse.sourceforge.net) (<http://prefuse.sourceforge.net>)

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



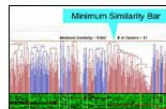
Software Packages



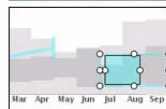
GeoZui3D
 Developed by [Colin Ware](#)'s Data Visualization Research Lab
 It is a Zooming User Interface - hence 'Zui'. It is georeferenced - hence GeoZui. It emphasizes interactive 3D solutions - hence GeoZui3D
[Original code](#) is available at free of charge for non-commercial users.



Worldmapper & User Trail & Chat Log Visualizations
 Developed by [Shashikant Penumarthy](#) and [Katy Börner](#) at the InfoVis Lab at IU.
 It is available for [use](#) and documented [online](#).

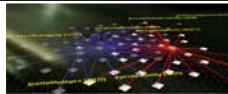


Hierarchical Clustering Explorer
 Developed by Jinwook Seo and Ben Shneiderman
[Original code](#) available via [HCIL@UMD](#)



Time Searcher
 Developed by Harry Hochheiser and Ben B. Shneiderman
[Original code](#) available via [HCIL@UMD](#)

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



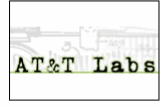
Open Source Toolkits



INRIA's InfoVis Toolkit
 by [Jean-Daniel Fekete](#)
 Interactive Graphics Toolkit written in Java to ease the development of Information Visualization applications and components.
Reference: Jean-Daniel Fekete, [The InfoVis Toolkit](#), Research Report RR-4818, INRIA Futurs, May 2003.



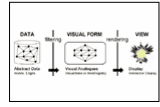
University of Maryland's Piccolo Toolkit
 by Jesse Grosjean and [Ben Bederson](#) at the [HCIL@UMD](#)
 Piccolo is an toolkit for the creation of robust graphical applications with features such as zooming and multiple representation. It is based on the Java2D and available as OSI Certified Open Source Software.



AT&T's GraphViz
 by Stephen North, Emden Gansner, John Ellson et al.
 Set of graph drawing tools for Unix or MS-Windows (win32), including a web service interface (webdot). Source code and binary executables for common platforms are available.

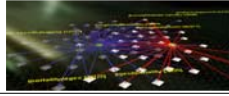


UC Irvine's Java Universal Network/Graph Framework (JUNG)
[\(http://jung.sourceforge.net/\)](http://jung.sourceforge.net/)
 by Scott White, Joshua O'Madadhain, Danyel Fisher and Yan-Biao Boey
 Java-based open-source software library designed to support the modeling, analysis, and visualization of data that can be represented as graphs. It comprises a wealth of algorithms developed in the fields of social network analysis, information visualization, knowledge discovery and data mining.



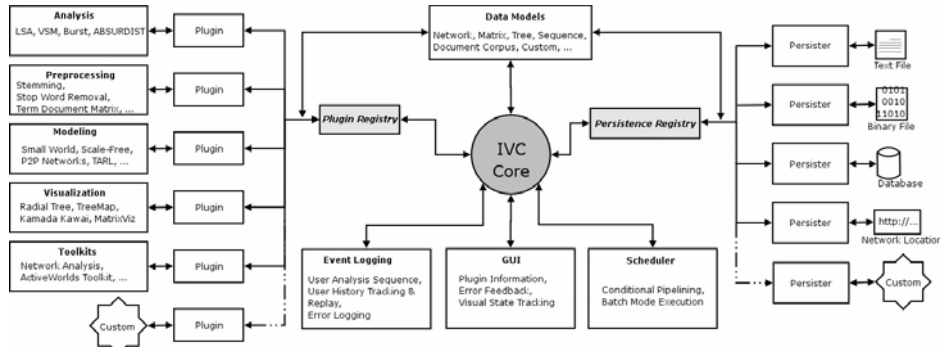
UC Berkeley's User Interface Research Group's prefuse
[\(http://prefuse.sourceforge.net/\)](http://prefuse.sourceforge.net/)
 by [Jeffrey Heer](#) and Alan Newberger. Advised by Stuart K. Card and James A. Landay.
 Interface toolkit for building highly interactive visualizations of structured and unstructured data.
Reference: [prefuse: a toolkit for interactive information visualization](#), Jeffrey Heer, Stuart K. Card and James A. Landay. Submitted paper draft, April 2004.

Katy Börner: InfoVis

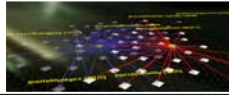


IVC Software Framework – Core

<http://iv.slis.indiana.edu/sw>

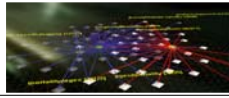


Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



Demo IVC Software Framework

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



Downloads via Sourceforge since June 21, 2004 (<http://sourceforge.net/projects/ivc>)

Statistics for the past 6 months.

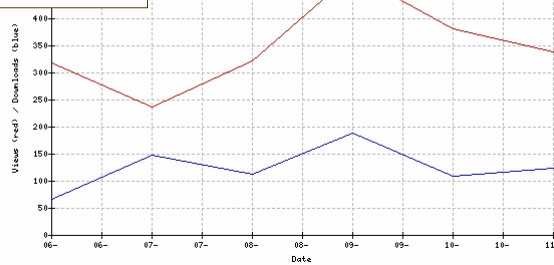
| Month | Rank | Page Views | D/I |
|----------------|----------------|------------|-----|
| November 2004 | 4236 (75.50) | 338 | 124 |
| October 2004 | 6630 (62.05) | 381 | 109 |
| September 2004 | 4908 (70.56) | 483 | 188 |
| August 2004 | 5500 (67.99) | 322 | 112 |
| July 2004 | 6820 (60.02) | 236 | 147 |
| June 2004 | 1081 (94.24) | 317 | 66 |

Statistics for All Time

| Lifespan | Rank | Page Views | D/I |
|----------|----------------|------------|-----|
| 149 days | 4863 (71.73) | 2,077 | 746 |

Usage Statistics

SourceForge.net Statistics: InfoVis Cyberinfrastructure
Page Views (red) and Downloads (blue) for the past 6 months



Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004

Workshop on Information Visualization Software Infrastructures

Sat. Oct. 9th, 2004, 10am-6pm
Room Hill Country A-B

Right before the [IEEE Symposium on Information Visualization \(InfoVis\) 2004](#).

Workshop Chairs

- **Jean-Daniel Fekete**, INRIA Futurs, France (Author of [The InfoVis Toolkit](#)),
Jean-Daniel.Fekete@inria.fr, Home page: <http://www.lri.fr/~fekete/index.en.html>
- **Katy Börner**, Indiana University, USA (Co-Author of the [InfoVis Cyberinfrastructure](#))
katy@indiana.edu, Home page: <http://ella.slis.indiana.edu/~katy/>

Description

Information visualization systems and toolkits are becoming available for a large range of visualization and interaction techniques and are used in diverse application domains. This workshop is aimed at gathering experts involved in building such infrastructures to share their views, understand the issues involved and trying to find ways to avoid fragmentation and improve collaborations.

To participate in the workshop, you should submit a semi-structured position paper explaining your view of what an infrastructure should provide, describe what you consider as the main challenges for such infrastructures and describe the capabilities of toolkits of systems you have already built, following a form available [here](#) by Sept 30th, 2004.

<http://www.indiana.edu/ivsi2004/>



The Tower of Babel by Pieter Bruegel

Information Visualization CyberInfrastructure

The InfoVis CyberInfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse data sets.

DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated.
<http://iv.slis.indiana.edu/db>

SOFTWARE

An open source IVC framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feels for the interface and even entire toolkits can be easily "plugged in" or "unplugged".
<http://iv.slis.indiana.edu/sw>

COMPUTING RESOURCES

The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex consisting of two Sun V1280 servers with 12 900MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V100 system with 4 cpus and 512M memory serves as the web front-end for the database servers.
<http://iv.slis.indiana.edu/cr>

LEARNING MODULES

A set of associated learning modules aims to equip learners with a practical skill set by providing code and advice to quickly modify and run different algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations.
<http://iv.slis.indiana.edu/lm>

InfoVis Lab, School of Library and Information Science, Indiana University (2006).
 For more information, contact Katy Börner at kborner@slis.indiana.edu
 This material is based upon work supported by the National Science Foundation under Grant No. IRI-0238261 and DUE-0333623.

IVC Learning Modules

<http://iv.slis.indiana.edu/iv>

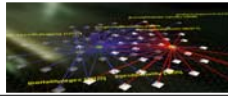
Learning Modules

Most information visualizations are highly interactive. While a number of excellent textbooks exist, the two-dimensional printouts on paper often cannot convey their true visual appearance and interactive performance. Several textbooks come with accompanying web sites that contain snapshots of user interfaces as well as animations and movies. However, none of them facilitates the exploration, application, evaluation, and comparison of algorithms.

This web page will provide access to a number of learning modules. Each learning module comes with an:

- ◆ Introduction
- ◆ Discussion of Existing Algorithms
- ◆ Learning Task
 - ◊ A challenging scenario to use the code in the **XML Toolkit** to visualize a data set to support a specific user group.
- ◆ Programming Exercise
 - ◊ An programming exercise plus an explanation of possible solutions.
- ◆ Opportunities & Challenges, and
- ◆ References to research papers, online demos, (commercial) applications

Katy Börner: InfoVis CyberInfrastructure, SLIS Colloquium, November 19th, 2004



Visualizing Tree Data

<http://iv.slis.indiana.edu/lm/lm-trees.html>

[Learning Modules](#) > Visualizing Tree Data

[Description](#) | [Usage Hints](#) | [Learning Task](#) | [Discussion](#) | [References](#) | [Acknowledgments](#)

Description

Many data sets come in tree format. There are family trees, organizational charts, classification hierarchies, and directory structures. The figure below shows an inheritance tree by Ernst Haeckel ('Stammbaum' in German). Read also [To Draw a Tree](#) by Pat Hanrahan.

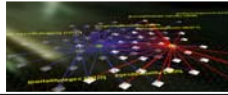


[Click image for larger version](#)

A tree graph is a set of straight line segments (edges) connected at their ends containing no closed loops (cycles). You can also call it a simple, undirected, connected, acyclic graph (or, equivalently, a connected forest). A tree with n nodes has $n-1$ graph edges. All trees are bipartite graphs.

Many trees have a root node and are called rooted trees. Trees without a root node are called free trees. Subsequently, we will only consider rooted trees. In rooted trees, all nodes except the root node have only one parent node. Nodes which have no children are called leaf nodes. All other nodes are referred to as intermediate nodes.

Katy Börner: InfoVis CyberInfrastructure, SLIS Colloquium, November 19th, 2004



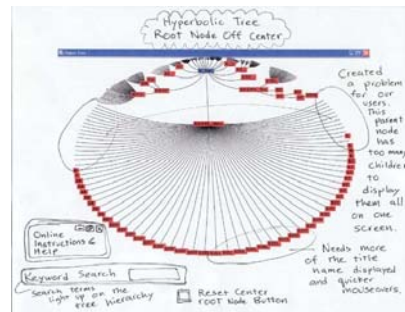
Student's Project Results

User & Task Analysis for Visualizing Tree Data

- Visualizing the structure of IU's Decision Support System
- Visualizing the co-occurrences of keywords in DLib Magazine articles.
- Visualization of the Java API
- Visualizing the the Library of Congress Classification System to retrieve legal materials in a library.

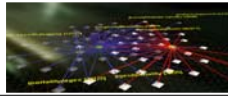
See Handin pages at

<http://ella.slis.indiana.edu/~katy/handin/L579-S04/cgi/handinlogin.cgi>



Katy Börner: InfoVis CyberInfrastructure, SLIS Colloquium, November 19th, 2004

Image by Peter Hook and Rongke Gao



Time Series Analysis & Visualization

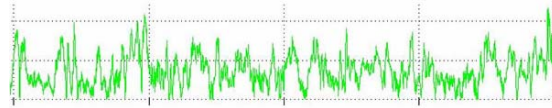
<http://iv.slis.indiana.edu/lm/lm-time-series.html>

[Learning Modules](#) > Visualizing Time Series Data

[Description](#) | [Usage Hints](#) | [Learning Task](#) | [Discussion](#) | [References](#) | [Acknowledgments](#)

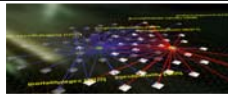
Description

A time series is a sequence of events/observations which are ordered in one dimension, e.g., time. Frequently, successive observations depend on each other and it makes sense to display them in a (time) sorted fashion, e.g., as a scatter plot. Alternatively, one could be interested to know how many observations of a certain value have been made. Here one would sort the observations by value, count the number of observations for each value and derive a histogram. Time series data can be continuous, i.e., there is an observation at every instant of time see figure below, or discrete, i.e., observations exist for regularly or irregularly spaced intervals.



Time series are recorded, analyzed and used in diverse domains of science. Check out the [Time Series Data Library](#) maintained by Rob Hyndman and Muhammad Akram for numerous data sets from Agriculture, Chemistry, Crime, Demography, Ecology, Finance, Health, Hydrology, Industry, Labour market, Macro-Economics, Meteorology, Micro-Economics, Physics, Production, Sales, Simulated series, Sport, Transport & Tourism or Utilities.

Katy Börner: InfoVis CyberInfrastructure, SLIS Colloquium, November 19th, 2004



Student's Project Results

Time Series Analysis & Visualization

- Using Timesearcher and the Burst Detection Algorithm to Analyze the Stock Market from 1925 to 1945
- Applying Burst and TimeSearcher to Chat Data
- Lab Access Trends
- Quest Atlantis Chat Log Data

See Handin pages at

<http://ella.slis.indiana.edu/~katy/handin/L579-S04/cgi/handinlogin.cgi>

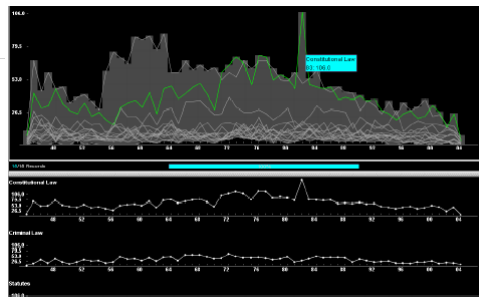
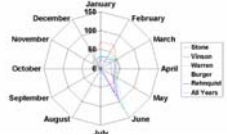
Katy Börner: InfoVis CyberInfrastructure, SLIS Colloquium, November 19th, 2004

Visualizing the Work of the United States Supreme Court Based on Time Data and Top Level West Topics

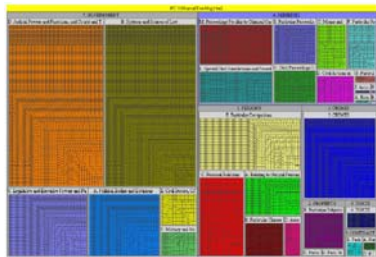
by Peter A. Hook & Rongke Gao



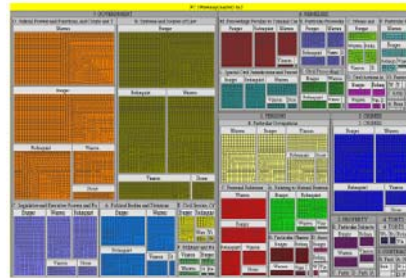
Topics Per Month Per Court



Top fifteen most occurring topics from 1944 to 2004 in Timesearcher



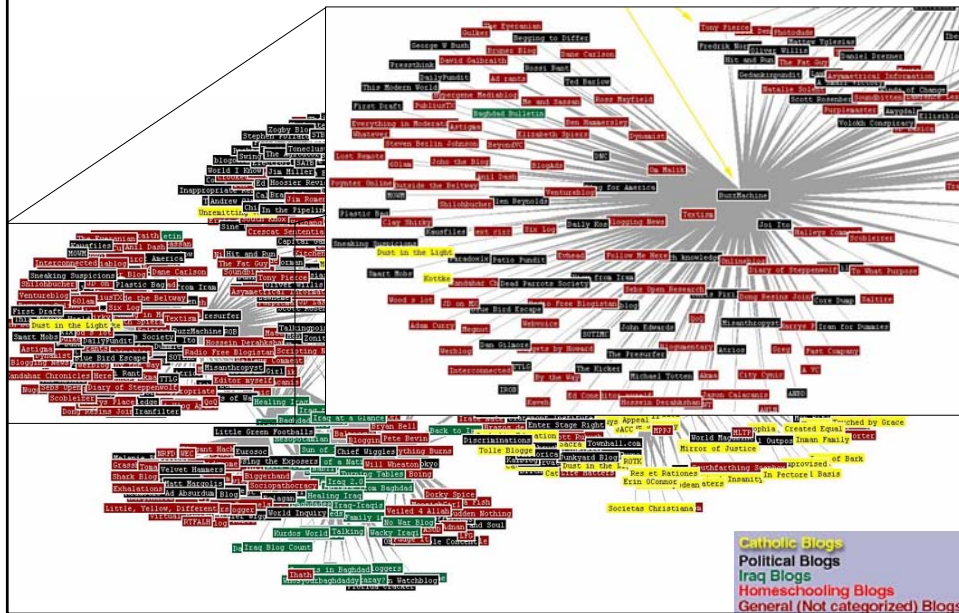
All topics grouped by West Category and Sub-Category grouped over the entire lengths of the data set

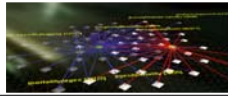


All topics by West Category and Sub-Category grouped corresponding to the five chief justices

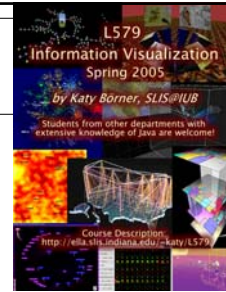
Visualizing Niches of the Blog Universe

BY Mike Tynorth and Elijah Wright





L579 Information Visualization Spring 2005



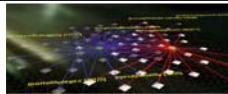
This course covers

- Perceptual basis of information visualization.
- Data mining algorithms that enable extraction of relationships in data.
- Visualization and interaction techniques.
- Discussions of systems that drive research and development, and
- Future trends and remaining fundamental problems in the field.

Students do weekly readings, provide a presentation on specific readings, do projects, and participate in class & online discussion.

Class Webpage: <http://ella.slis.indiana.edu/~katy/L579>

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004

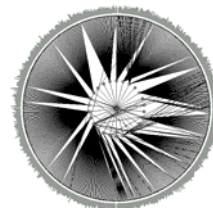
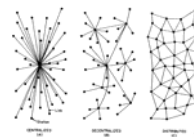


L597 Structural Data Mining and Modeling Fall 2005

This course

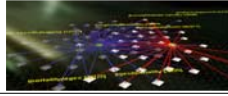
- Introduces students to major methods, theories, and applications of structural data mining and modeling.
- Covers elementary graph theory and matrix algebra, data collection, structural data mining, data modeling, and applications.

Upon taking this course students will be able to analyze and describe real networks (power grids, WWW, social networks, etc.) as well as relevant phenomena such as disease propagation, search, organizational performance, social power, and the diffusion of innovations.



Format: Lectures and 4-5 labs.

Class Webpage: <http://ella.slis.indiana.edu/~katy/L597>



Future Work

IVC Database

- Create tables/upload Citeseer, 110 year Physical Review journals dataset, etc.
- Optimize online interface and make it available to other researchers.
- Create connections to R and other packages for large scale (network) data analysis.
- Document, document, document.

IVC Software Framework

- Release IVC core as 'alpha'.
- Integrate a lot more algorithms.

IVC Learning Modules

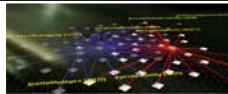
- Write new learning modules as new algorithms become available.
- User test learning modules.

Outreach

There will be a [Data Analysis, Modeling and Visualization](#) Tutorial @ [Electronic Imaging](#), San Jose, CA, Jan 16th, 2005 which uses the IVC infrastructure.

Do RESEARCH using this infrastructure!

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004



Acknowledgements

The Information Visualization Software Repository was created in 2000 and has since then been used to teach the Information Visualization class at Indiana University. Katy Börner, Yuezheng Zhou, and Jason Baumgartner implemented the very first algorithms. In Summer 2003, Jason Baumgartner, Nihar Sheth, and Nathan J. Deckard lead a project to design a XML toolkit that enables the serialization and parallelization of commonly used data analysis and visualization algorithms. Contributions of software packages and implementation work are acknowledged on the respective software pages.

Support comes from the School of Library and Information Science, Indiana University's High Performance Network Applications Program, an Academic Equipment Grant by SUN Microsystems, SBC (formerly Ameritech) Fellow Grant, and the National Science Foundation under DUE-0333623 and IIS-0238261.



Craig A. Stewart, Mary Papakhian, Anurag Shankar all UTTS generously made the Research Database Complex available for this project and provided very insightful comments.

Stephanie Burks, Principal Unix Systems Administrator, Research and Technical Services, UTTS has been instrumental in setting up the computing infrastructure and administration of the Oracle database.

Katy Börner: InfoVis Cyberinfrastructure, SLIS Colloquium, November 19th, 2004

Information Visualization CyberInfrastructure

The InfoVis CyberInfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse data sets.

DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated. (<http://rxrx.indiana.edu/ib3/>)



COMPUTING RESOURCES

The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1200 servers with 12 x 900MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V480 system with 4 CPUs and 6GB memory serves as the web front-end for the database servers. (<http://rxrx.indiana.edu/ib3/>)



InfoVis Lab, School of Library and Information Science, Indiana University (2001). For more information, contact Katy Börner at kborner@indiana.edu

This material is based upon work supported by the National Science Foundation under Grant No. IRI-0238261 and DUE-0333623.

SOFTWARE

An open source VC framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feels for the interface and entire toolkits can be easily "plugged in" or "unplugged". (<http://rxrx.indiana.edu/ib3/>)



LEARNING MODULES

A set of associated learning modules aims to equip learners with a practical skill set by providing code and advice to quickly modify and run different algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations. (<http://rxrx.indiana.edu/ib3/>)



Photo: Andrew Carter/Sony.com