

An Introduction to Information Visualization Techniques for Exploring Large Database

Jing Yang
Fall 2005

1

Social Visualization

Class 12, Part A

Reference: A large number of slides in this class
come from John Stasko's Infovis class slides.
They are used with his permission.

2

Definition

- Social Visualization
 - “Visualization of social information for social purposes”
---Judith Donath, MIT
 - Visualizing data that concerns people or is somehow people-centered

3

This slide is from John Stasko's Infovis class slides

Example Domains

- Social visualization might depict
 - Baby names
 - Conversations
 - Newsgroup activities
 - Email patterns
 - Chat room activities
 - Presence at specific locations
 - Social networks
 - Life histories

4

This slide is partially from Stasko's Infovis class slides.

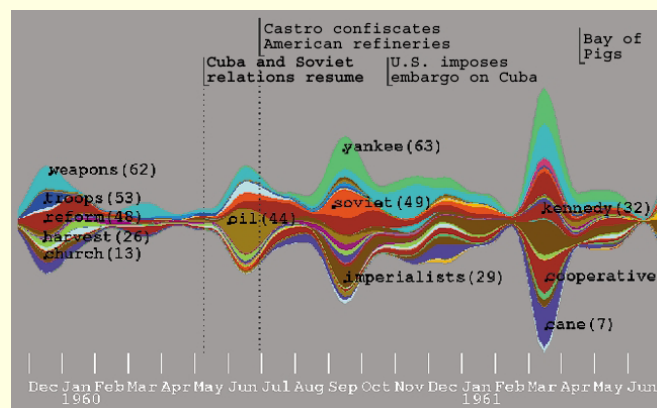
Baby Name Visualization

- Baby Names, Visualization, and Social Data Analysis [Wattenberg Infovis 2005]
- NameVoyager – a web-based visualization applet
 - Let users interactively explore name data, historical name popularity figures
 - Demo
 - More than 500,000 site visits in the first two week
 - Average of 10,000 visits per day after two months
- Lesson – To design a successful exploratory data analysis tool, one good strategy is to create a system that enables “social” data analysis

5

Baby Name Visualization

- Similar to Themeriver



6

Social Network Visualization

- Vizster: Visualizing Online Social Networks [Heer Infovis 05]
- Online social networks – millions of members publicly articulate mutual “friendship” relations
 - Friendser.com, Tribe.net, and orkut.com
- Vizster
 - Playful end-user exploration and navigation of large-scale online social networks
 - Explore connectivity, support visual search and analysis, and automatically identifying and visualizing community structures
 - Video

7

Social Network Visualization

- Vizster: Visualizing Online Social Networks [Heer Infovis 05]
- Usage observation
 - 500-person all-night event
 - Many party-goers are familiar with the friendster system
 - Interactive kiosk and a projection of the visualization onto a large screen

8

Email Visualization

- THREAD ARCS: An Email Thread Visualization [Kerr Infovis 2003]
- Thread Arcs
- Combine the chronology of messages with the structure of a conversational thread
- Help people learn various attributes of conversations and find relevant messages

9

THREAD ARCS: An Email Thread Visualization [Kerr Infovis 2003]

- Basic ideas:

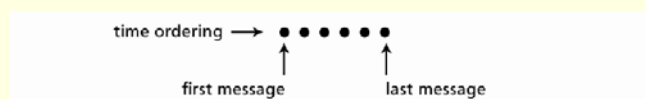


Figure 1. Chronology of message nodes in a line of six messages.

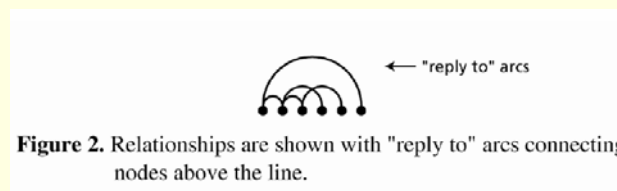


Figure 2. Relationships are shown with "reply to" arcs connecting nodes above the line.

10

THREAD ARCS: An Email Thread Visualization [Kerr Infovis 2003]

■ Design choices:



Figure 4. The relationship between messages are clearer when arcs are drawn above and below nodes (B).

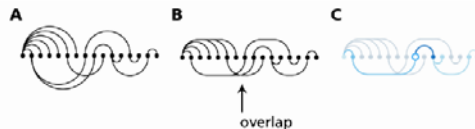


Figure 7. Unconstrained (A) and constrained arc heights (B) for a thread of sixteen messages. A selection highlighting scheme for the ninth message (C).

11

THREAD ARCS: An Email Thread Visualization [Kerr Infovis 2003]

■ Highlight strategies:

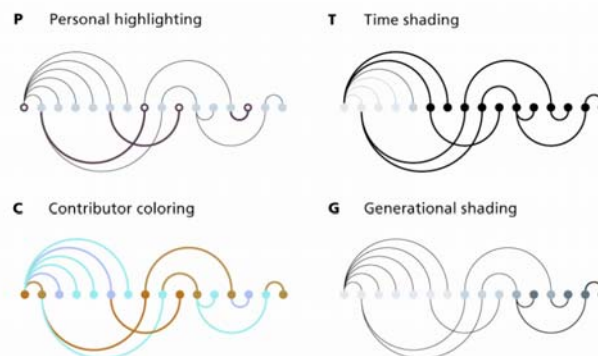


Figure 16. Different message attributes highlighting schemes.

12

THREAD ARCS: An Email Thread Visualization [Kerr Infovis 2003]

■ Prototype:



Figure 18. Thread Arcs integrated into email client prototype.

13

Chat Room Visualization

■ Chat Circles [Viegas and Donath CHI'99]

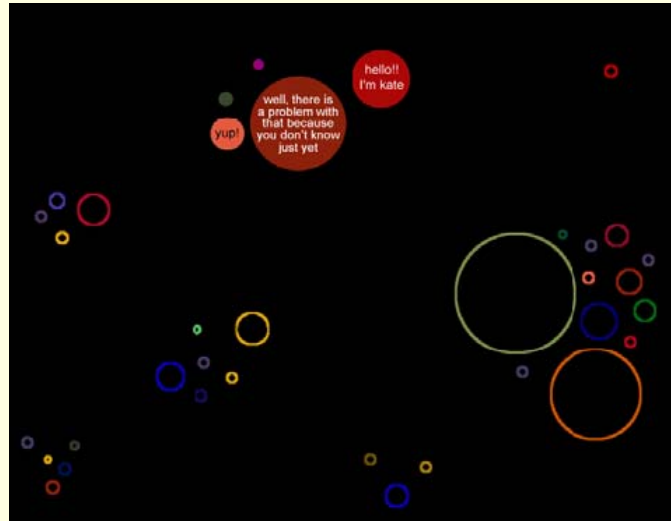
<http://chatcircles.media.mit.edu/about.html>

You can try it out!

- GUI for chat rooms
- Represent people using circles
- Mimics cocktail party in certain ways

14

Chat Circles [Viegas and Donath CHI'99]



15

Chat Circles [Viegas and Donath CHI'99]



16

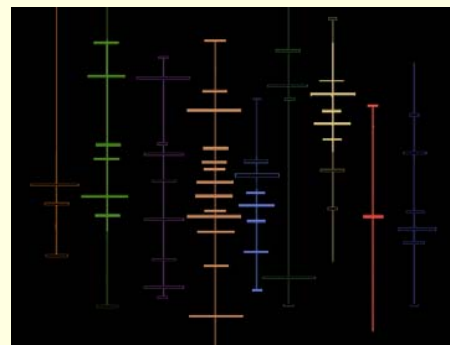
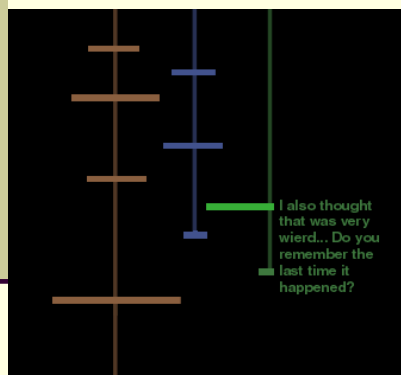
Chat Circles [Viegas and Donath CHI'99]

- Each participant is a colored circle
- Circle grows with each posted message, slowly shrinks/fades as goes idle
- Will stay there as small circle while connected
- Comments appear inside circles
- Can only “hear” what is going on nearby

17

Chat Circles [Viegas and Donath CHI'99]

- History interface



18

Chat Circles [Viegas and Donath CHI'99]

- Mapping
 - Individual users on x-axis
 - Time goes up on y-axis
 - Tick marks are postings, mouse over reveals them
 - Solid tick marks were within earshot of you, hollow ones weren't
- Try it live
<http://chatcircles.media.mit.edu/>

19

Chat Circles [Viegas and Donath CHI'99]

- Each participant is a colored circle
- Circle grows with each posted message, slowly shrinks/fades as goes idle
- Will stay there as small circle while connected
- Comments appear inside circles
- Can only "hear" what is going on nearby

20

Discussion Group Visualization

- Discussion group:
 - Web-based message boards
 - Usenet newsgroups
 - Chatrooms
- Questions:
 - Do participants really get involved?
 - How much interaction is there?
 - Do participants welcome newcomers?
 - Who are the experts?

21

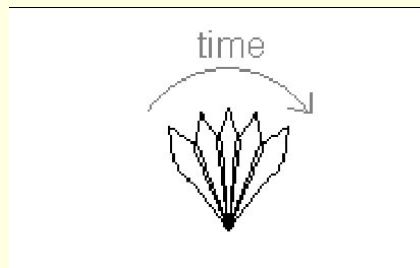
People Garden [Xiong and Donath UIST'99]

- Visualization technique for portraying online interaction environments (Virtual Communities)
- Provides both individual and societal views
- Utilizes garden and flower metaphors

22

Data Portrait: Petals

- Fundamental view of an individual

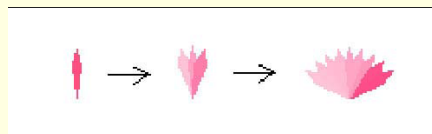


- His/Her postings are represented as petals of the flower, arranged by time in a clockwise

23

Data Portrait: Postings

- Time of Posting

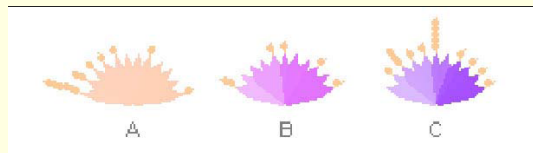


- New posts are added to the right
- Slide everything back so it stays symmetric
- Each petal fades over time showing time since posting
- A marked difference in saturation of adjacent petals denotes a gap in posting

24

Data Portrait: Responses

- Data Portrait: Responses

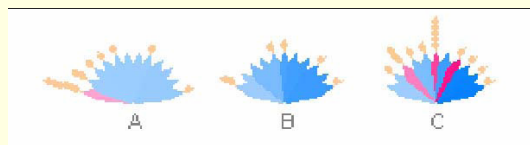


- Small circle drawn on top of a posting to represent each follow-up response

25

Data Portrait: Color

- Initial post vs. reply



- Color can represent original/reply
- Here magenta is original post, blue is reply

26

Garden

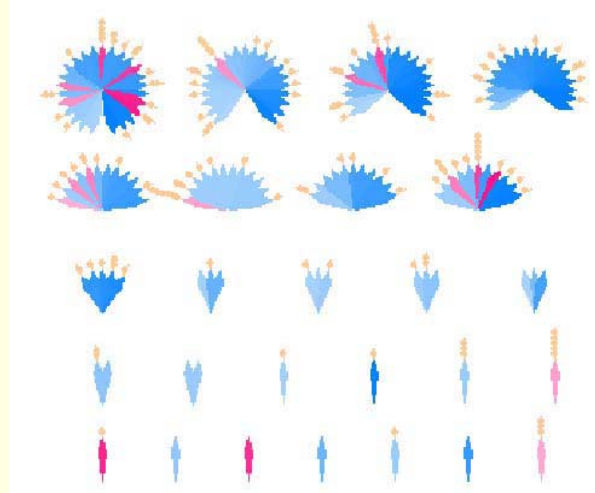
- Combine many portraits to make a garden
- Message board with 1200 postings over 2 months
- Each flower is a different user
- Height indicates length of time at the board



27

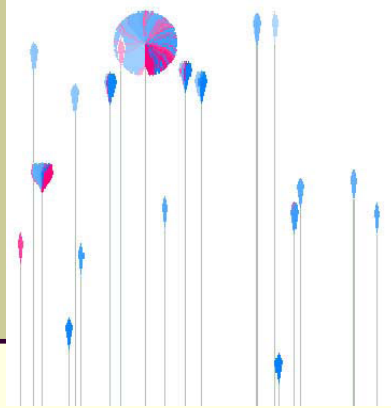
Alternate Garden View

- Sorted by number of postings

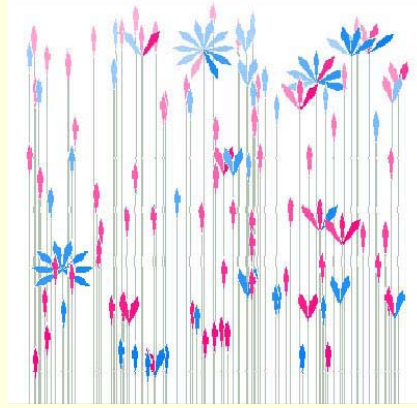


28

Interpreting Displays



Group with one dominating person



More democratic group

29

Software Visualization

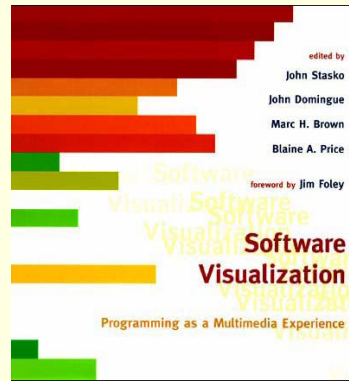
Class 12, Part B

30

Definition

- “The use of the crafts of typography, graphic design, animation, and cinematography with modern human computer interaction and computer graphics technology to facilitate both the human understanding and effective use of computer software.”

Price, Baecker and Small, '98



31

Challenge

- Software clearly is abstract data
- Unlike much information visualization, however, software is often dynamic, thus requiring our visualizations reflect the time dimension
 - – History views
 - – Animation
 - – ...

32

Sub-domains

- Two main sub-areas of software visualization
 - Program visualization - Use of visualization to help programmers, coders, developers. Software engineering focus
 - Algorithm visualization - Use of visualization to help teach algorithms and data structures. Pedagogy focus

33

Program Visualization

- Can be as simple as enhanced views of program source
- Can be as complex as views of the execution of a highly parallel program, its data structures, run-time heap, etc.

34

Enhanced Code Views

```
/**
 * Gets the normalized value of a data.
 * @param valueString the data in a string
 */
double CDoubleDimension::GetNorValue(std::string valueString) {
    double value = atof(valueString.c_str());

    if (min == max)
    {
        if (value == min)
            return 1;
        else
            return 0;
    }

    double v = (value - min) / (max - min);

    if (v < 0)
        v = 0;
    else if (v > 1)
        v = 1;

    return v;
}
```

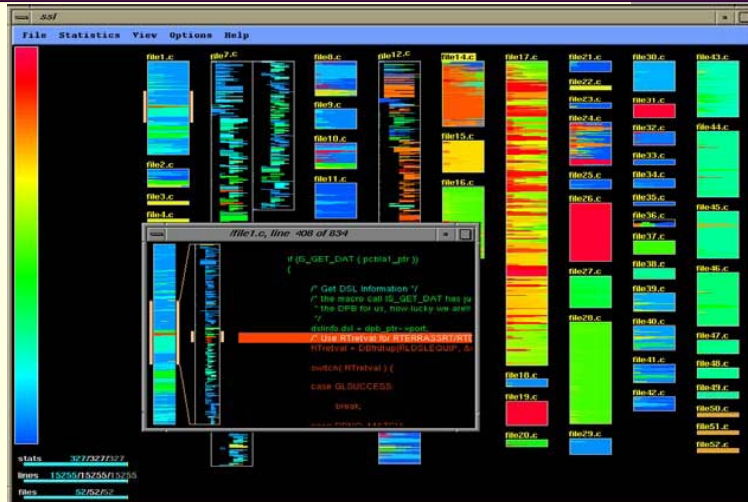
35

SeeSoft System [Eick et al. IEEE ToSE '92]

- Pulled-back, far away view of source code
- Map one line of source to one line of pixels
 - Can indicate line indentation, etc.
 - Use color to represent the programmer, age, or functionality of each line.
- Like taping your source code to the wall, walking far away, then looking back at it

36

SeeSoft System View



37

Use

- Tracking (typically means mapping this data attribute to color)
- Code modification (when, by whom)
- Bug fixes
- Code coverage or hotspots
- Interactive, can change color mappings, can brush views, can compare files, ...

38

Tarantula [Eagan et al. Infovis'01]

- Utilizes SeeSoft code view methodology
- Takes results of test suite run and helps developer find program faults
- Clever color mapping is the key!

39

Color Mapping of Tarantula

- Color reflects a statement's relative success rate of its execution by the test suite.
 - Color spectrum: from red to yellow to green
- Statements executed by a failed test case become more red
- Statements executed by a passed test case become more green
- Statements shown as red are highly suspect
- Statements shown as green convey a strong confidence in their correctness
- Statements shown as yellow convey a sense of ambiguousness,

40

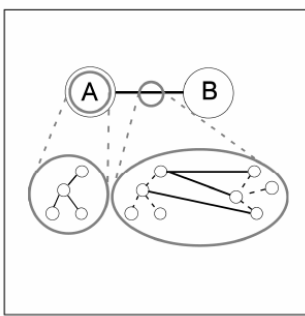
Tarantula View



41

Software Structure Visualization

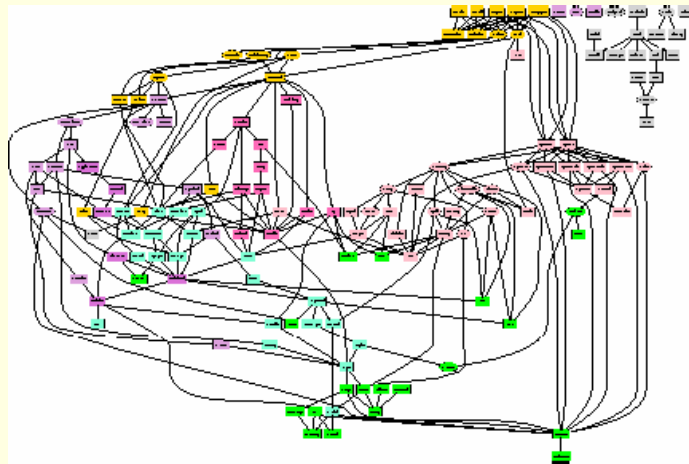
- Call graph visualization
- Flow chart visualization
- Graph visualization!



A call graph

42

Sample Call Graph View



43

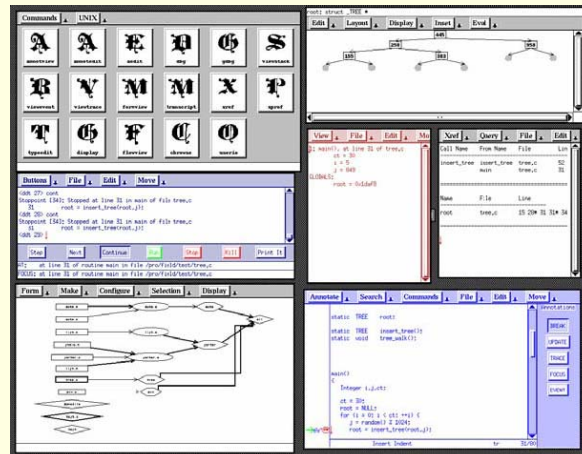
FIELD [Reiss Software Pract & Exp'90]

- Program development and analysis environment with a wide assortment of different program views
 - Integrated a variety of UNIX tools
 - Utilized central message server architecture in which tools communicated through message passing

44

FIELD [Reiss Software Pract & Exp'90]

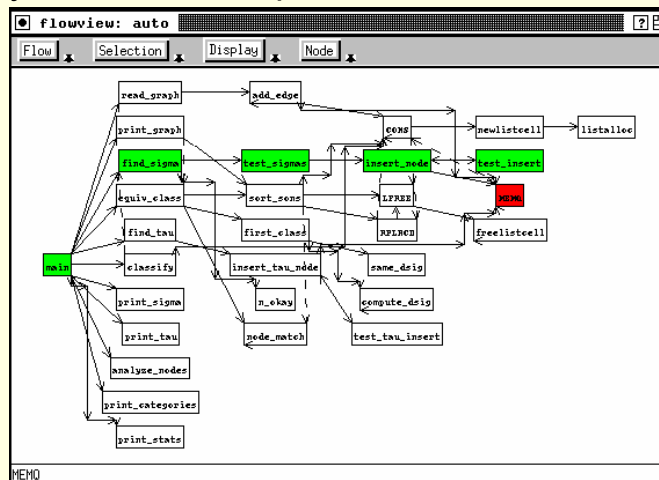
Interface



45

FIELD [Reiss Software Pract & Exp'90]

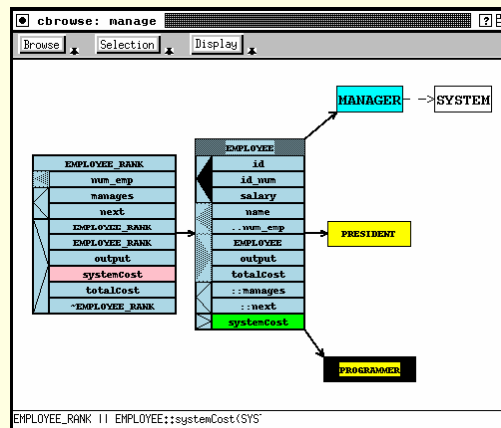
Dynamic Call Graph View



46

FIELD [Reiss Software Pract & Exp'90]

■ Class browser

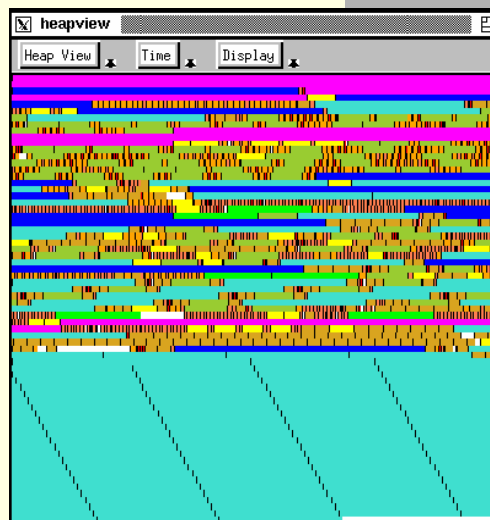


47

FIELD [Reiss Software Pract & Exp'90]

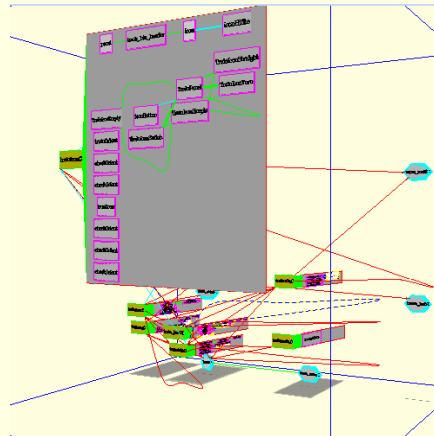
■ Heap View

Color could be
When allocated
Block size
Where allocated



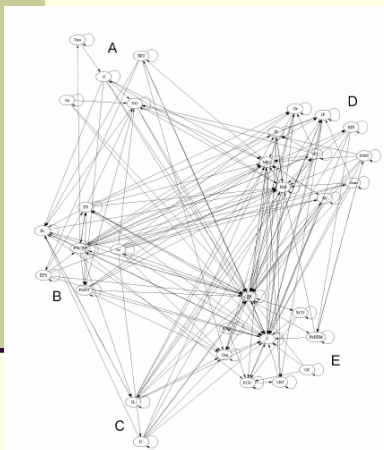
FIELD [Reiss Software Pract & Exp'90]

■ 3D call graph

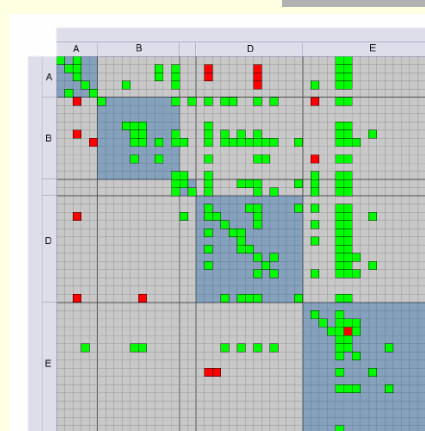


49

Multilevel Call Matrices [vanHan Infovis 2003]



Node-line diagram



Call Matrix

50

Multilevel Call Matrices [vanHan Infovis 2003]

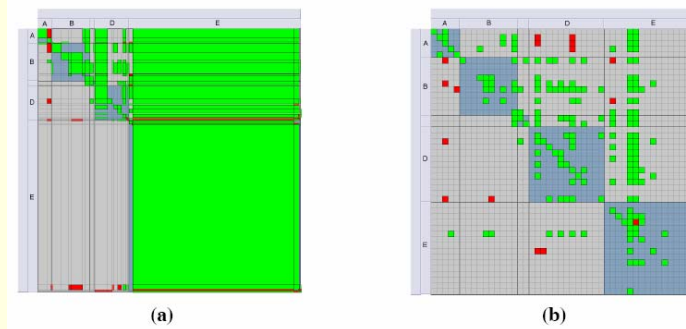


Figure 2: Recursive matrix subdivision according to relative component size (a), Matrix subdivision according to number of subcomponents at a specific abstraction level (b)

51

Multilevel Call Matrices [vanHan Infovis 2003]

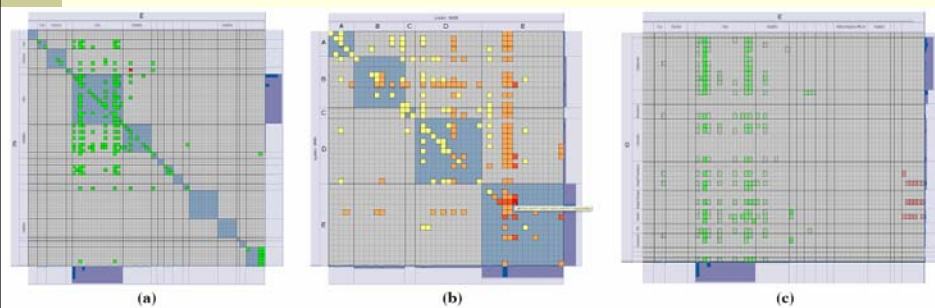


Figure 5: Displaying additional data: call permissions (a), call neighborhood (red calls are closer to call under pointer) and call density (c)

52

PV System [Kimelman et al. Vis94]

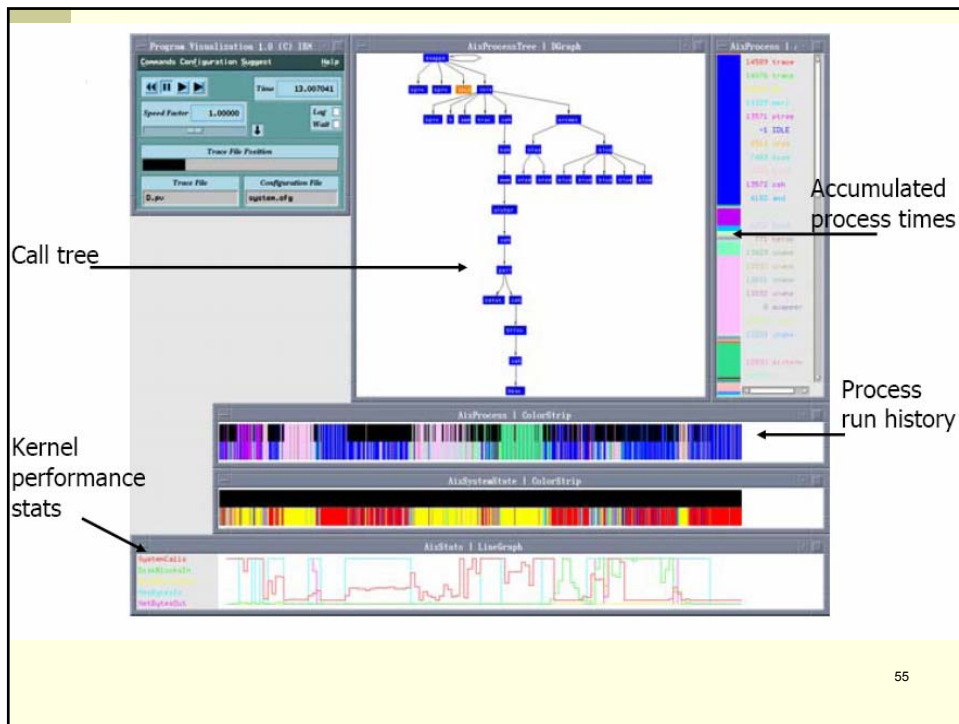
- Used for understanding application and system behavior for purposes of debugging and tuning
- Users look for trends, anomalies, and correlations
- Ran on RISC/6000 workstations using AIX
- Trace-driven, can be viewed on-line or off

53

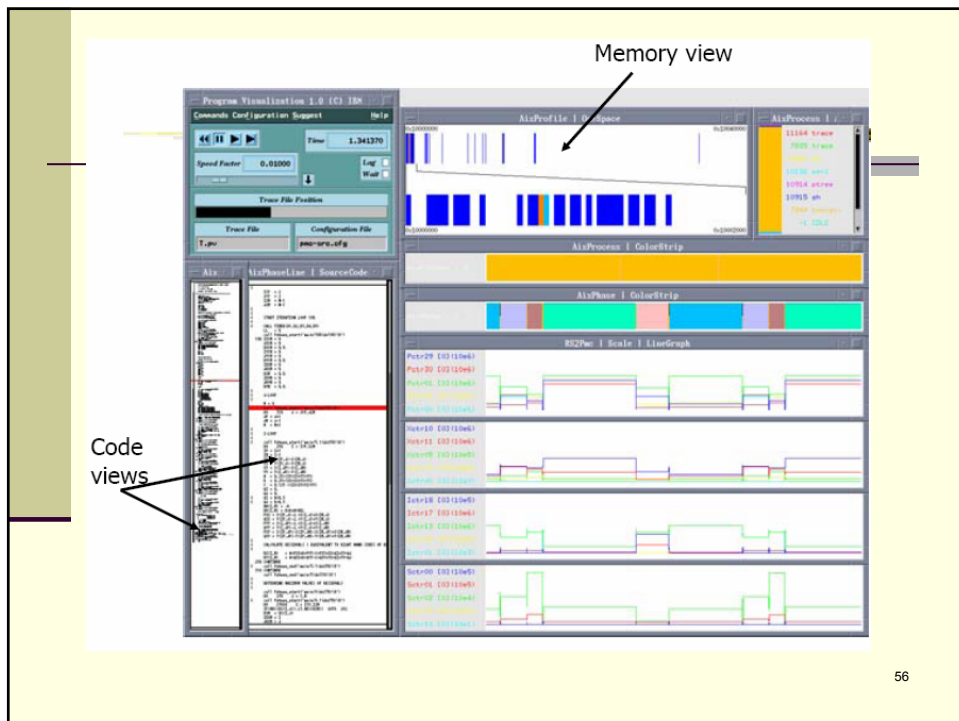
Different Views

- Hardware-level performance info
 - Instruction execution rates, cache utilization,
 - processor utilization
- Operating system level activity
 - Context switches, system calls, address space activity
- Communication library level activity
 - Message passing, interprocessor communication
- Language run-time activity
 - Dynamic memory allocation, parallel loop scheduling
- Application-level activity
 - Data structure accesses, algorithm phase transitions

54



55



56

Commercial Systems

- A number of commercial program development environments have begun to incorporate program visualization tools such as these
 - Majority are PC-based
- Has not become wide-spread

57

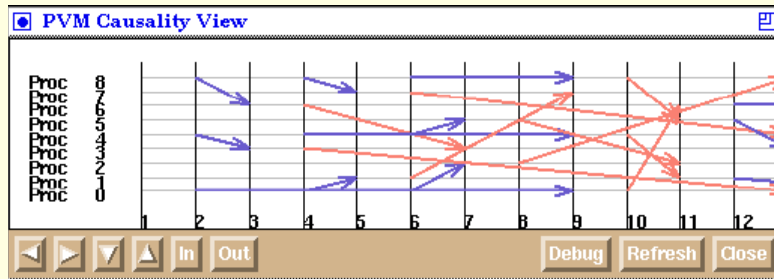
Concurrent Programs

- Understanding parallel programs is even more difficult than serial
- Visualization and animation seem natural for illustrating concurrency
- Temporal mapping of program execution to animation becomes critical
- Example system: POLKA [stasko & Kraemer JPDC '93]

58

Message Passing Systems

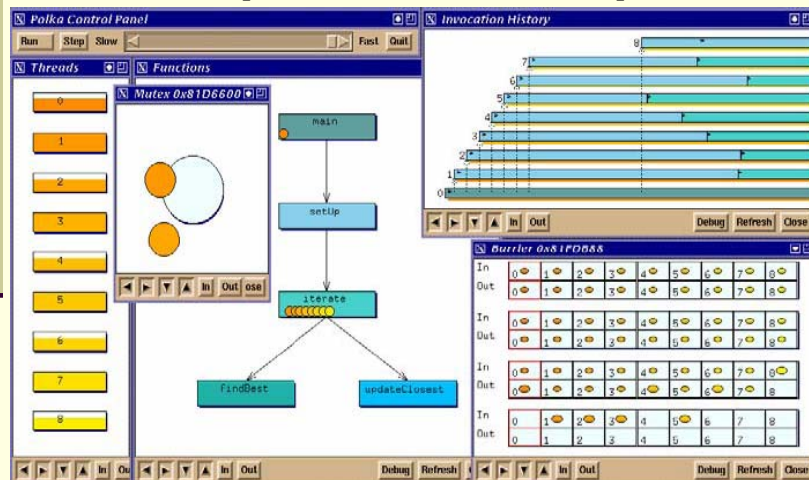
- PVM/Conch [Topol et al. JPDSN '98]



59

Shared Memory Threads

- Pthreads [Zhao & Stasko TR '95]



60

Algorithm Visualization

- Learning about algorithms is one of the most difficult things for computer science students
 - Very abstract, complex, difficult to grasp
- Idea: Can we make the data and operations of algorithms more concrete to help people understand them?

61

Algorithm Animation

- Common name for area
- Dynamic visualizations of the operations and data of computer algorithm as it executes

62

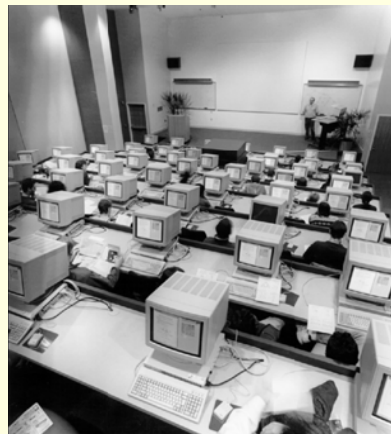
Sorting Out Sorting

- Seminal work in area
- 30 minute video produced by Ron Baecker at Toronto in 1981
- Illustrates and compares nine sorting algorithms as they run on different data sets
- **Video**
(<http://kmdi.utoronto.ca/RMB/publications.html>)

63

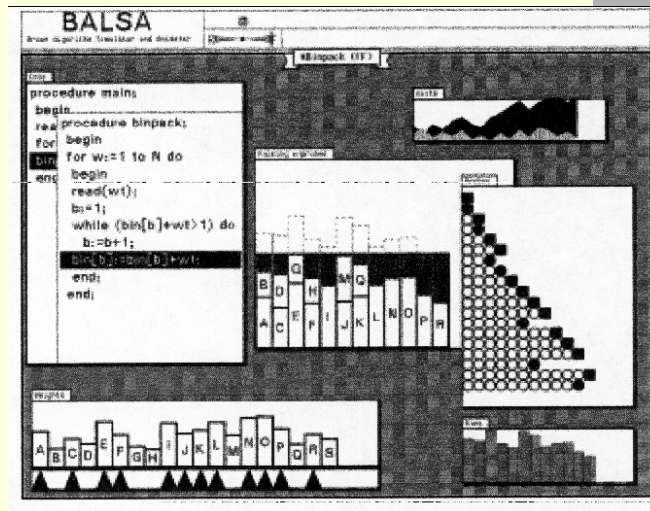
Balsa [M. Brown Computer '88]

- First main system in area
- Used in “electronic classroom” at Brown
- Introduced use of multiple views and interesting event model



64

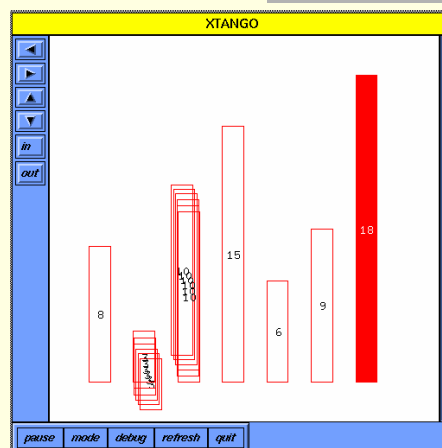
Example Animation



65

Tango [Stasko Computer '90]

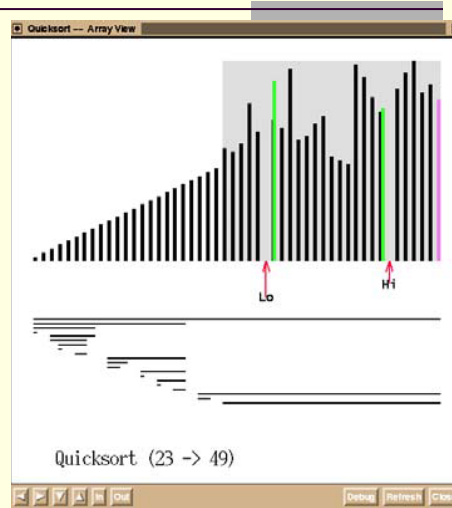
- Smooth animation
- Simplification of the design/programming Process
- Formal model of the animation



66

POLKA [Stasko & Kraemer JPDC '93]

- Improved animation design model
- Object-oriented paradigm
- Multiple animation windows
- Much richer visualization/animation capabilities



67

A Useful Link

- <http://www.cc.gatech.edu/gvu/softviz/SoftViz.html>

68

Text and Document Visualization

Class 12, Part C

69

Text is Everywhere

- We use documents as primary information artifact in our lives
- Our access to documents has grown tremendously in recent years due to networking infrastructure
 - WWW
 - Digital libraries
 - ...

70

Big Question

- What can information visualization provide to help users in gathering information from text and document collections?

71

InfoVis Tasks

- Two main tasks that Information Visualization can assist with in this area
 - Enhance a person's ability to read, understand and gain knowledge from a document
 - Understand the contents of a document or collection of documents without reading them

72

Specific Tasks for Document Collections

- What are the main themes of a document?
- How are certain words or themes distributed through a document?
- Which documents contain text on topic XYZ?
- Which documents are of interest to me?
- Are there other documents that might be close enough to be worthwhile?

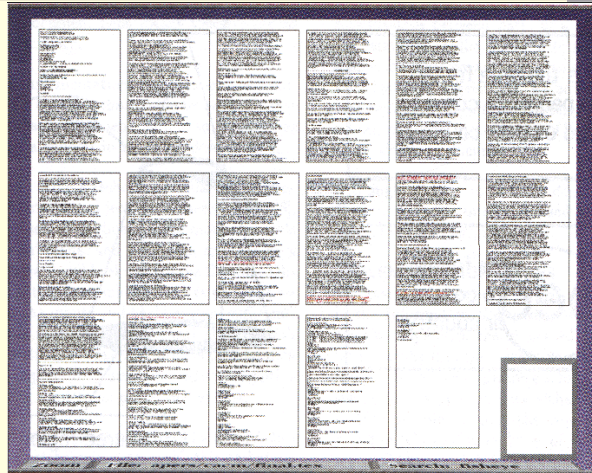
73

Simple Taxonomy

	Enhanced presentation (syntax)	Concepts and relationships (semantics)
Single document		
Collection of documents		

74

Enhanced Presentation of a Document



Text is too small to read

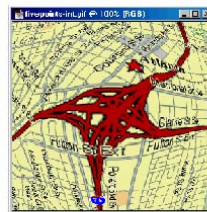
75

Enhanced Presentation of a Document

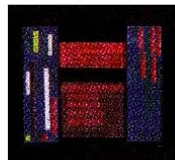
Magnifying
lens



Fisheye
view



Bifocal
display



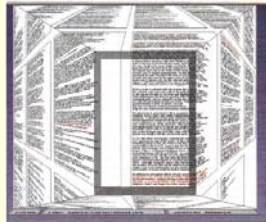
Perspective
wall



76

Enhanced Presentation of a Document

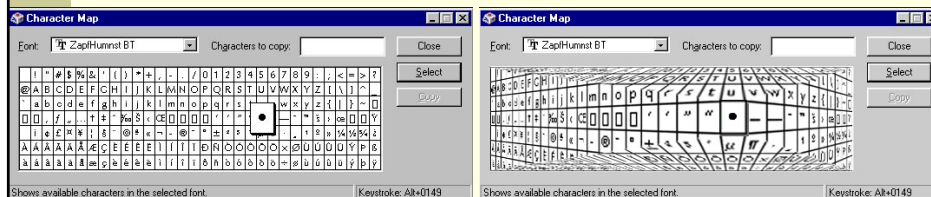
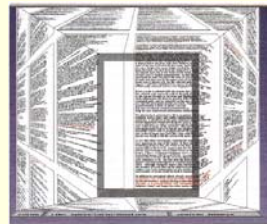
Document Lens



77

Enhanced Presentation of a Document

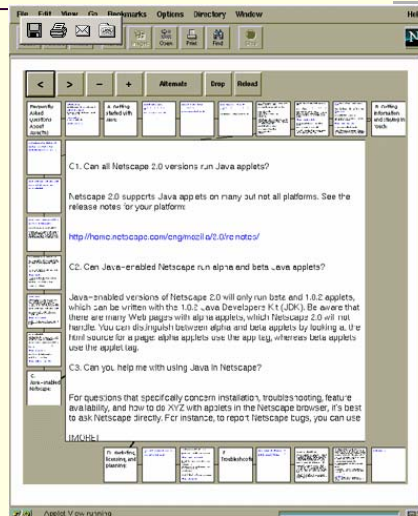
Document Lens



78

Enhanced Presentation of a Document

Zoom Browser



79

Enhanced Presentation of Labels

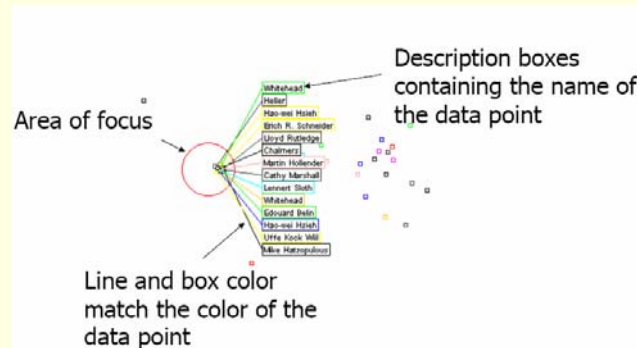
- Dynamic Visualization of Graphs with Extended Labels [Wong et al. Infovis 2005]

video

80

Enhanced Presentation of Labels

- Excentric Labeling [Fekete and Plaisant CHI '99]



81

Concepts and Relationships in Individual Document

- TOPIC ISLANDSTM – A Wavelet-Based Text Visualization System [Miller Vis' 98]
 - Construct digital signals from words within a document
 - Apply wavelet transforms to the signals
 - Analyze narrative flow using resultant wavelet energy
 - Use MDS to map themes

82

Topic Islands [Miller Vis 98]

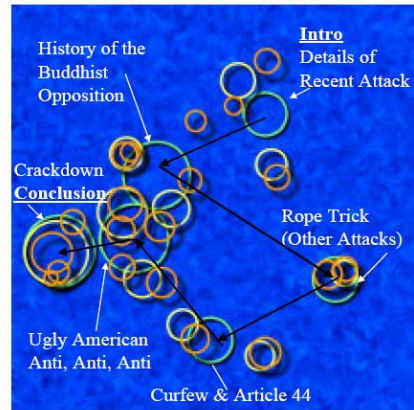


Figure 14: The top view of the islands for the Buddhist Article shows the interplay of theme map. See also CP 8.

83

Topic Islands [Miller Vis 98]

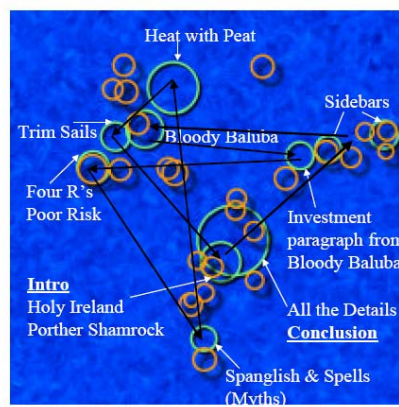


Figure 13: The top view of the islands for the Ireland article show the interplay of theme map and the narrative. See CP7.

84

Document Collections

- Problem or challenge is how to present the contents/semantics/themes/etc of the documents to someone who does not have time to read them all
- Who cares?
 - Researchers, news people,...

85

Improving Text Searches

- What's wrong with the common search?
 - Query responses do not include include:
 - How strong the match is
 - How frequent each term is
 - How each term is distributed in the document
 - Overlap between terms
 - Length of document
 - Document ranking is opaque
 - Inability to compare between results
 - Input limits term relationships

86

TileBars [Hearst CHI'95]

■ Goal

- Minimize time and effort for deciding which documents to view in detail

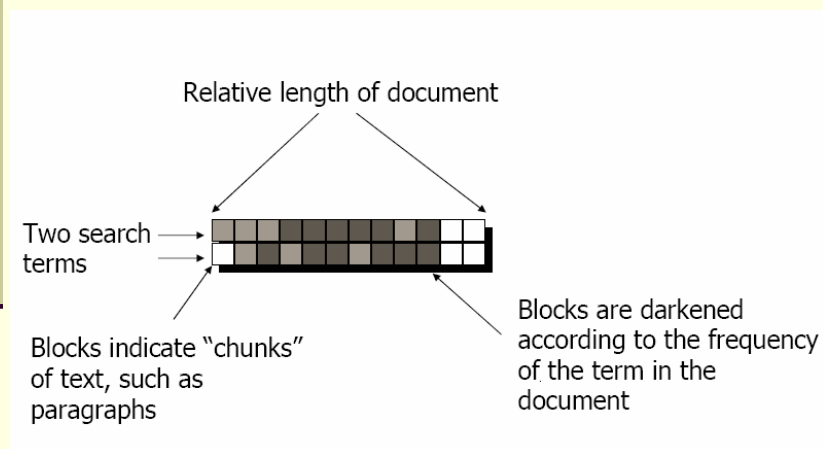
■ Idea

- Show the role of the query terms in the retrieved documents, making use of document structure

87

TileBars [Hearst CHI'95]

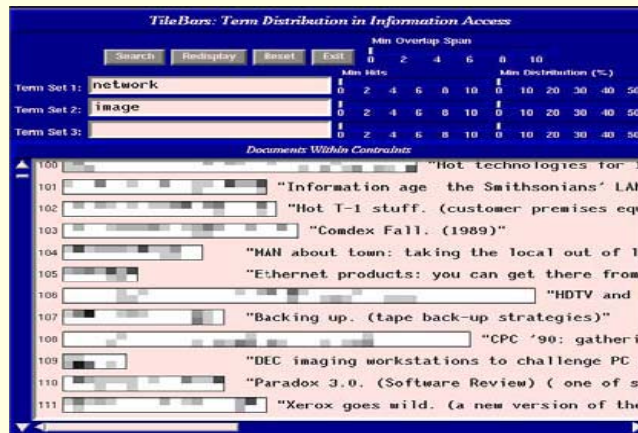
■ Techniques



88

TileBars [Hearst CHI'95]

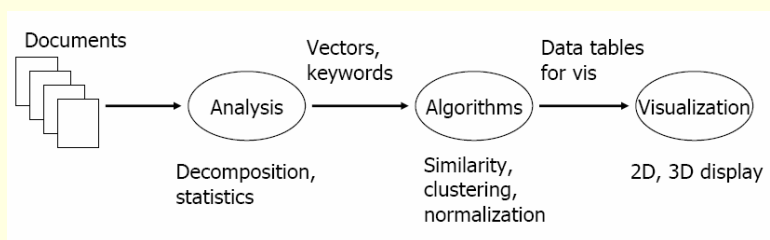
■ Interface



■ <http://elib.cs.berkeley.edu/tilebars/about.html#using>

89

More Complex Process



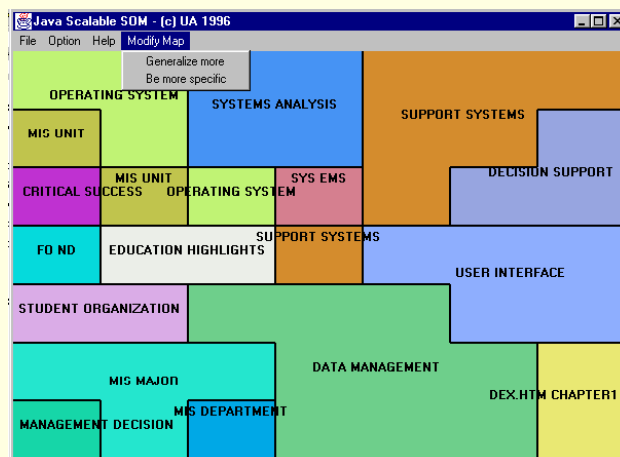
90

Visualizing Documents

- Break each document into its words
- Two documents are “similar” if they share many words
- Use algorithm for clustering similar documents together and dissimilar documents far apart

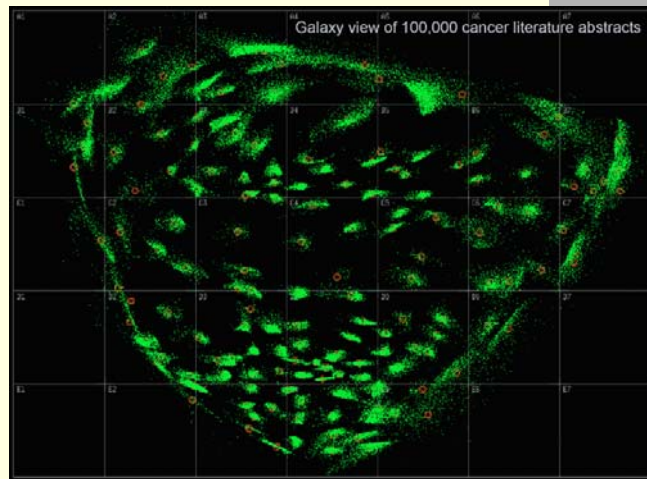
91

Use SOM Map



92

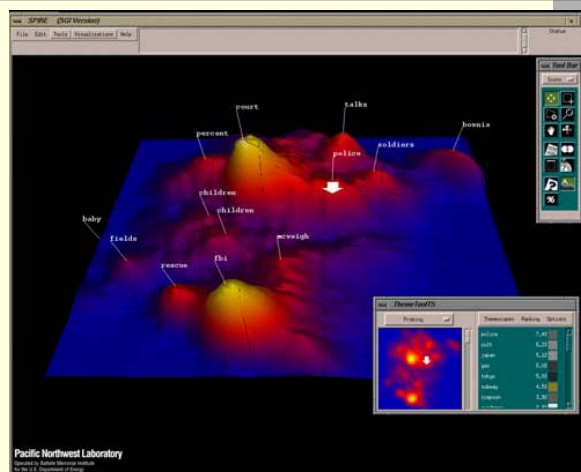
Galaxy



Galaxy of PNNL

93

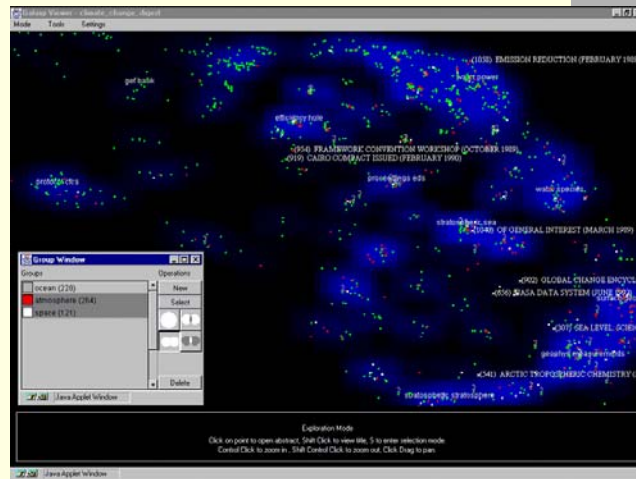
ThemeScape



Themescape of PNNL

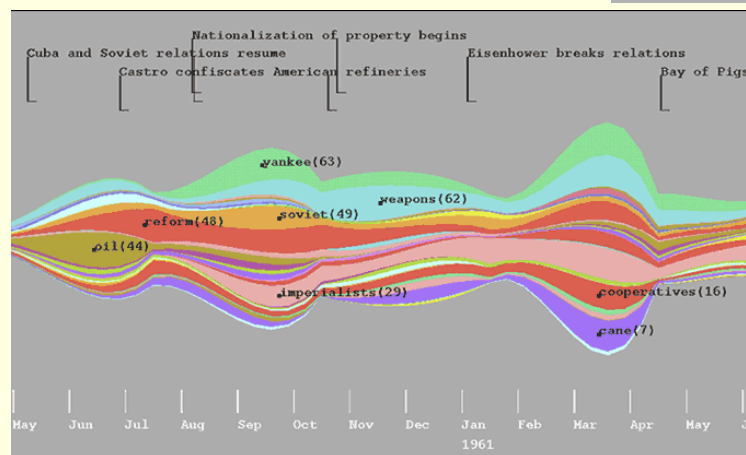
94

WebTheme



95

ThemeRiver



96