Topic Notes

Text and Document Visualization 2

CS 7450 - Information Visualization

October 31, 2012

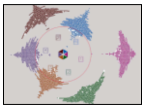
John Stasko

Example Tasks & Goals

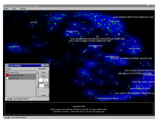
- Which documents contain text on topic XYZ?
- Which documents are of interest to me?
- Are there other documents that are similar to this one (so they are worthwhile)?
- How are different words used in a document or a document collection?
- What are the main themes and ideas in a document or a collection?
- Which documents have an angry tone?
- How are certain words or themes distributed through a document?
- Identify "hidden" messages or stories in this document collection.
- How does one set of documents differ from another set?
- Quickly gain an understanding of a document or collection in order to subsequently do XYZ.
- Find connections between documents.

This Week’s Agenda

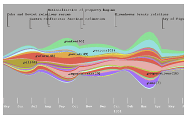
Visualization for IR
Helping search



Visualizing text
Showing words, phrases, and sentences



Visualizing document sets
Words & sentences
Analysis metrics
Concepts & themes



Last Time

Recall

Related Topic - Sensemaking

- Sensemaking
  - Gaining a better understanding of the facts at hand in order to take some next steps
  - (Better definitions in VA lecture)
- InfoVis can help make a large document collection more understandable more rapidly

Today’s Agenda

- Move to collections of documents
  - Still do words, phrases, sentences
  - Add
    - More context of documents
    - Document analysis metrics
    - Document meta-data
    - Document entities
    - Connections between documents
    - Documents concepts and themes

Various Document Metrics

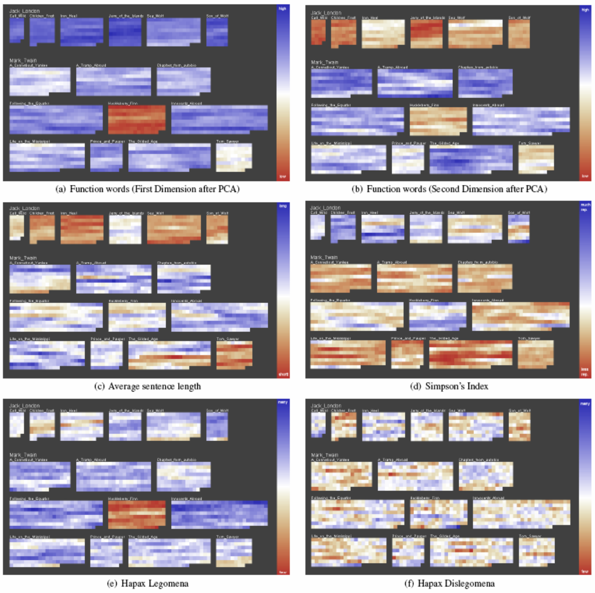
- Goals?
- Different variables for literary analysis
  - Average word length
  - Syllables per word
  - Average sentence length
  - Percentage of nouns, verbs, adjectives
  - Frequencies of specific words
  - Hapax Legomena – number of words that occur once

Keim & Oelke
VAST ‘07

Vis

Each block represents a contiguous set of words, eg, 10,000 words

Do partial overlap in blocks for a smoother appearance



(a) Function words (First Dimension after PCA)

(b) Function words (Second Dimension after PCA)

(c) Average sentence length

(d) Simpson's Index

(e) Hapax Legomena

(f) Hapax Dislegomena

Figure 2: Fingerprints of books of Mark Twain and Jack London. Different measures for authorship attribution are tested. If a measure is able to discriminate between the two authors, the visualizations of the books that are written by the same author will equal each other more than the visualizations of books written by different authors. It can easily be seen that this is not true for every measure (e.g. Hapax Dislegomena). Furthermore, it is interesting to observe that the book *Huckleberry Finn* sticks out in a number of measures as if it is not written by Mark Twain.

The Bible

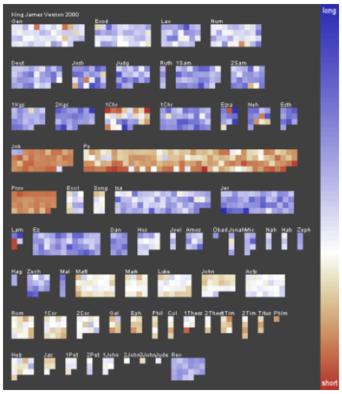


Figure 4: Visual Fingerprint of the Bible. Each pixel represents one chapter of the bible and color is mapped to the average verse length. Interesting characteristics such as the generally shorther verses of the poetry books, the inhomogeneity of the 1. Book of Chronicles or the difference between the Old Testament and the New Testament can be perceived.

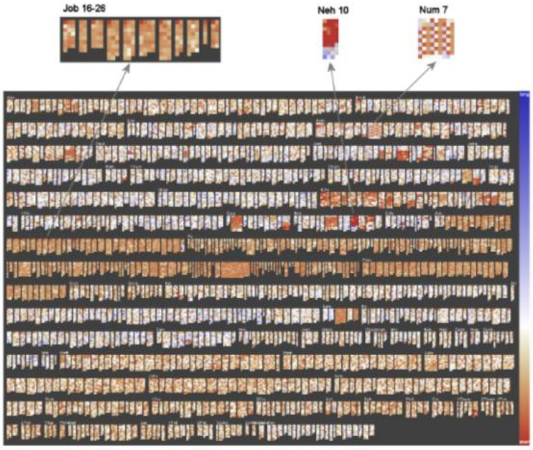


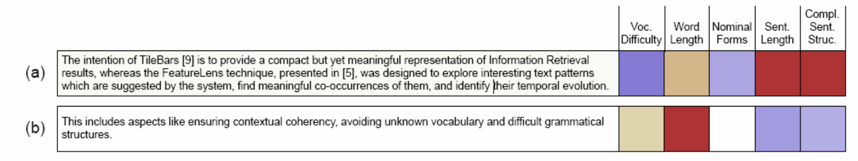
Figure 5: Visual Fingerprint of the Bible. More detailed view on the bible in which each pixel represents a single verse and verses are grouped to chapters. Color is again mapped to verse length. The detailed view reveals some interesting patterns that are camouflaged in the averaged version of fig. 4.

Follow-On Work

- Focus on readability metrics of documents
- Multiple measures of readability
  - Provide quantitative measures
- Features used:
  - Word length
  - Vocabulary complexity
  - Nominal forms
  - Sentence length
  - Sentence structure complexity

Oelke & Keim
VAST '10

Visualization & Metrics



		Voc. Difficulty	Word Length	Nominal Forms	Sent. Length	Compl. Sent. Struc.
(a)	The intention of TileBars [9] is to provide a compact but yet meaningful representation of Information Retrieval results, whereas the FeatureLens technique, presented in [5], was designed to explore interesting text patterns which are suggested by the system, find meaningful co-occurrences of them, and identify their temporal evolution.	🟪	🟫	🟪	🟥	🟥
(b)	This includes aspects like ensuring contextual coherency, avoiding unknown vocabulary and difficult grammatical structures.	🟫	🟥	⬜	🟪	🟪

Figure 5: Two example sentences whose overall readability score is about the same. The detail view reveals the different reasons why the sentences are difficult to read.

Uses heatmap style vis (blue-readable, red-unreadable)

Interface

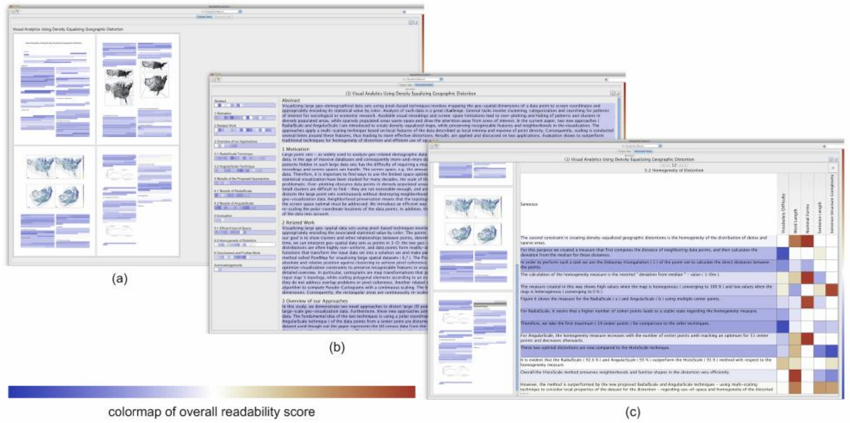


Figure 3: Screenshot of the VisRA tool on 3 different aggregation levels. (a) Corpus View (b) Block View (c) Detail View. To display single features, the colormap is generated as described in section 3.4 and figure 2.

Their Paper (Before & After)

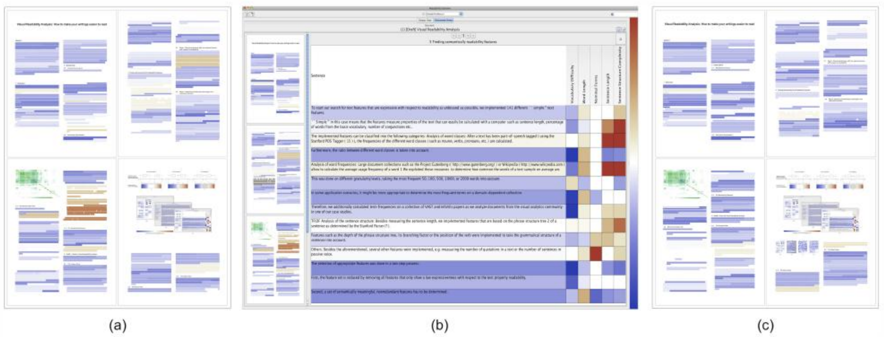


Figure 6: Revision of our own paper. (a) The first four pages of the paper as structure thumbnails before the revision. (b) Detail view for one of the sections. (c) Structure thumbnails of the same pages after the revision.

Comment from the Talk

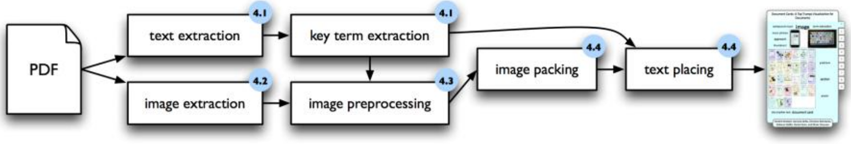
- In academic papers, you want your abstract to be really readable
- Would be cool to compare rejected papers to accepted papers

Overviews of Documents

- Can we provide a quick browsing, overview UI, maybe especially useful for small screens?

Document Cards

- Compact visual representation of a document
- Show key terms and important images



PDF

text extraction 4.1

key term extraction 4.1

image extraction 4.2

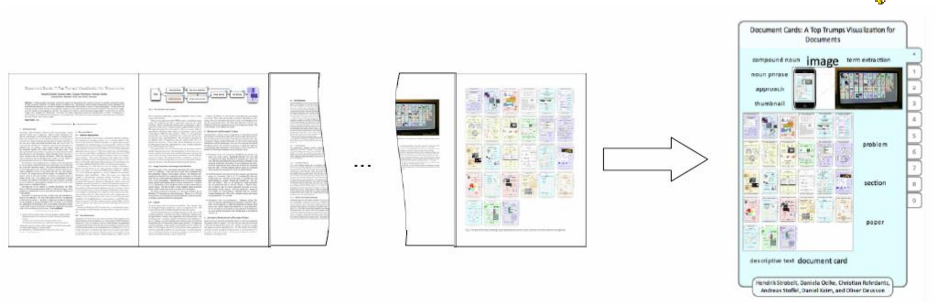
image preprocessing 4.3

image packing 4.4

text placing 4.4

Strobelt et al
TVCG (InfoVis) '09

Representation

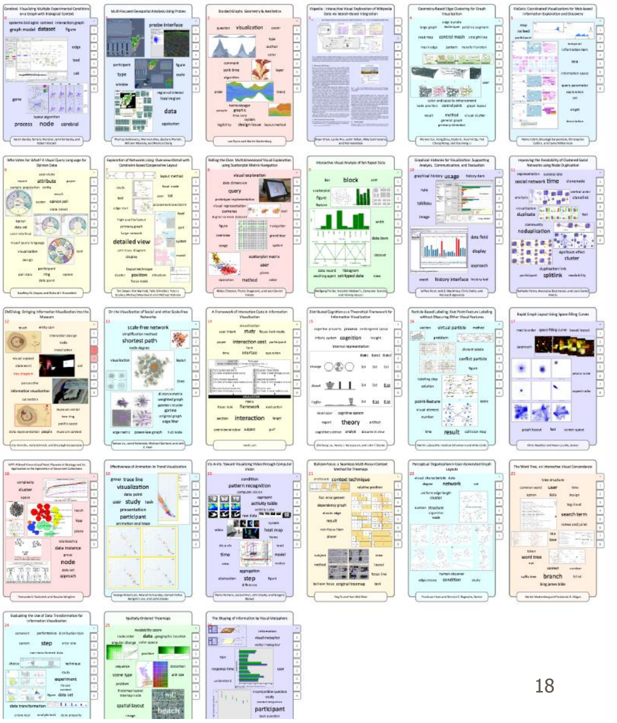

...
Document Cards: A Top Trumps Visualization for Documents
compound noun
image
term extraction
noun phrase
approach
thumbnail
problem
section
paper
descriptive text
document card
Hendrik Strobelt, Daniela Oelke, Christian Rohrdantz, Andreas Stoffel, Daniel Keim, and Oliver Deussen

Layout algorithm searches for empty space rectangles to put things

Interaction

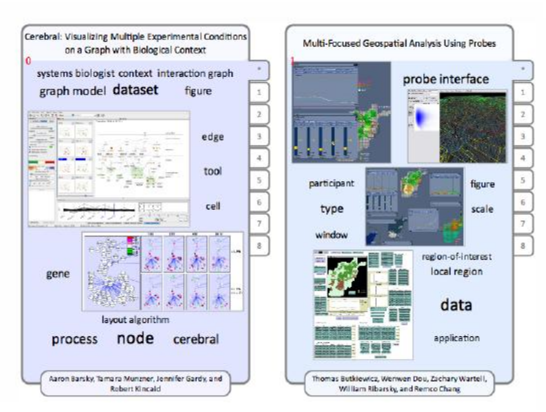
- Hover over non-image space shows abstract in tooltip
- Hover over image and see caption as tooltip
- Click on page number to get full page
- Click on image goes to page containing it
- Clicking on a term highlights it in overview and all tooltips

InfoVis '08 Proceedings



Video

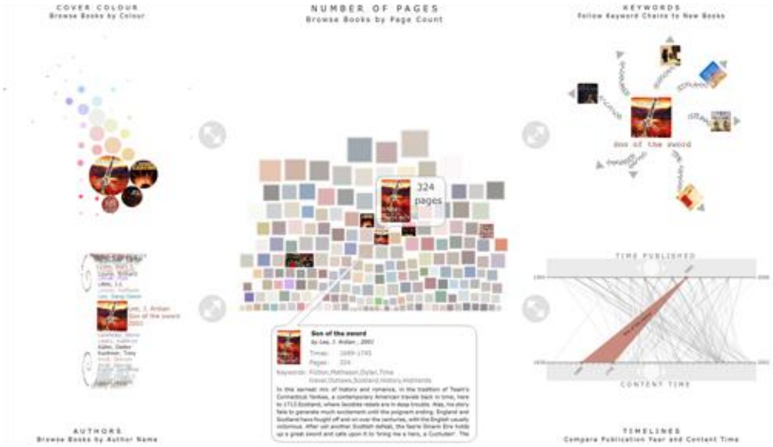
Zooming In


Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context
systems biologist context interaction graph
graph model dataset figure
edge
tool
cell
gene
layout algorithm
process node cerebral
Aaron Barsky, Tamara Munzner, Jennifer Gardy, and Robert Kincaid
Multi-Focused Geospatial Analysis Using Probes
probe interface
participant
type
window
figure
scale
region-of-interest
local region
data
application
Thomas Butkiewicz, Wenwen Dou, Zachary Wartell, William Ribarsky, and Remco Chang
Who Votes For What? A Visual Query Language for Opinion Data
user study
report attribute paper
sample population entity
result
sector opinion poll
state street
typical
data set
user interface
visual query language
visualization
task
design
participant
poll data ring system
data point
Geoffrey M. Draper, and Richard F. Riesenfeld
1
2
3
4
5
6
7
8

Bohemian Bookshelf

Video

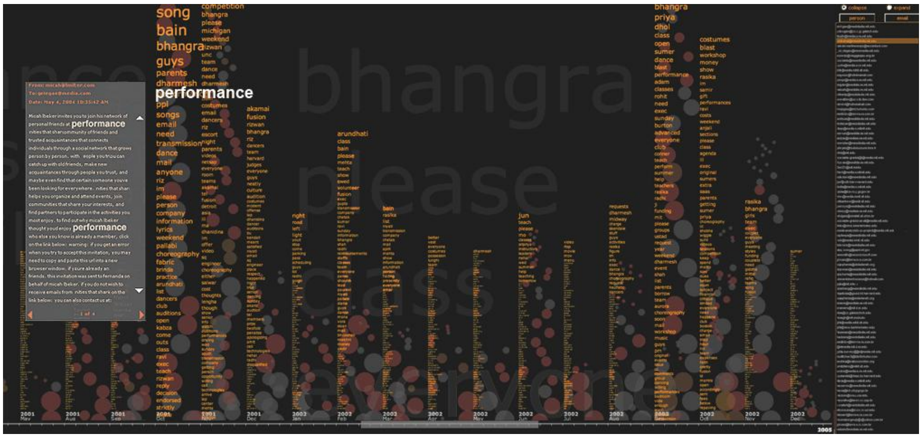
Serendipitous browsing


COVER COLOUR
Browse Books by Colour
NUMBER OF PAGES
Browse Books by Page Count
KEYWORDS
Follow Keyword Chains to New Books
Son of the sword
324 pages
TIME PUBLISHED
CONTENT TIME
AUTHORS
Browse Books by Author Name
TIMELINES
Compare Publication Year and Content Time

Thudt et al
CHI ‘12

Themail

More to come on
Social Visualization day



Collection of email

Viegas et al
CHI '06

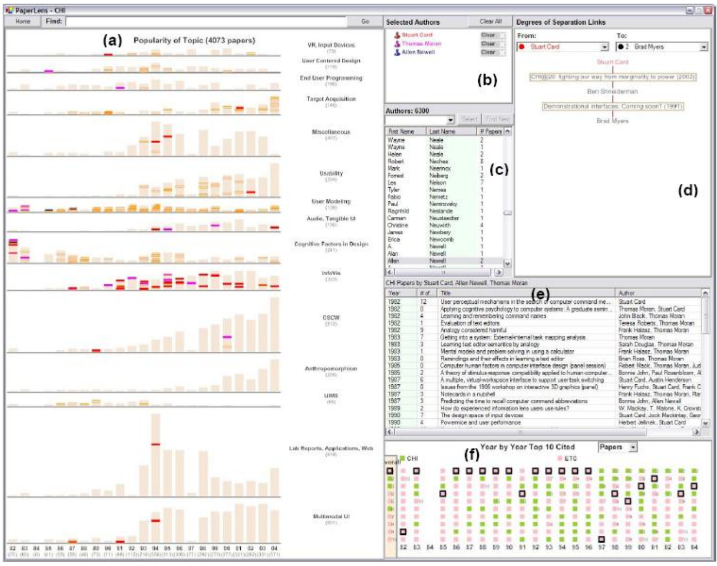
PaperLens

- Focus on academic papers
- Visualize doc metadata such as author, keywords, date, …
- Multiple tightly-coupled views
- Analytics questions
- Effective in answering questions regarding:
  - Patterns such as frequency of authors and papers cited
  - Themes
  - Trends such as number of papers published in a topic area over time
  - Correlations between authors, topics and citations

Lee et al
CHI '05 Short

PaperLens

Video



a) Popularity of topic
b) Selected authors
c) Author list
d) Degrees of separation of links
e) Paper list
f) Year-by-year top ten cited papers/ authors – can be sorted by topic

NetLens

Kang et al
Information Visualization '07

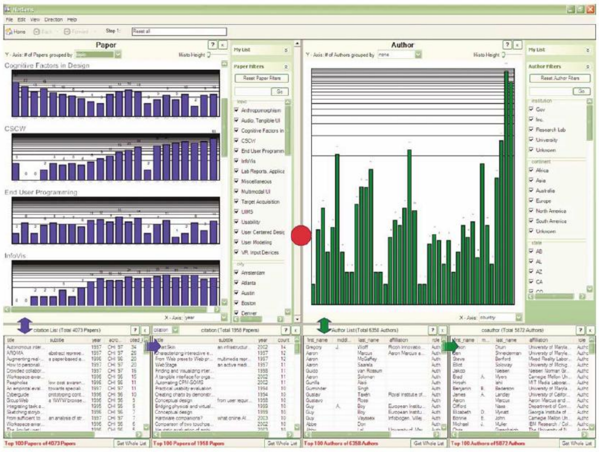


Figure 1 NetLens has two symmetric windows. The left is for Content (papers) and the right for Actors (authors). Each side is further divided into panels; overview at the top, filters on the right, and lists at the bottom. Here, the Content side has two lists to reflect papers and their citations or references, and the lists on the Actor side show authors and their co-authors, respectively. The paper overview panel shows the distribution of papers (in logarithmic scale) over time, grouped by topics. Users can see which topics have their number of papers increase or decrease over 22 years. On the right side, the overview of the authors shows the distribution of countries of origin in logarithmic scale.

More Document Info

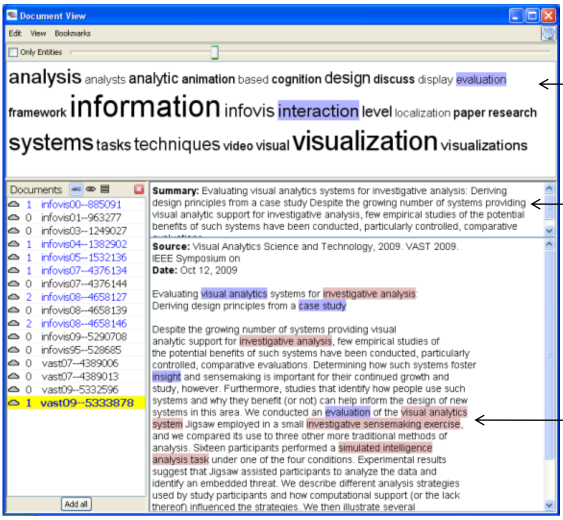
- Highlight entities within documents
  - People, places, organizations
- Document summaries
- Document similarity and clustering
- Document sentiment

Jigsaw

- Targeting sense-making scenarios
- Variety of visualizations ranging from word-specific, to entity connections, to document clusters
- Primary focus is on entity-document and entity-entity connection
- Search capability coupled with interactive exploration

Stasko, Görg, & Liu
Information Visualization '08

Document View



analysis analysts analytic animation based cognition design discuss display evaluation framework information infovis interaction level localization paper research systems tasks techniques video visual visualization visualizations

1 infovis00--885091
0 infovis01--963277
0 infovis03--1249027
1 infovis04--1382902
1 infovis05--1532136
1 infovis07--4376134
0 infovis07--4376144
2 infovis08--4658127
0 infovis08--4658139
2 infovis08--4658146
0 infovis09--5290708
0 infovis95--528685
0 vast07--4389006
0 vast07--4389013
0 vast09--5332596
1 vast09--5333878

Summary: Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study Despite the growing number of systems providing visual analytic support for investigative analysis, few empirical studies of the potential benefits of such systems have been conducted, particularly controlled, comparative

Source: Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on
Date: Oct 12, 2009

Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study

Despite the growing number of systems providing visual analytic support for investigative analysis, few empirical studies of the potential benefits of such systems have been conducted, particularly controlled, comparative evaluations. Determining how such systems foster insight and sensemaking is important for their continued growth and study, however. Furthermore, studies that identify how people use such systems and why they benefit (or not) can help inform the design of new systems in this area. We conducted an evaluation of the visual analytics system Jigsaw employed in a small investigative sensemaking exercise, and we compared its use to three other more traditional methods of analysis. Sixteen participants performed a simulated intelligence analysis task under one of the four conditions. Experimental results suggest that Jigsaw assisted participants to analyze the data and identify an embedded threat. We describe different analysis strategies used by study participants and how computational support (or the lack thereof) influenced the strategies. We then illustrate several

Wordcloud overview

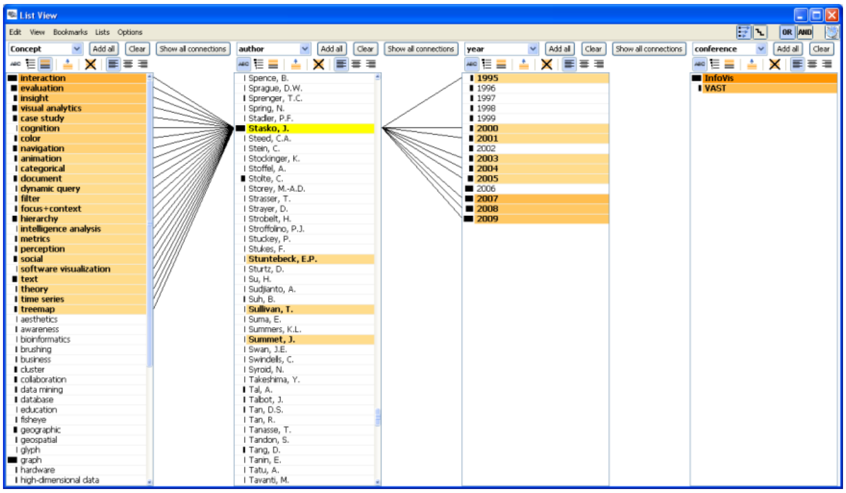
Doc List

Document summary

Selected document's text with entities identified

List View

Entities listed by type



Concept

interaction
evaluation
insight
visual analytics
case study
cognition
color
navigation
animation
categorical
document
dynamic query
filter
focus+context
hierarchy
intelligence analysis
metrics
perception
social
software visualization
text
theory
time series
treemap
aesthetics
awareness
bioinformatics
brushing
business
cluster
collaboration
data mining
database
education
fisheye
geographic
geospatial
glyph
graph
hardware
high-dimensional data

author

Spence, B.
Sprague, D.W.
Sprenger, T.C.
Spring, N.
Stadler, P.F.
Stasko, J.
Steed, C.A.
Stein, C.
Stockinger, K.
Stoffel, A.
Stolte, C.
Storey, M.-A.D.
Strasser, T.
Strayer, D.
Strobelt, H.
Stroffolino, P.J.
Stuckey, P.
Stukes, F.
Stuntebeck, E.P.
Sturtz, D.
Su, H.
Sudjianto, A.
Suh, B.
Sullivan, T.
Suma, E.
Summers, K.L.
Summet, J.
Swan, J.E.
Swindells, C.
Syroid, N.
Takeshima, Y.
Tal, A.
Talbot, J.
Tan, D.S.
Tan, R.
Tanasse, T.
Tandon, S.
Tang, D.
Tanin, E.
Tatu, A.
Tavanti, M.

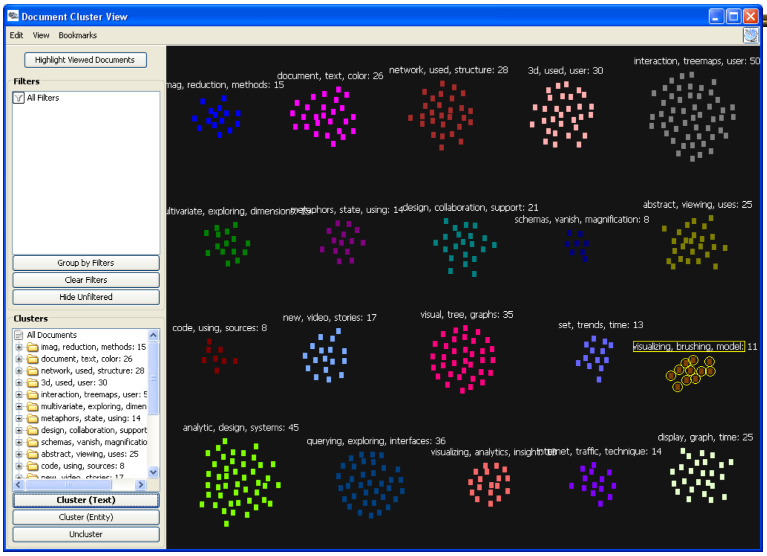
year

1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009

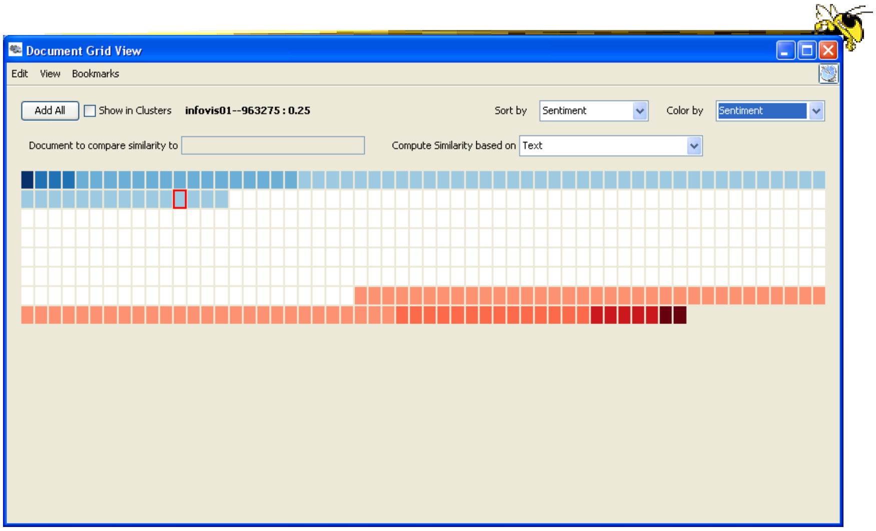
conference

InfoVis
VAST

Document Cluster View


Document Cluster View
Edit View Bookmarks
Highlight Viewed Documents
Filters
All Filters
Group by Filters
Clear Filters
Hide Unfiltered
Clusters
All Documents
imag, reduction, methods: 15
document, text, color: 26
network, used, structure: 28
3d, used, user: 30
interaction, treemaps, user: 5
multivariate, exploring, dimen
metaphors, state, using: 14
design, collaboration, support
schemas, vanish, magnificatio
abstract, viewing, uses: 25
code, using, sources: 8
new, video, stories: 17
Cluster (Text)
Cluster (Entity)
Uncluster
mag, reduction, methods: 15
document, text, color: 26
network, used, structure: 28
3d, used, user: 30
interaction, treemaps, user: 50
ultivariate, exploring, dimensions: 18
metaphors, state, using: 14
design, collaboration, support: 21
schemas, vanish, magnification: 8
abstract, viewing, uses: 25
code, using, sources: 8
new, video, stories: 17
visual, tree, graphs: 35
set, trends, time: 13
visualizing, brushing, model: 11
analytic, design, systems: 45
querying, exploring, interfaces: 36
visualizing, analytics, insight: 18
internet, traffic, technique: 14
display, graph, time: 25

Document Grid View



infovis01--963275 : 0.25

Sort by Sentiment

Color by Sentiment

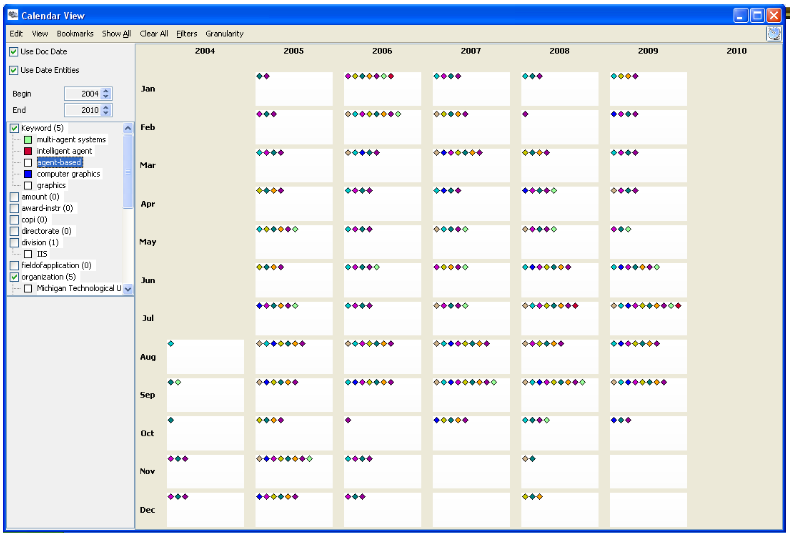
Document to compare similarity to

Compute Similarity based on Text

Here showing sentiment analysis of docs

Calendar View

Temporal context of entities & docs



☑ Use Doc Date

☑ Use Date Entities

Begin 2004

End 2010

- ☑ Keyword (5)
  - 🟩 multi-agent systems
  - 🟥 intelligent agent
  - □ agent-based
  - 🟦 computer graphics
  - □ graphics
- □ amount (0)
- □ award-instr (0)
- □ copi (0)
- □ directorate (0)
- □ division (1)
  - □ IIS
- □ fieldofapplication (0)
- ☑ organization (5)
  - □ Michigan Technological U

Video

Jigsaw

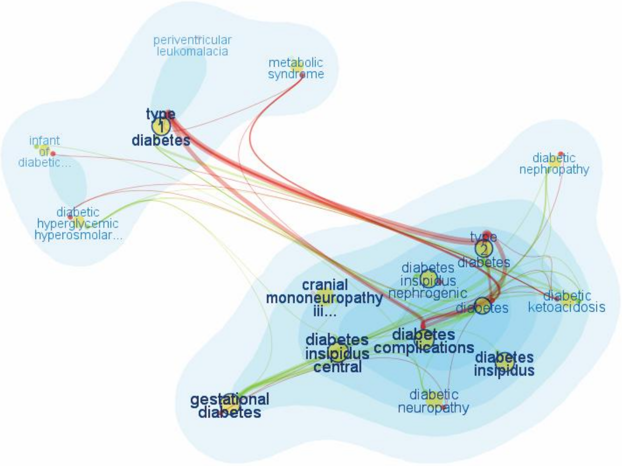
- Much more to come on Visual Analytics day…

FacetAtlas

- Show entities and concepts and how they connect in a document collection
- Visualizes both local and global patters
- Shows
  - Entities
  - Facets – classes of entities
  - Relations – connections between entities
  - Clusters – groups of similar entities in a facet

Cao et al
TVCG (InfoVis) '10

Visualization



Video

Up to Higher Level

- How do we present the contents, semantics, themes, etc of the documents
  - Someone may not have time to read them all
  - Someone just wants to understand them
- Who cares?
  - Researchers, fraud investigators, CIA, news reporters

Vector Space Analysis

- How does one compare the similarity of two documents?
- One model
  - Make list of each unique word in document
    - Throw out common words (a, an, the, …)
    - Make different forms the same (bake, bakes, baked)
  - Store count of how many times each word appeared
  - Alphabetize, make into a vector

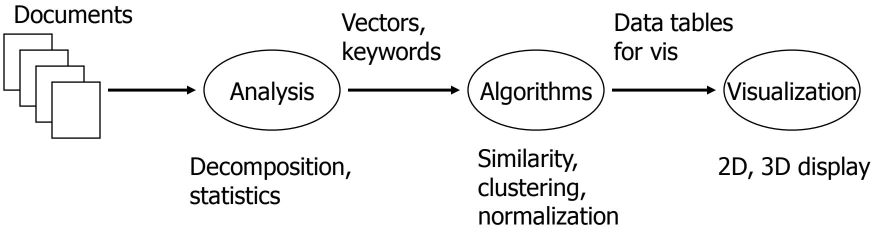
Vector Space Analysis

- Model (continued)
  - Want to see how closely two vectors go in same direction, inner product
  - Can get similarity of each document to every other one
  - Use a mass-spring layout algorithm to position representations of each document
- Some similarities to how search engines work

Wiggle

- Not all terms or words are equally useful
- Often apply TFIDF
  - Term frequency, inverse document frequency
- Weight of a word goes up if it appears often in a document, but not often in the collection

Process



Documents

Analysis

Decomposition, statistics

Vectors, keywords

Algorithms

Similarity, clustering, normalization

Data tables for vis

Visualization

2D, 3D display

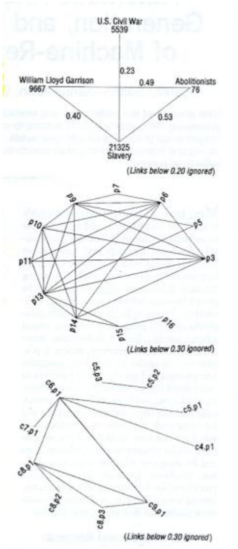
Smart System

- Uses vector space model for documents
  - May break document into chapters and sections and deal with those as atoms
- Plot document atoms on circumference of circle
- Draw line between items if their similarity exceeds some threshold value

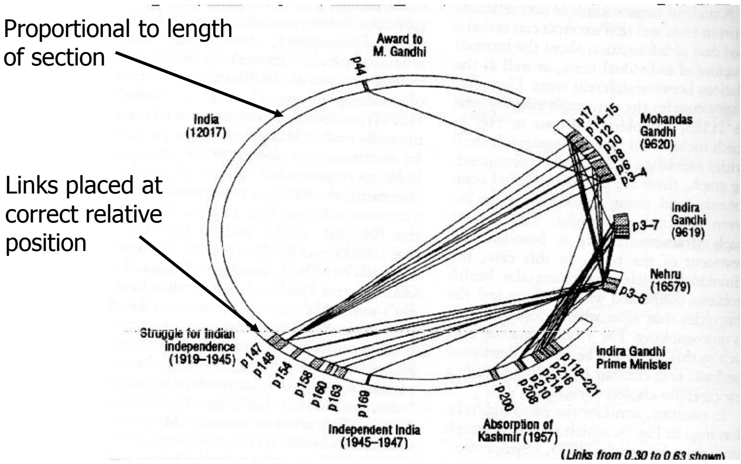
Salton et al
Science '95

Text Relation Maps

- Label on line can indicate similarity value
- Items evenly spaced
- Doesn't give viewer idea of how big each section/document is



Improved Design


Proportional to length of section

Links placed at correct relative position

Text Themes

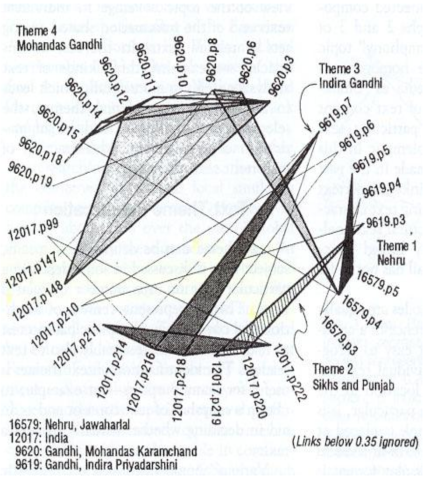
- Look for sets of regions in a document (or sets of documents) that all have common theme
  - Closely related to each other, but different from rest
- Need to run clustering process

Algorithm

- Recognize triangles in relation maps
  - Three with edges above threshold
- Make a new vector that is centroid of 3
- Triangles merged whenever centroid vectors are sufficiently similar

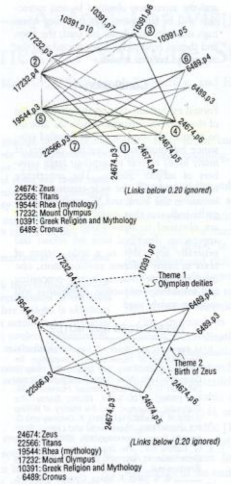
Text Theme Example

- Triangles shown
- Colored in to help presentation



Skimming and Summarization

- Can use graph traversal to follow specific themes throughout collection
- Walk along connected edges



VIBE System

- Smaller sets of documents than whole library
- Example: Set of 100 documents retrieved from a web search
- Idea is to understand contents of documents relate to each other

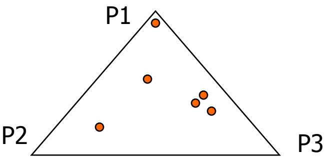
Olsen et al
Info Process & Mgmt '93

Focus

- Points of Interest
  - Terms or keywords that are of interest to user
    - Example: cooking, pies, apples
- Want to visualize a document collection where each document's relation to points of interest is show
- Also visualize how documents are similar or different

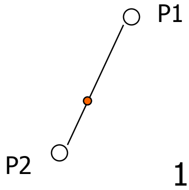
Technique

- Represent points of interest as vertices on convex polygon
- Documents are small points inside the polygon
- How close a point is to a vertex represents how strong that term is within the document



Algorithm

- Example: 3 POIs
- Document (P1, P2, P3) (0.4, 0.8, 0.2)
- Take first two

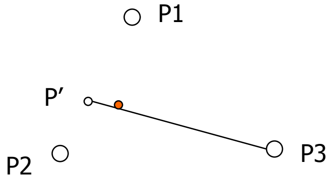


$$\frac{0.4}{0.4+0.8} = 0.333$$

1/3 of way from P2 to P1

Algorithm

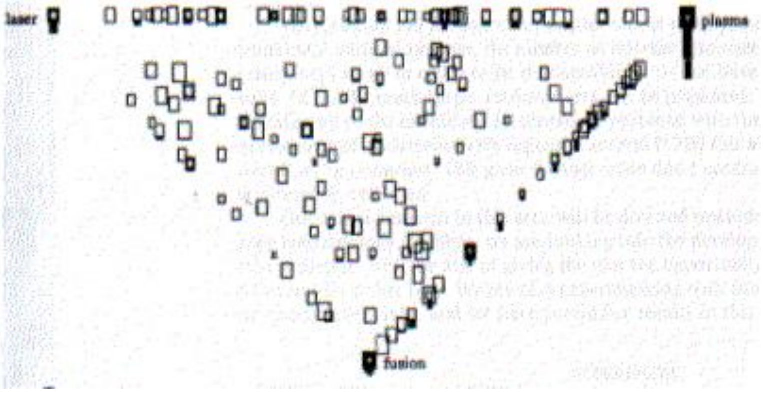
- Combine weight of first two 1.2 and make a new point, P’
- Do same thing for third point



$$\frac{1.2}{1.2+0.2} = 0.86$$

0.14 of way from P’ to P3

Sample Visualization



VIBE Pro's and Con's

- Effectively communications relationships
- Straightforward methodology and vis are easy to follow
- Can show relatively large collections

- Not showing much about a document
- Single items lose "detail" in the presentation
- Starts to break down with large number of terms

Visualizing Documents

- VIBE presented documents with respect to a finite number of special terms
- How about generalizing this?
  - Show large set of documents
  - Any important terms within the set become key landmarks
  - Not restricted to convex polygon idea

Basic Idea

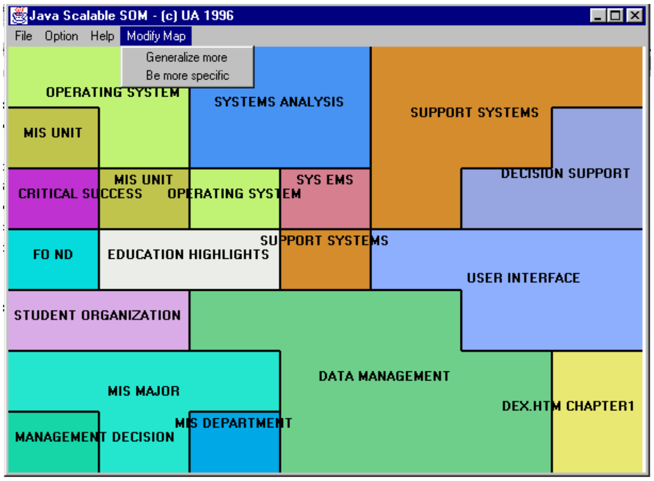
- Break each document into its words
- Two documents are "similar" if they share many words
- Use mass-spring graph-like algorithm for clustering similar documents together and dissimilar documents far apart

Kohonen's Feature Maps

- AKA Self-Organizing Maps
- Expresses complex, non-linear relationships between high dimensional data items into simple geometric relationships on a 2-d display
- Uses neural network techniques

Lin
Visualization '92

Map Display of SOM



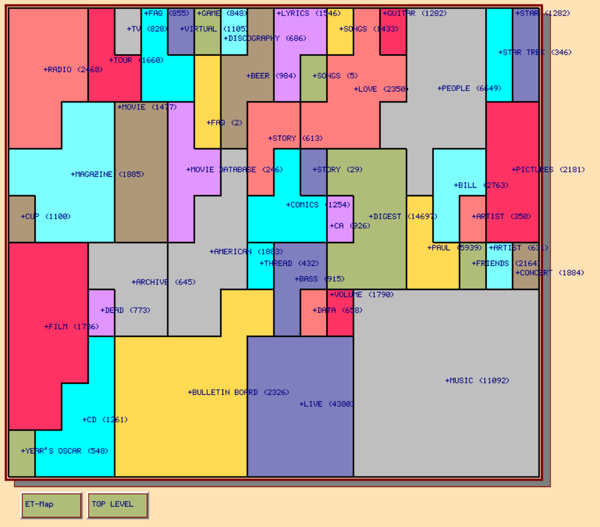
Map Attributes

- Different, colored areas correspond to different concepts in collection
- Size of area corresponds to its relative importance in set
- Neighboring regions indicate commonalities in concepts
- Dots in regions can represent documents

More Maps

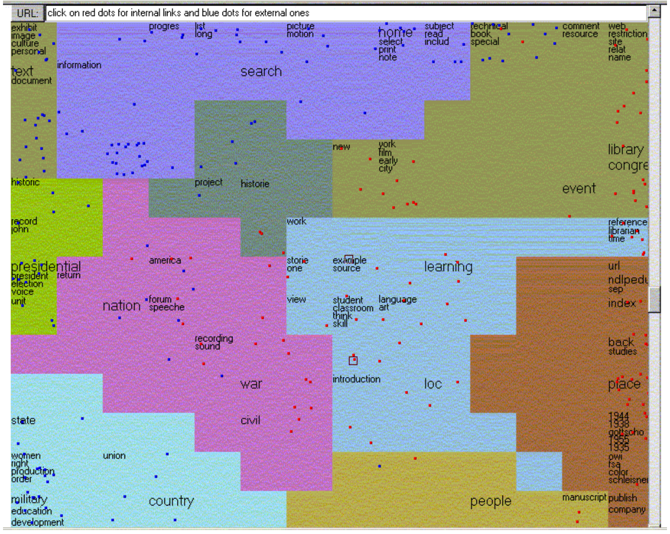
ET-Map

Artificial Intelligence Lab, University of Arizona.



More Maps

Interactive demos


URL: click on red dots for internal links and blue dots for external ones

Xia Lin

Work at PNNL

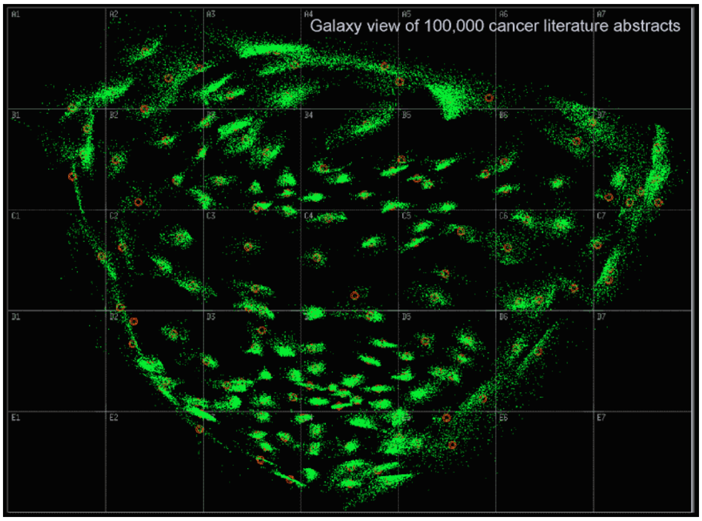
http://www.pnl.gov/infoviz

- Group has developed a number of visualization techniques for document collections
  - Galaxies
  - Themescapes
  - ThemeRiver
  - …

Wise et al
InfoVis '95

Galaxies

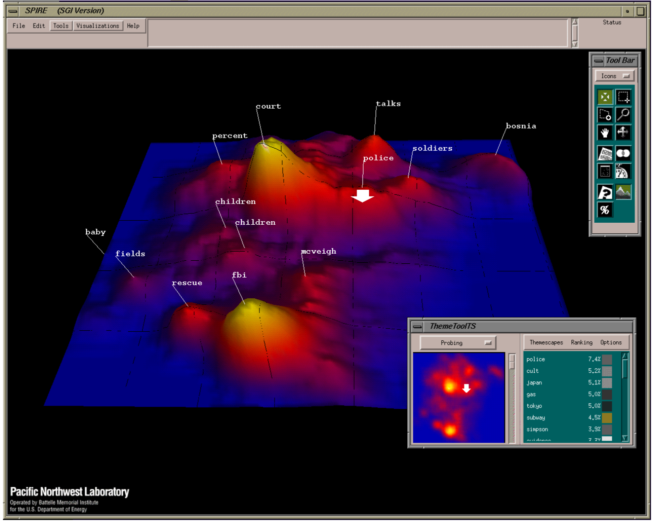
Presentation of documents where similar ones cluster together



Themescapes

- Self-organizing maps didn’t reflect density of regions all that well -- Can we improve?
- Use 3D representation, and have height represent density or number of documents in region

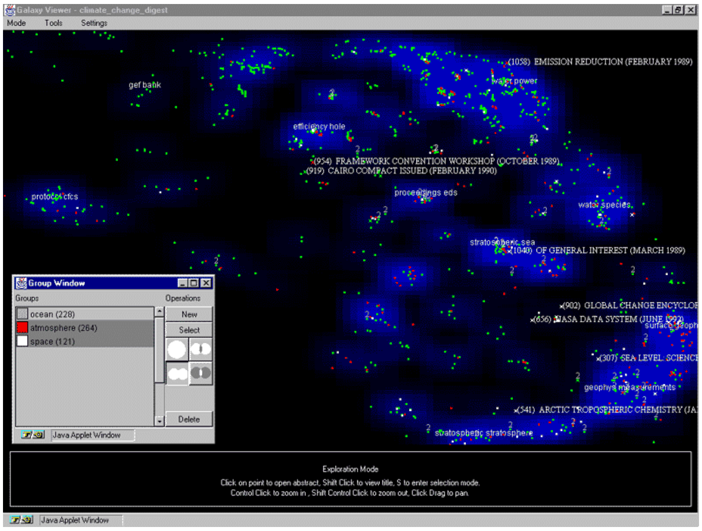
Themescape



police	7.4%
cult	5.2%
japan	5.1%
gas	5.0%
tokyo	5.0%
subway	4.5%
simpson	3.9%

Video

WebTheme


Galaxy Viewer - climate_change_digest
Mode
Tools
Settings
(1058) EMISSION REDUCTION (FEBRUARY 1989)
water power
gef bank
efficiency hole
(954) FRAMEWORK CONVENTION WORKSHOP (OCTOBER 1989)
(919) CAIRO COMPACT ISSUED (FEBRUARY 1990)
protocol cfcs
proceedings eds
water species
stratospheric sea
(1040) OF GENERAL INTEREST (MARCH 1989)
(902) GLOBAL CHANGE ENCYCLOP
(656) NASA DATA SYSTEM (JUNE 1992)
surface geoph
(307) SEA LEVEL SCIENC
geophys measurements
(541) ARCTIC TROPOSPHERIC CHEMISTRY (JA
stratospheric stratosphere
Group Window
Groups
Operations
New
Select
Delete
Java Applet Window
Java Applet Window

- ▨ ocean (228)
- 🟥 atmosphere (264)
- □ space (121)

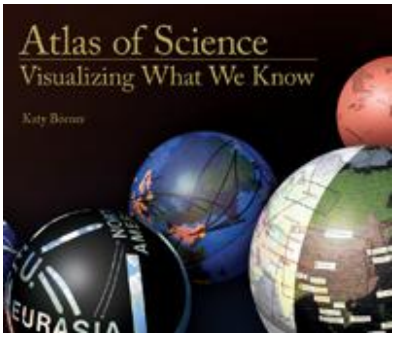
Exploration Mode

Click on point to open abstract, Shift Click to view title, S to enter selection mode.
Control Click to zoom in , Shift Control Click to zoom out, Click Drag to pan.

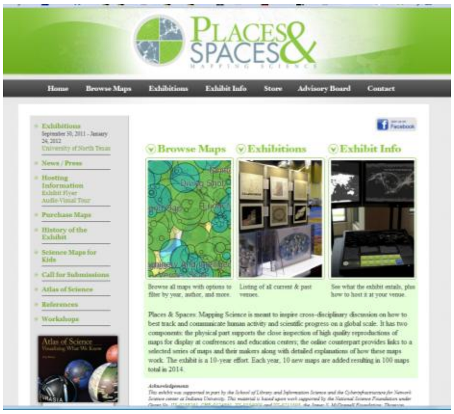
Related Topic

- Maps of Science
- Visualize the relationships of areas of science, emerging research disciplines, the impact of particular researchers or institutions, etc.
- Often use documents as the “input data”

Wonderful Book and Website



K. Börner



- Exhibitions
  September 30, 2011 - January 24, 2012
  University of North Texas
- News / Press
- Hosting Information
  Exhibit Flyer
  Audio-Visual Tour
- Purchase Maps
- History of the Exhibit
- Science Maps for Kids
- Call for Submissions
- Atlas of Science
- References
- Workshops

Browse Maps — Browse all maps with options to filter by year, author, and more.

Exhibitions — Listing of all current & past venues.

Exhibit Info — See what the exhibit entails, plus how to host it at your venue.

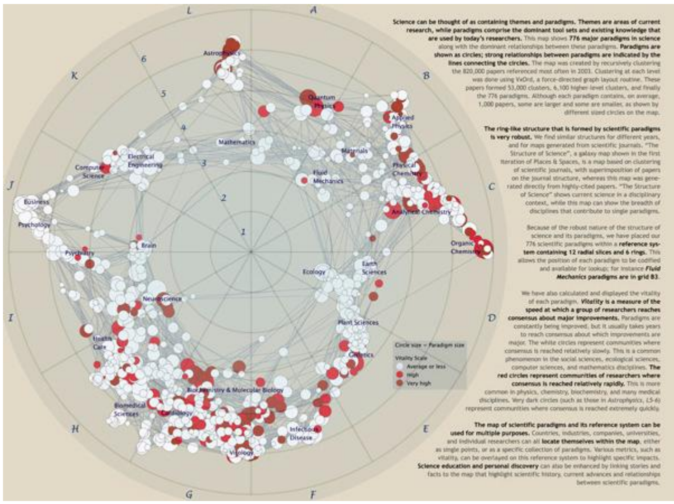
Places & Spaces: Mapping Science is meant to inspire cross-disciplinary discussion on how to best track and communicate human activity and scientific progress on a global scale. It has two components: the physical part supports the close inspection of high quality reproductions of maps for display at conferences and education centers; the online counterpart provides links to a selected series of maps and their makers along with detailed explanations of how these maps work. The exhibit is a 10-year effort. Each year, 10 new maps are added resulting in 100 maps total in 2014.

Acknowledgements

This exhibit was supported in part by the School of Library and Information Science and the Cyberinfrastructure for Network Science center at Indiana University. This material is based upon work supported by the National Science Foundation under Grant No. [illegible] the James S. McDonnell Foundation, Thomson

http://scimaps.org

Some Examples



Science can be thought of as containing themes and paradigms. Themes are areas of current research, while paradigms comprise the dominant tool sets and existing knowledge that are used by today's researchers. This map shows **776 major paradigms in science** along with the dominant relationships between these paradigms. **Paradigms are shown as circles; strong relationships between paradigms are indicated by the lines connecting the circles.** The map was created by recursively clustering the 820,000 papers referenced most often in 2003. Clustering at each level was done using VxOrd, a force-directed graph layout routine. These papers formed 53,000 clusters, 6,100 higher-level clusters, and finally the 776 paradigms. Although each paradigm contains, on average, 1,000 papers, some are larger and some are smaller, as shown by different sized circles on the map.

The ring-like structure that is formed by scientific paradigms is very robust. We find similar structures for different years, and for maps generated from scientific journals. "The Structure of Science", a galaxy map shown in the first iteration of Places & Spaces, is a map based on clustering of scientific journals, with superimposition of papers on the journal structure, whereas this map was generated directly from highly-cited papers. "The Structure of Science" shows current science in a disciplinary context, while this map can show the breadth of disciplines that contribute to single paradigms.

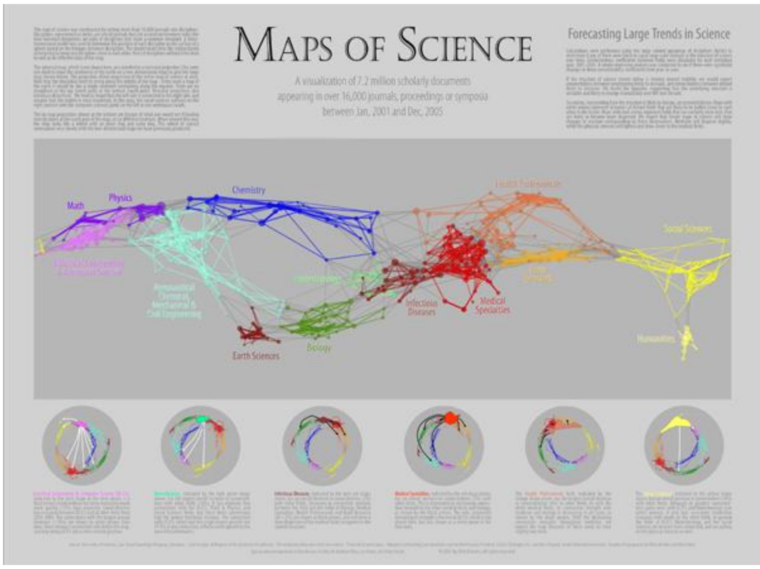
Because of the robust nature of the structure of science and its paradigms, we have placed our 776 scientific paradigms within a **reference system containing 12 radial slices and 6 rings.** This allows the position of each paradigm to be codified and available for lookup; for instance ***Fluid Mechanics* paradigms are in grid B3**.

We have also calculated and displayed the vitality of each paradigm. **Vitality is a measure of the speed at which a group of researchers reaches consensus about major improvements.** Paradigms are constantly being improved, but it usually takes years to reach consensus about which improvements are major. The white circles represent communities where consensus is reached relatively slowly. This is a common phenomenon in the social sciences, ecological sciences, computer sciences, and mathematics disciplines. **The red circles represent communities of researchers where consensus is reached relatively rapidly.** This is more common in physics, chemistry, biochemistry, and many medical disciplines. Very dark circles (such as those in *Astrophysics*, *L5-6*) represent communities where consensus is reached extremely quickly.

The map of scientific paradigms and its reference system can be used for multiple purposes. Countries, industries, companies, universities, and individual researchers can all **locate themselves within the map**, either as single points, or as a specific collection of paradigms. Various metrics, such as vitality, can be overlayed on this reference system to highlight specific impacts. **Science education and personal discovery** can also be enhanced by linking stories and facts to the map that highlight scientific history, current advances and relationships between scientific paradigms.

Boyack & Klavans

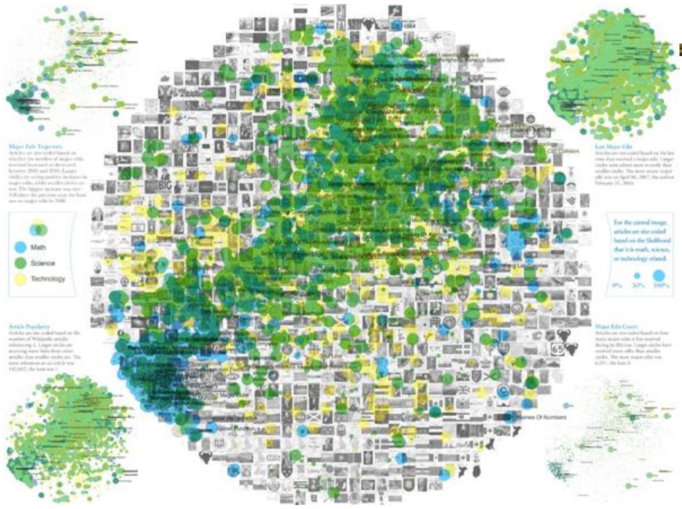
http://scimaps.org/maps/map/map_of_scientific_pa_55/


MAPS OF SCIENCE
A visualization of 7.2 million scholarly documents appearing in over 16,000 journals, proceedings or symposia between Jan, 2001 and Dec, 2005
Forecasting Large Trends in Science
Math
Physics
Chemistry
Earth Sciences
Biology
Infectious Diseases
Medical Specialties
Social Sciences
Humanities

Klavans & Boyack

http://scimaps.org/maps/map/maps_of_science_fore_50/

Science Related Wikipedia Activity



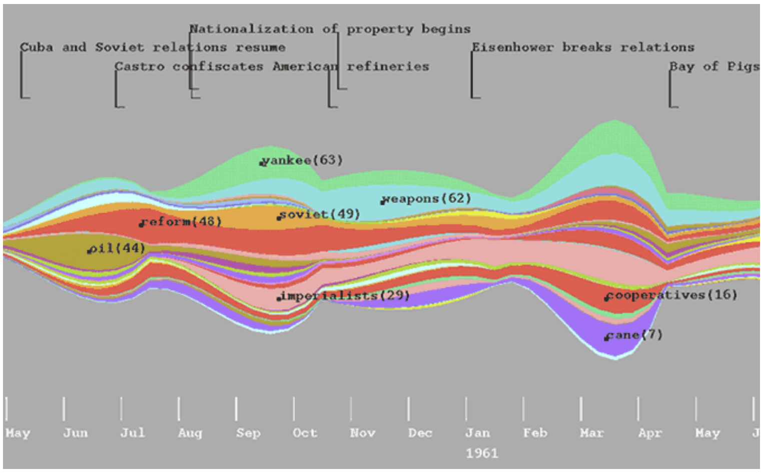
Allgood, Herr, Holloway & Boyack

http://scimaps.org/maps/map/science_related_wiki_49/

Temporal Issues

- Semantic map gives no indication of the chronology of documents
- Can we show themes and how they rise or fall over time?

ThemeRiver



Havre, Hetzler, & Nowell
InfoVis '00

Representation

- Time flows from left->right
- Each band/current is a topic or theme
- Width of band is "strength" of that topic in documents at that time

More Information

- What's in the bands?
- Analysts may want to know about what each band is about

Topic Modeling

- Hot topic in text analysis and visualization
- Latent Dirichlet Allocation
- Unsupervised learning
- Produces "topics" evident throughout doc collection, each modeled by sets of words/terms
- Describes how each document contributes to each topic

TIARA

- Keeps basic ThemeRiver metaphor
- Embed word clouds into bands to tell more about what is in each
- Magnifier lens for getting more details
- Uses Latent Dirichlet Allocation to do text analysis and summarization

Liu et al
CIKM '09, KDD '10, VAST '10

Representation

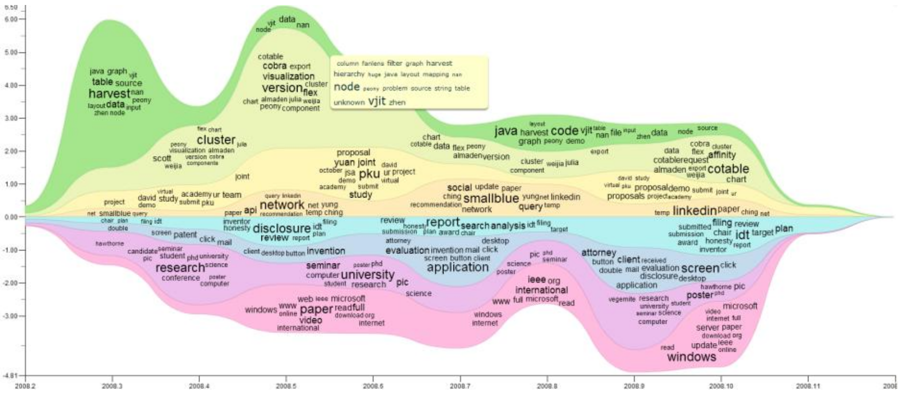
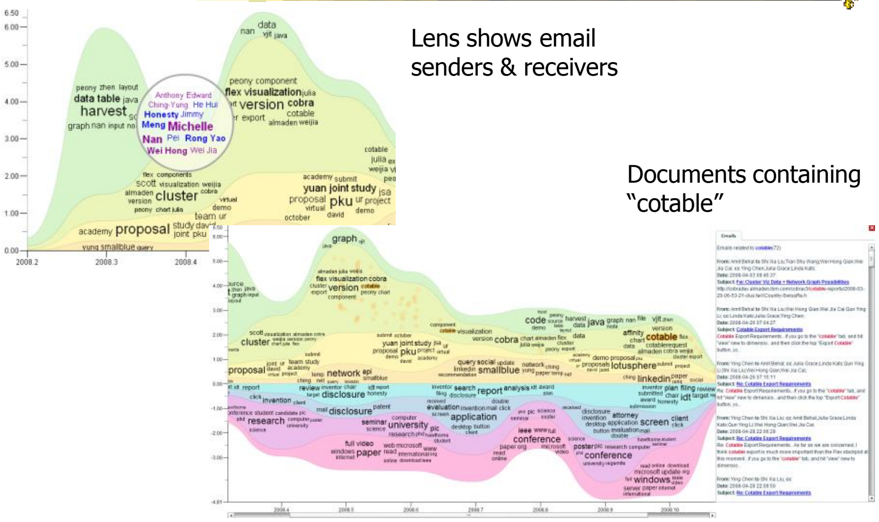


Figure 1. Annotated TIARA-created visual summary of 10,000 emails in the year of 2008. Here, the x-axis encodes the time dimension, the y-axis encodes the importance of each topic. Each layer represents a topic, which is described by a set of keywords. These topic keywords are distributed along the time, summarizing the topic content and the content evolution over time. The tool tip shows the aggregated content of the top-most topic (green one).

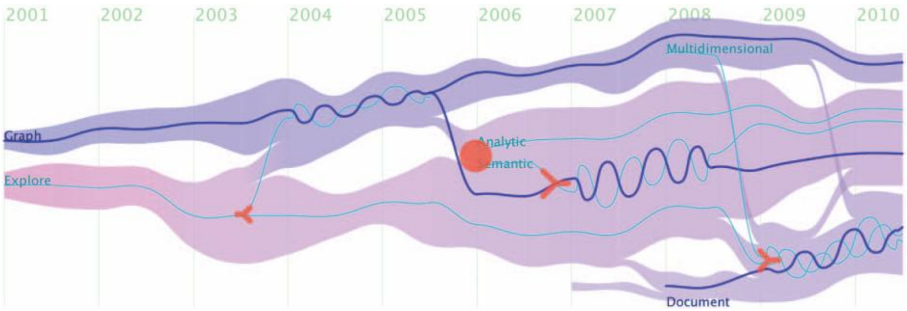
Features



Lens shows email senders & receivers

Documents containing “cotable”

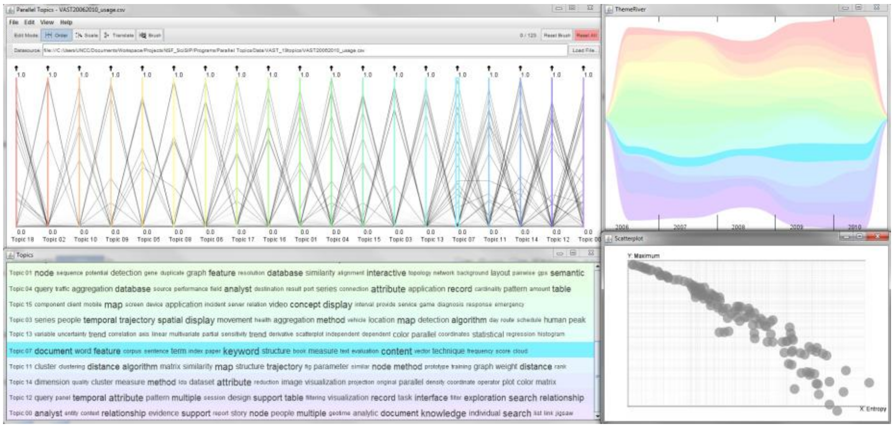
TextFlow



Showing how topics merge and split

Cui et al
TVCG (InfoVis) ‘11

ParallelTopics



Topic 01 node sequence potential detection gene duplicate graph feature resolution database similarity alignment interactive topology network background layout pairwise gps semantic

Topic 04 query traffic aggregation database source performance field analyst destination result port series connection attribute application record cardinality pattern amount table

Topic 15 component client mobile map screen device application incident server relation video concept display interval provide service game diagnosis response emergency

Topic 03 series people temporal trajectory spatial display movement health aggregation method vehicle location map detection algorithm day route schedule human peak

Topic 13 variable uncertainty trend correlation axis linear multivariate partial sensitivity trend derivative scatterplot independent dependent color parallel coordinates statistical regression histogram

Topic 07 document word feature corpus sentence term index paper keyword structure book measure text evaluation content vector technique frequency score cloud

Topic 11 cluster clustering distance algorithm matrix similarity map structure trajectory flp parameter similar node method prototype training graph weight distance rank

Topic 14 dimension quality cluster measure method ica dataset attribute reduction image visualization projection original parallel density coordinate operator plot color matrix

Topic 12 query panel temporal attribute pattern multiple session design support table filtering visualization record task interface filter exploration search relationship

Topic 00 analyst entity context relationship evidence support report story node people multiple geotime analytic document knowledge individual search list link jigsaw

Dou et al
VAST ‘11

HW 6

- Visualize Amazon review collection
- Some nice ideas (see examples)

- What attributes can be shown?

HW 5

- Commercial tool evaluation
  - InfoZoom, Spotfire, Tableau

- Overall, pretty good
- Our observations

HW 7

- Draw a graph
- 10-vertex abstract graph provided
- You draw a node-link representation
- Follow the directions!

- Due Monday (no late ones)
- Don't spend a lot of time

Upcoming

- Graphs & Networks 1
  - Reading

    Lee et al '06

- Graphs & Networks 2
  - Reading

    Perer & Shneiderman '06