VISUALIZATION

Visualizing Text

Overview

- Introduction
- Visualizing search results
- Visualizing documents
- Visualizing document collections

Overview

- Introduction
- Visualizing search results
- Visualizing documents
- Visualizing document collections

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
  - Email
- We often have to make critical decisions based on our understanding of documents
- However, few people have enough time to read everything,
- It is more urgent to scan, understand, operate, and navigate the enormous corpus of documents, and thus to efficiently acquire useful information and knowledge.
- What can information visualization provide to help users in understanding and gathering information from text and document collections?

Text Visualization – Task and 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.
- Understand the history of changes in a document.
- Find connections between documents.

Related Topic - IR

- Information Retrieval
  - Active search process that brings back particular/specific items (will discuss that some today, but not always focus)
- InfoVis, conversely, seems to be most useful when
  - Perhaps not sure precisely what you're looking for
  - More of a browsing task than a search one

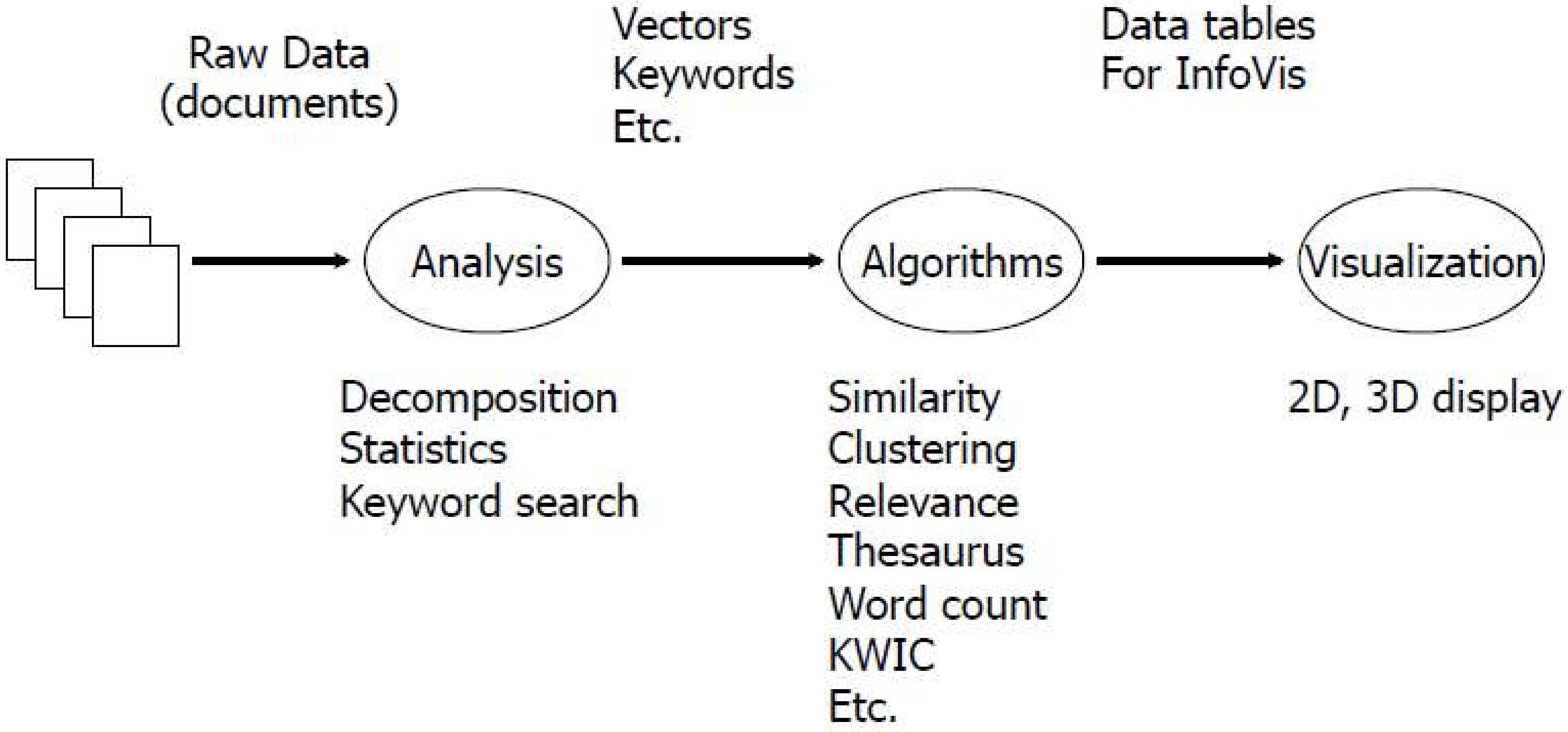
Related Topic - Sensemaking

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

Challenge

- Text is nominal data
  - Does not seem to map to geometric/graphical presentation as easily as ordinal and quantitative data
- Unstructured text does NOT have any explicit meta-data
  - Just that infinitely big collection of nominal data
  - Meta-data is sometimes extracted from raw text (Text Mining)
- The “Raw data --> Data Table” mapping now becomes more important

Text Visualization - Process



Raw Data
(documents)

Vectors
Keywords
Etc.

Data tables
For InfoVis

Analysis

Algorithms

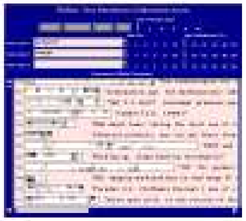
Visualization

Decomposition
Statistics
Keyword search

Similarity
Clustering
Relevance
Thesaurus
Word count
KWIC
Etc.

2D, 3D display

Types of Text Visualization

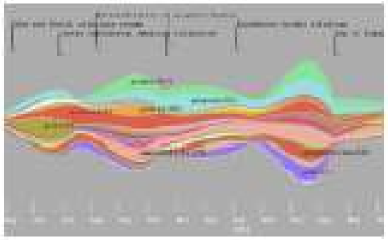


Visualization for IR
Helping search

Visualizing text
Showing words, phrases, and sentences

→

Visualizing document sets
Words, entities & sentences
Analysis metrics
Concepts & themes



Overview

- Introduction
- Visualizing search results
- Visualizing documents
- Visualizing document collections

Visualization for IR

- Can InfoVis help IR?
- Assume there is some active search or query
    - Show results visually
    - Show how query terms relate to results
    - …
- Visualizing the results of search operations is another big area in text infovis

Search - Problems

- Query responses do not 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

Tile Bars

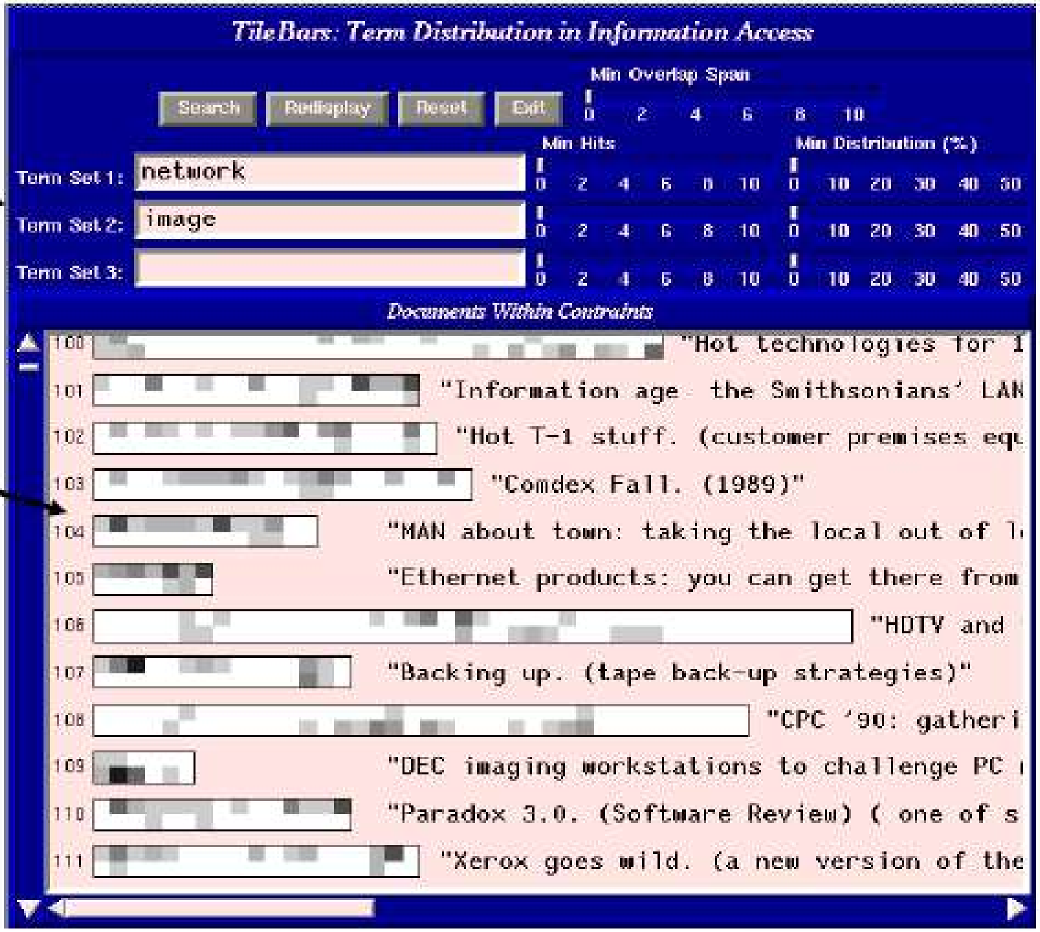
- Goal
  - Minimize time and effort for deciding which documents to view in detail
- Idea
  - Show the role of the query terms in the retrieved documents, making use of document structure
  - Graphical representation of term distribution and overlap using Tile bars
- Simultaneously indicate:
  - Relative document length
  - Frequency of term sets in document
  - Distribution of term sets with respect to the document and each other

- TileBars: Visualization of Term Distribution Information in Full Text Information Access Marti A. Hearst   Proceedings of CHI ’95, Denver, CO, May 1995

Tile Bar

Search terms

Presentation



TileBars: Term Distribution in Information Access

Term Set 1: network

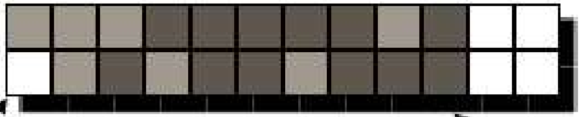
Term Set 2: image

Term Set 3:

Documents Within Contraints

100 "Hot technologies for 1
101 "Information age the Smithsonians' LAN
102 "Hot T-1 stuff. (customer premises equ
103 "Comdex Fall. (1989)"
104 "MAN about town: taking the local out of l
105 "Ethernet products: you can get there from
106 "HDTV and
107 "Backing up. (tape back-up strategies)"
108 "CPC '90: gatheri
109 "DEC imaging workstations to challenge PC
110 "Paradox 3.0. (Software Review) (one of s
111 "Xerox goes wild. (a new version of the

Two search terms

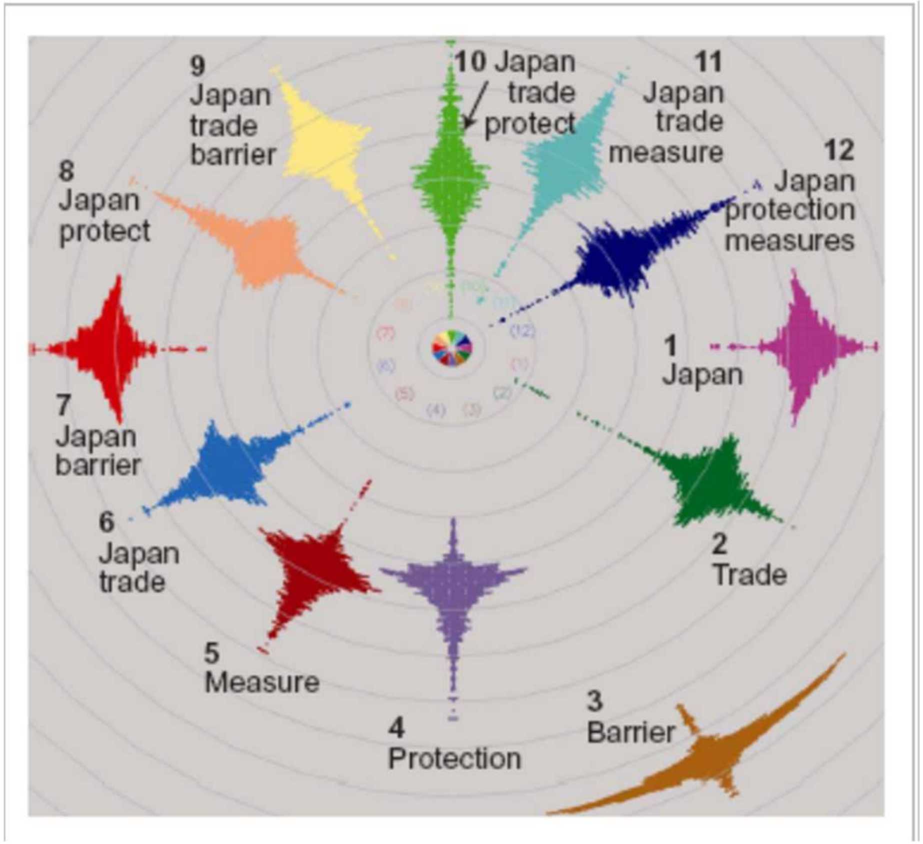


Blocks indicate "chunks" of text, such as paragraphs

Blocks are darkened according to the frequency of the term in the document

Sparkler

- Visually presenting and exploring the results for different queries on the same topic simultaneously
- Glyphs representing documents arranged along a line
  - Distance from the center indicates degree of relevance to the query.
  - When there were multiple documents with the same relevance score, they were spread out horizontally from the line, forming a visualization of the distribution of relevance scores.
- The Sparkler for the different queries arranged along a circle
  - Each query's visualization assigned a different color.
- Selecting a document in one sparkler caused its position to be highlighted in the other Sparkler visualizations.



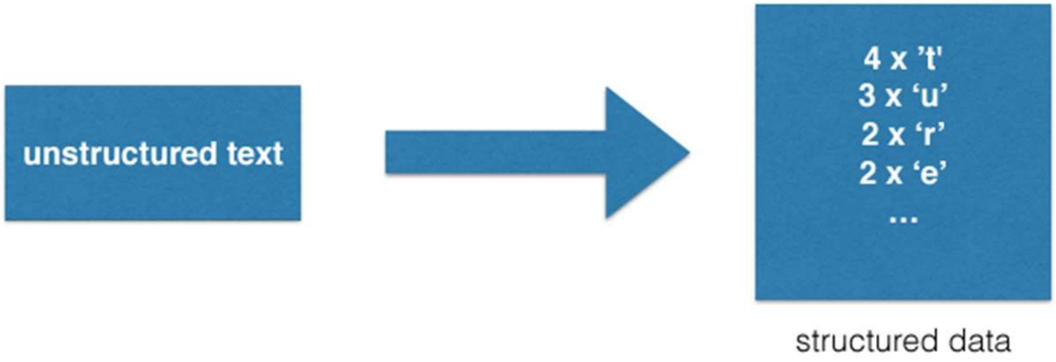
Havre S, Hetzler E, Perrine K, Jurrus E, Miller N. Interactive visualization of multiple query results. InfoVis 2001

Overview

- Introduction
- Visualizing search results
- Visualizing documents
- Visualizing document collections

Visualizing Single Documents

- How do we represent the words, phrases, and sentences in a document or set of documents?
  - Main goal of understanding versus search
  - Visualizing text (features) requires a transformation step: discretization, aggregation, normalization



unstructured text

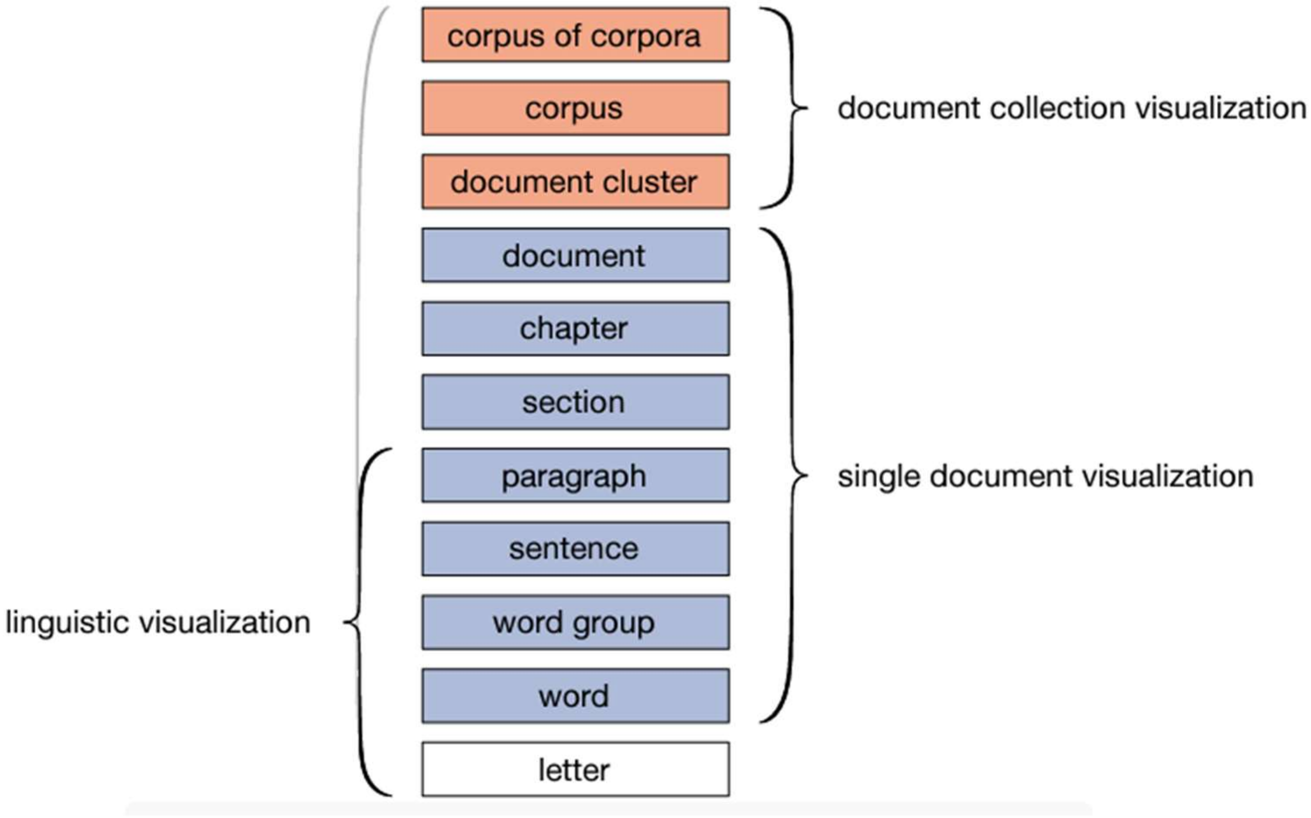
4 x 't'
3 x 'u'
2 x 'r'
2 x 'e'
...

structured data

Typical steps of processing to derive Text Features

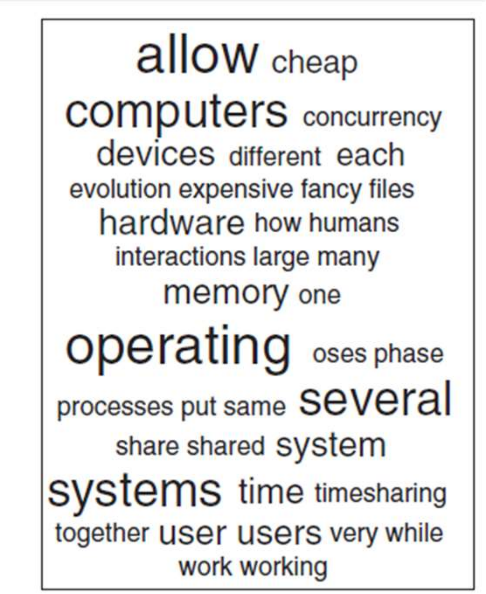
- Large collections require pre-processing of text to extract information.
- Typical steps are:
  - Cleaning
  - Sentence splitting
  - Changing to lower case
  - Stopword removal
  - Stemming
  - PoS tagging
  - Named entity recognition
  - Deep parsing – trying to “understand” text

Text Units Hierarchy



Vocabulary – based Visualization

- Idea is to show word/concept importance through visual means
- Tag Clouds: Provide layouts of raw tokens, colored, and sized by the corresponding word frequency within a single document.



Tag Cloud/Wordle

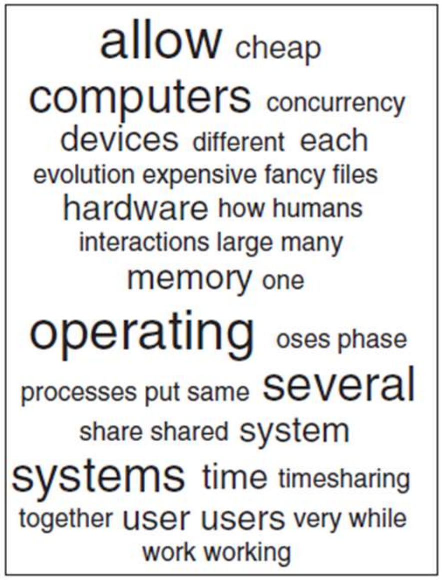
- Tightly packed words Idea is to show word/concept importance through visual means

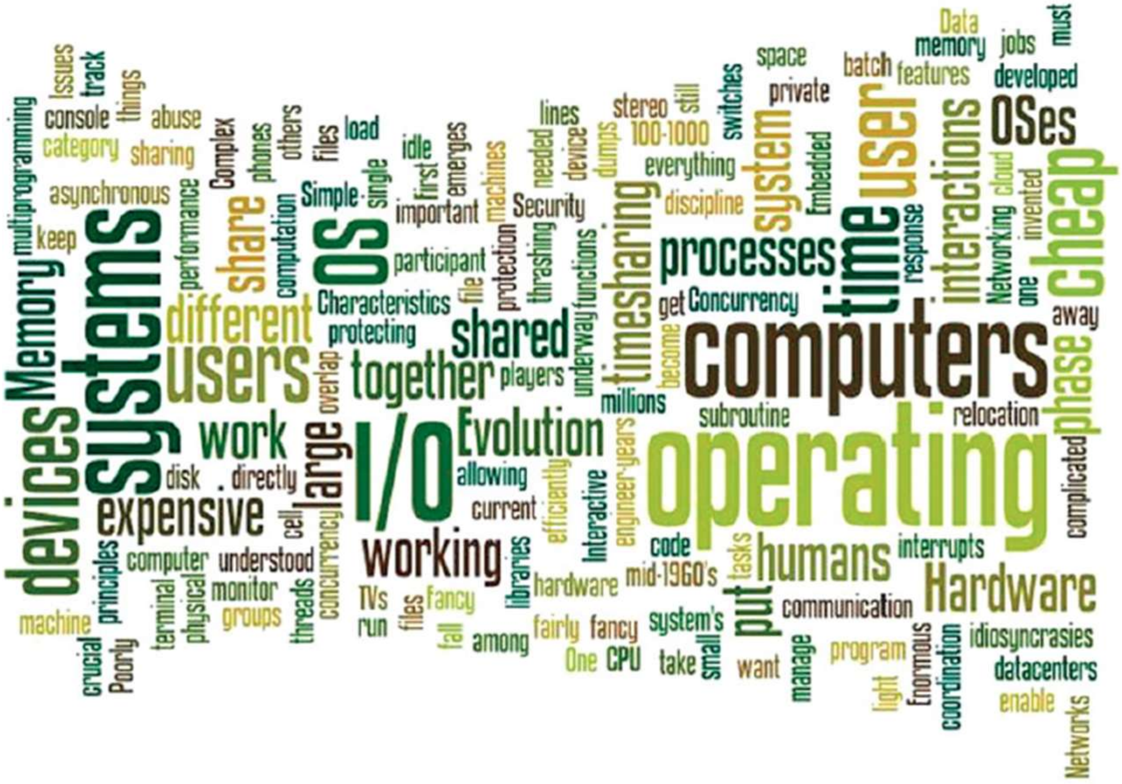
Tag Clouds

- Provide layouts of raw tokens, colored, and sized by the corresponding word frequency within a single document.
- Words displayed alphabetically or most frequent words in middle

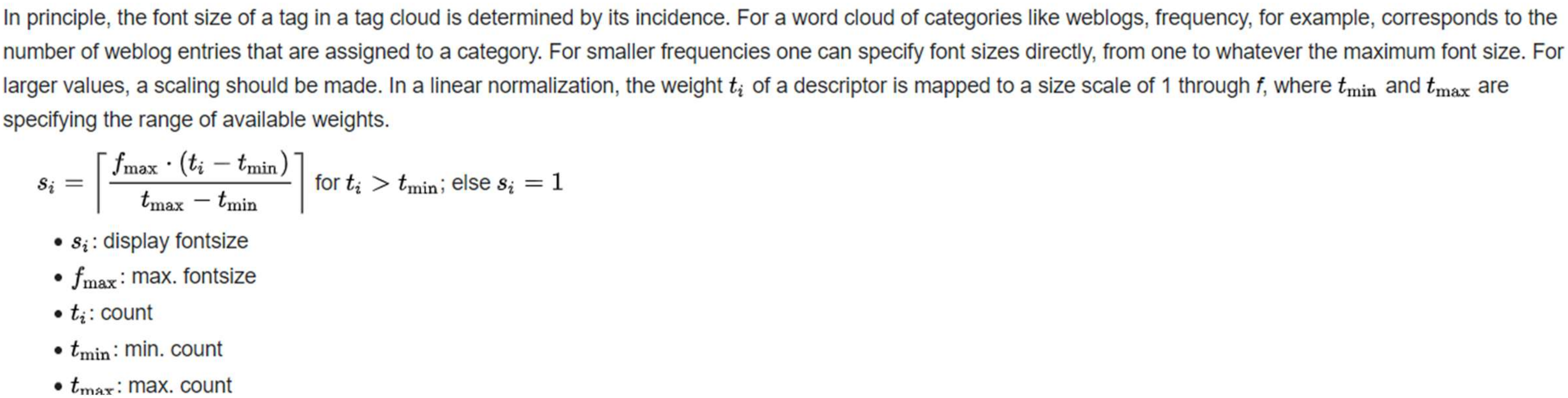
Wordle

- Wordle adds a layout algorithm to allow users to modify the font, color, or configuration.





Tag Cloud - Creation



In principle, the font size of a tag in a tag cloud is determined by its incidence. For a word cloud of categories like weblogs, frequency, for example, corresponds to the number of weblog entries that are assigned to a category. For smaller frequencies one can specify font sizes directly, from one to whatever the maximum font size. For larger values, a scaling should be made. In a linear normalization, the weight t_i of a descriptor is mapped to a size scale of 1 through *f*, where t_{min} and t_{max} are specifying the range of available weights.

$$s_i = \left\lceil \frac{f_{\mathrm{max}} \cdot (t_i - t_{\mathrm{min}})}{t_{\mathrm{max}} - t_{\mathrm{min}}} \right\rceil \text{ for } t_i > t_{\mathrm{min}}; \text{ else } s_i = 1$$

- s_i: display fontsize
- f_{max}: max. fontsize
- t_i: count
- t_{min}: min. count
- t_{max}: max. count

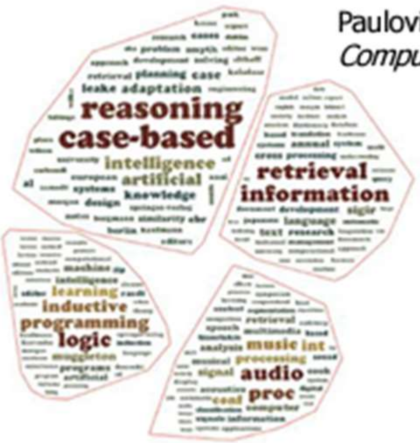
Implementations of tag clouds also include text parsing and filtering out unhelpful tags such as common words, numbers, and punctuation.

Word Cloud - Evaluation

- Actually not a great visualization. Why?
  - Hard to find a particular word
  - Long words get increased visual emphasis
  - Font sizes are hard to compare
  - Alphabetical ordering not ideal for many tasks
- Word Clouds are really more overview-style visualizations
  - Don’t really support queries, searches, drill-down
- However very popular
  - Serve as social signifiers that provide a friendly atmosphere that provide a point of entry into a complex site
  - Act as individual and group mirrors
  - Easy to understand

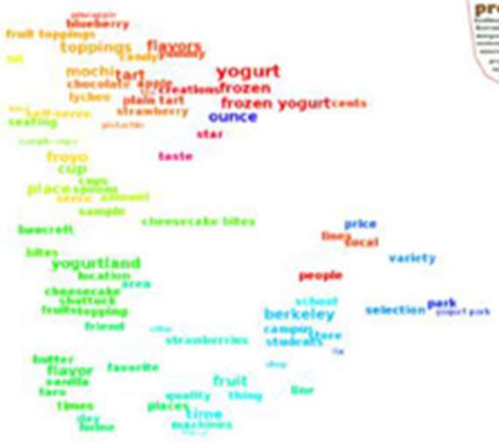
Semantic/Context Word Clouds

Group by
related concepts

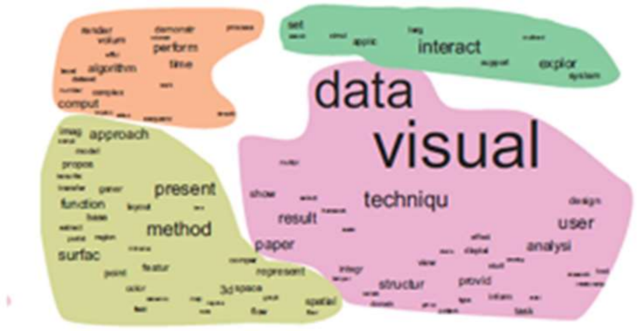


Paulovich et al
Computer Graphics Forum '12

Wang et al
Graphics Interface '14



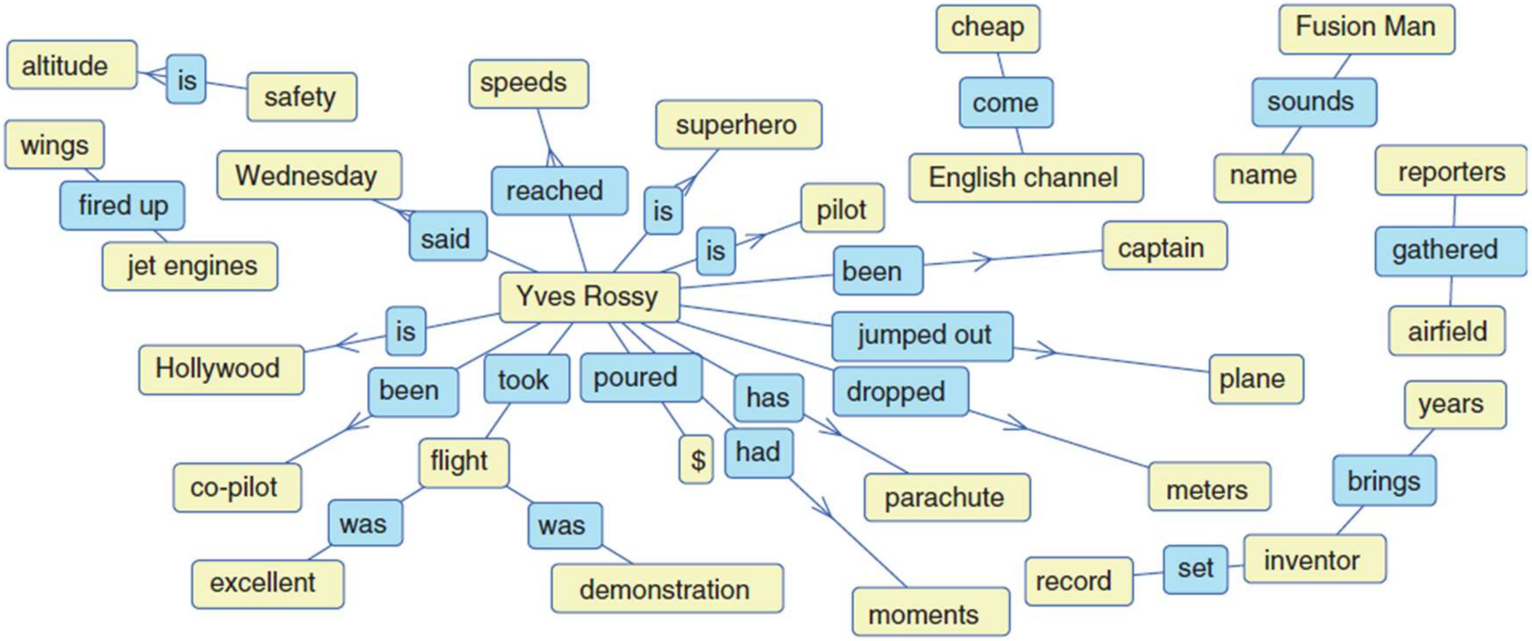
Wu et al
Computer Graphics Forum '11



Beyond Individual Words

- Can we show combinations of words, phrases, and sentences?

Semantic Graph



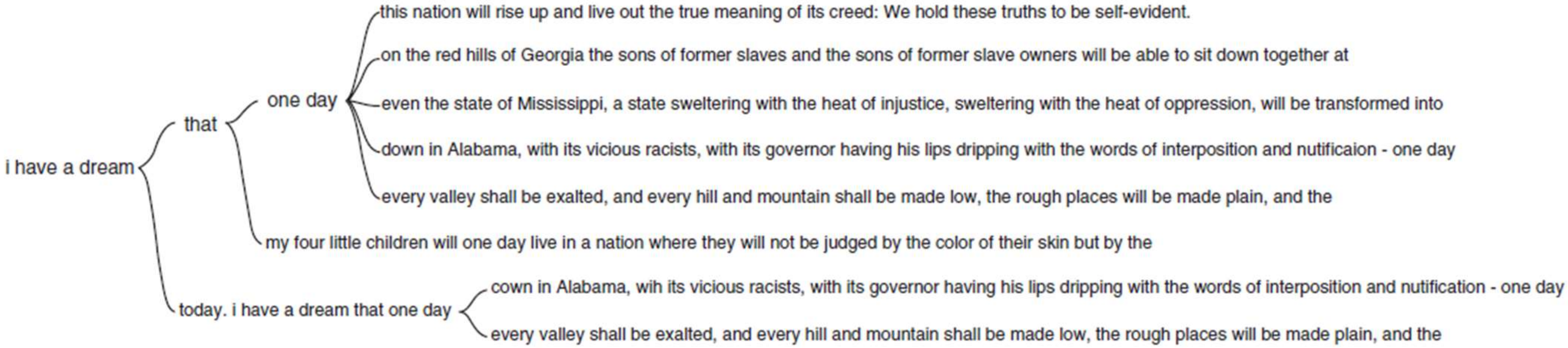
- SemanticGraphs (Rusu et al InfoVis 2009) is a visualization based on the semantic representation of a document in the form of a semantic graph.
- The document is summarized with the semantic graph and the list of extracted triplets.
  - It extracts subject–verb–object for each sentence by the parse tree.
  - Then, it links the triplets to their corresponding entity, which needs to resolve pronominal anaphors as well as to attach the associate WordNet synset.

Word Tree

- The WordTree visualization provides the representation of both word frequency and context.
- Size is used to represent frequency of the term or phrase.
- The root of the tree is a user-selected word or phrase
- The branches represent the contexts in which the word or phrase is used in the document.
- Users can click on a branch, choose a different search term or re-center the tree.
- Clicking on phrase makes it the focus
- Shows context of a word or words
  - Follow word with all the phrases that follow it

- *Wattenberg & Viégas TVCG (InfoVis) '08*

Example - Word Tree



i have a dream

that

one day

this nation will rise up and live out the true meaning of its creed: We hold these truths to be self-evident.

on the red hills of Georgia the sons of former slaves and the sons of former slave owners will be able to sit down together at

even the state of Mississippi, a state sweltering with the heat of injustice, sweltering with the heat of oppression, will be transformed into

down in Alabama, with its vicious racists, with its governor having his lips dripping with the words of interposition and nutificaion - one day

every valley shall be exalted, and every hill and mountain shall be made low, the rough places will be made plain, and the

my four little children will one day live in a nation where they will not be judged by the color of their skin but by the

today. i have a dream that one day

cown in Alabama, wih its vicious racists, with its governor having his lips dripping with the words of interposition and nutification - one day

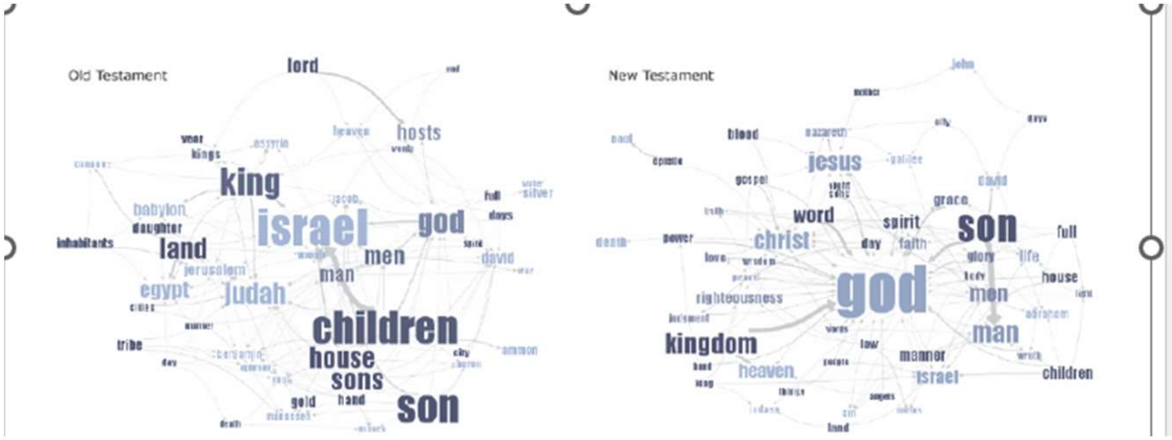
every valley shall be exalted, and every hill and mountain shall be made low, the rough places will be made plain, and the

WordTree shows all occurrences of ‘I have a dream’ in Martin Luther King’s historical speech

Phrase Nets

- A phrase net examine unstructured text documents by displaying a graph whose nodes are words and whose edges indicate that two words are linked by a user-specified relation.
  - These relations may be defined either at the syntactic or lexical level; different relations often produce very different perspectives on the same text.
  - Taken together, these perspectives often provide an illuminating visual overview of the key concepts and relations in a document
- Presents pairs of terms from phrases such as
  - X and Y
  - X’s Y
  - X at Y
  - X (is|are|was|were) Y
- Uses special graph layout algorithm with compression and simplification

Example – Phrase Nets



Matching the pattern “X of Y” to compare old and new testaments. Israel takes the central place in the Old Testament and God acts at he the main pattern in the New Testament.

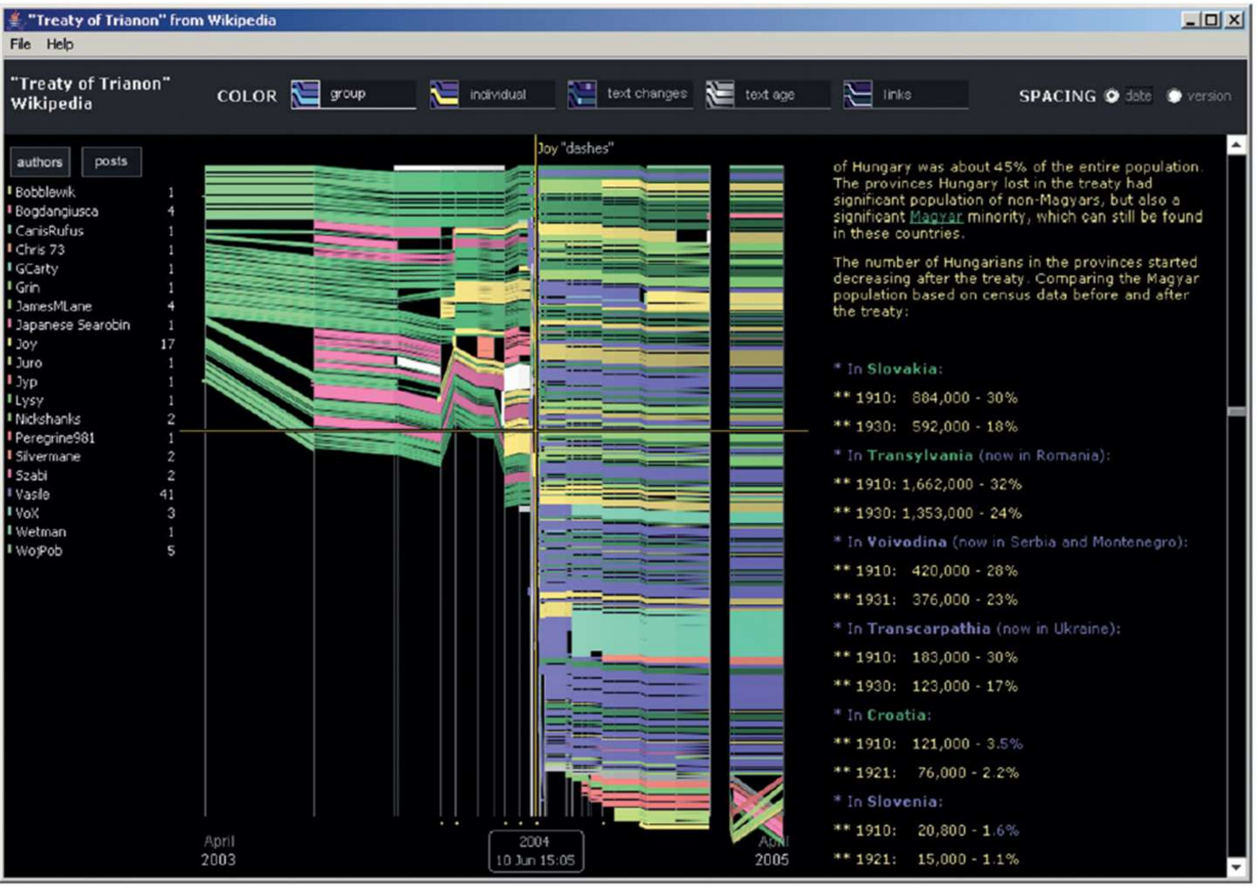
van Ham et al TVCG (InfoVis) ‘09

Visualizing different versions of a document

- History Flow is designed to show changes between multiple document versions on Wikipedia..
- It also reveals some complex patterns of cooperation and confliction, such as vandalism, anonymity versus named authorship, etc.
- Each version of the document is represented by a line with length which is related to the length of its text
- The lines are ordered by date.
- Sections of each line are colored differently according to the different authors.
- Link sections of text that keep the same between adjacent version by drawing shaded connections

Viegas FB, Wattenberg M, Dave K.
Studying cooperation and conflict between authors with history flo visualizations
Proceedings of the SIGCHI 2004

Example – History Flow



authors	posts
Bobblewik	1
Bogdangiusca	4
CanisRufus	1
Chris 73	1
GCarty	1
Grin	1
JamesMLane	4
Japanese Searobin	1
Joy	17
Juro	1
Jyp	1
Lysy	1
Nickshanks	2
Peregrine981	1
Silvermane	2
Szabi	2
Vasile	41
VoX	3
Wetman	1
WojPob	5

of Hungary was about 45% of the entire population. The provinces Hungary lost in the treaty had significant population of non-Magyars, but also a significant Magyar minority, which can still be found in these countries.

The number of Hungarians in the provinces started decreasing after the treaty. Comparing the Magyar population based on census data before and after the treaty:

* In **Slovakia**:

** 1910:   884,000 - 30%

** 1930:   592,000 - 18%

* In **Transylvania** (now in Romania):

** 1910: 1,662,000 - 32%

** 1930: 1,353,000 - 24%

* In **Voivodina** (now in Serbia and Montenegro):

** 1910:   420,000 - 28%

** 1931:   376,000 - 23%

* In **Transcarpathia** (now in Ukraine):

** 1910:   183,000 - 30%

** 1930:   123,000 - 17%

* In **Croatia**:

** 1910:   121,000 - 3.5%

** 1921:   76,000 - 2.2%

* In **Slovenia**:

** 1910:   20,800 - 1.6%

** 1921:   15,000 - 1.1%

- Version history for the Wikipedia entry *Treaty of Trianon*.
- The left side reveals different authors that are colored differently.
- The center part shows the visualization;
- In the right side, a text view closely linked with the visualization shows the detailed content.
- Users can locate on the visualization by moving a set of crosshairs
- The text view shows the corresponding version and position

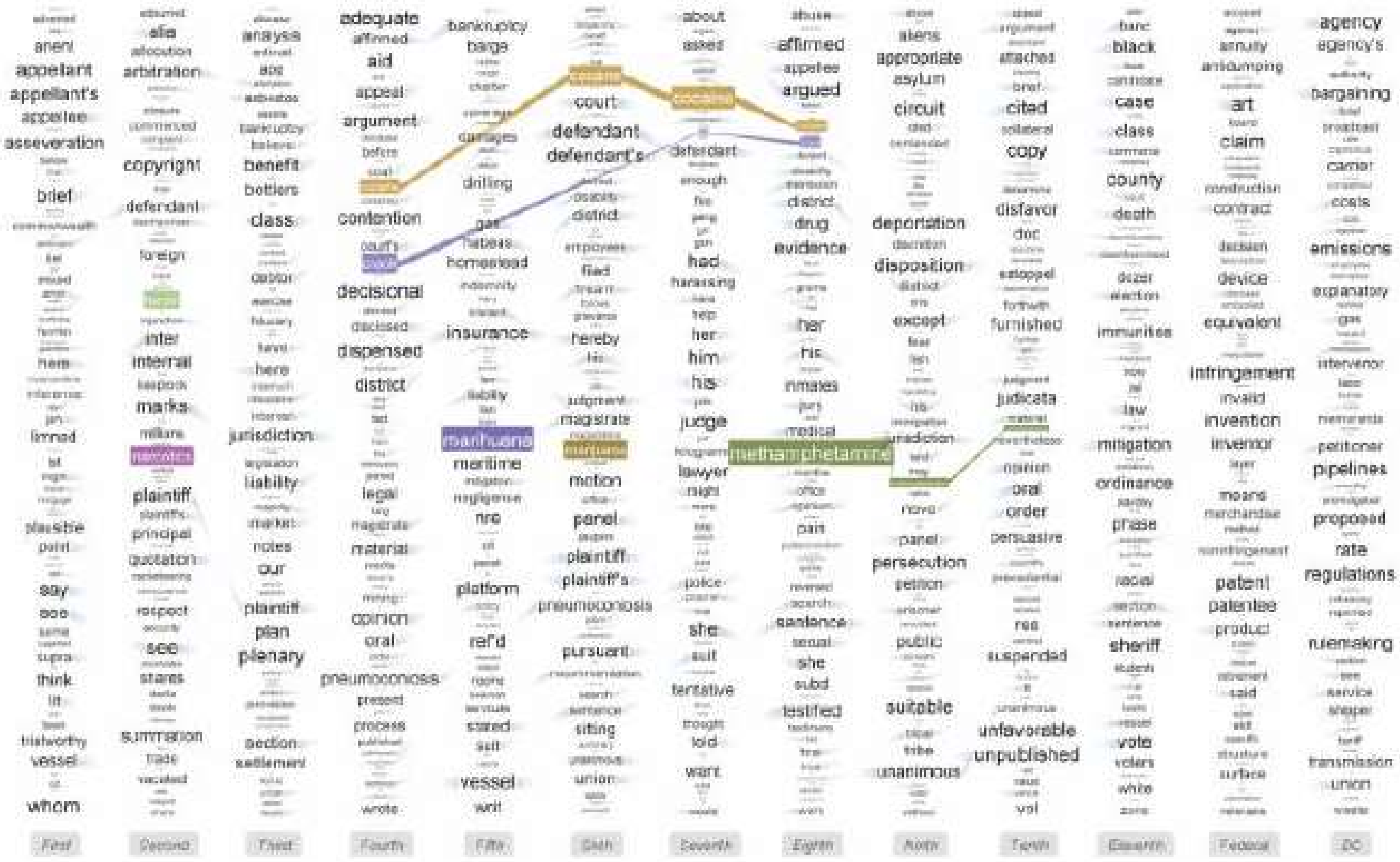
Overview

- Introduction
- Visualizing search results
- Visualizing documents
- Visualizing document collections

Comparing Multiple Documents

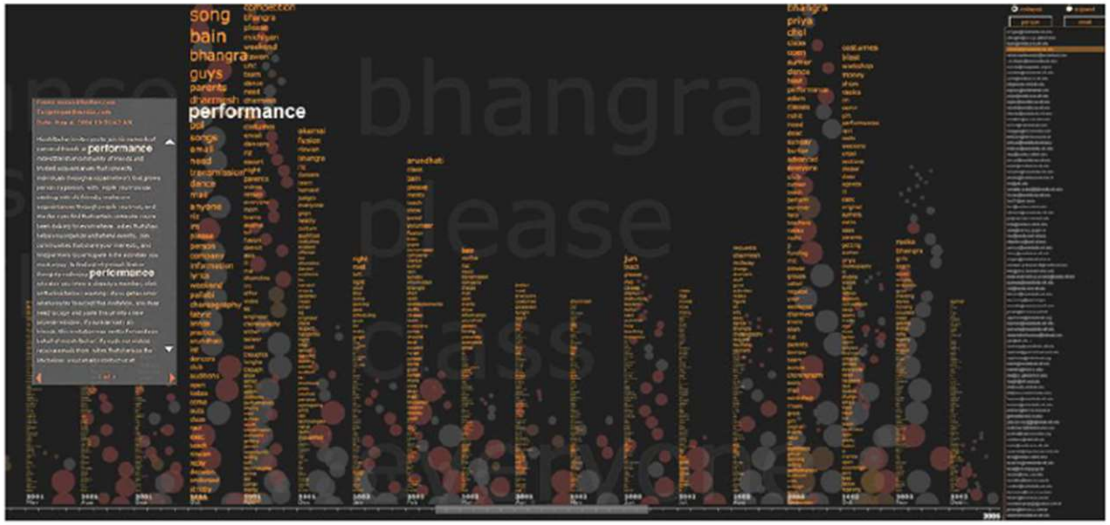
- Move to collections of documents
  - Still do words, phrases, sentences
  - Add
    - More context of documents
    - Document analysis metrics
    - Document meta-data
    - Document entities
    - Connections between documents
    - Documents concepts and themes

Parallel Tag Cloud



The visualization technique combines graphical elements from parallel coordinates and traditional tag clouds to provide rich overviews of a document collection (C. Collins, F. B. Viegas and M. Wattenberg, "Parallel Tag Clouds to explore and analyze faceted text corpora," *2009 IEEE Symposium on Visual Analytics Science and Technology*)

Themail

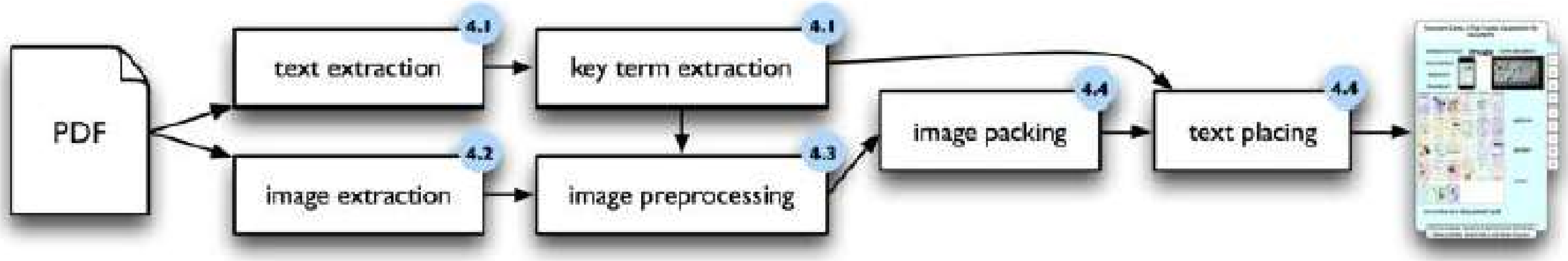


A visualization that portrays relationships using the interaction histories preserved in email archives. Using the content of exchanged messages, it shows the words that characterize one’s correspondence with an individual and how they change over the period of the relationship.

Viegas et al CHI 2006

Document Cards

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

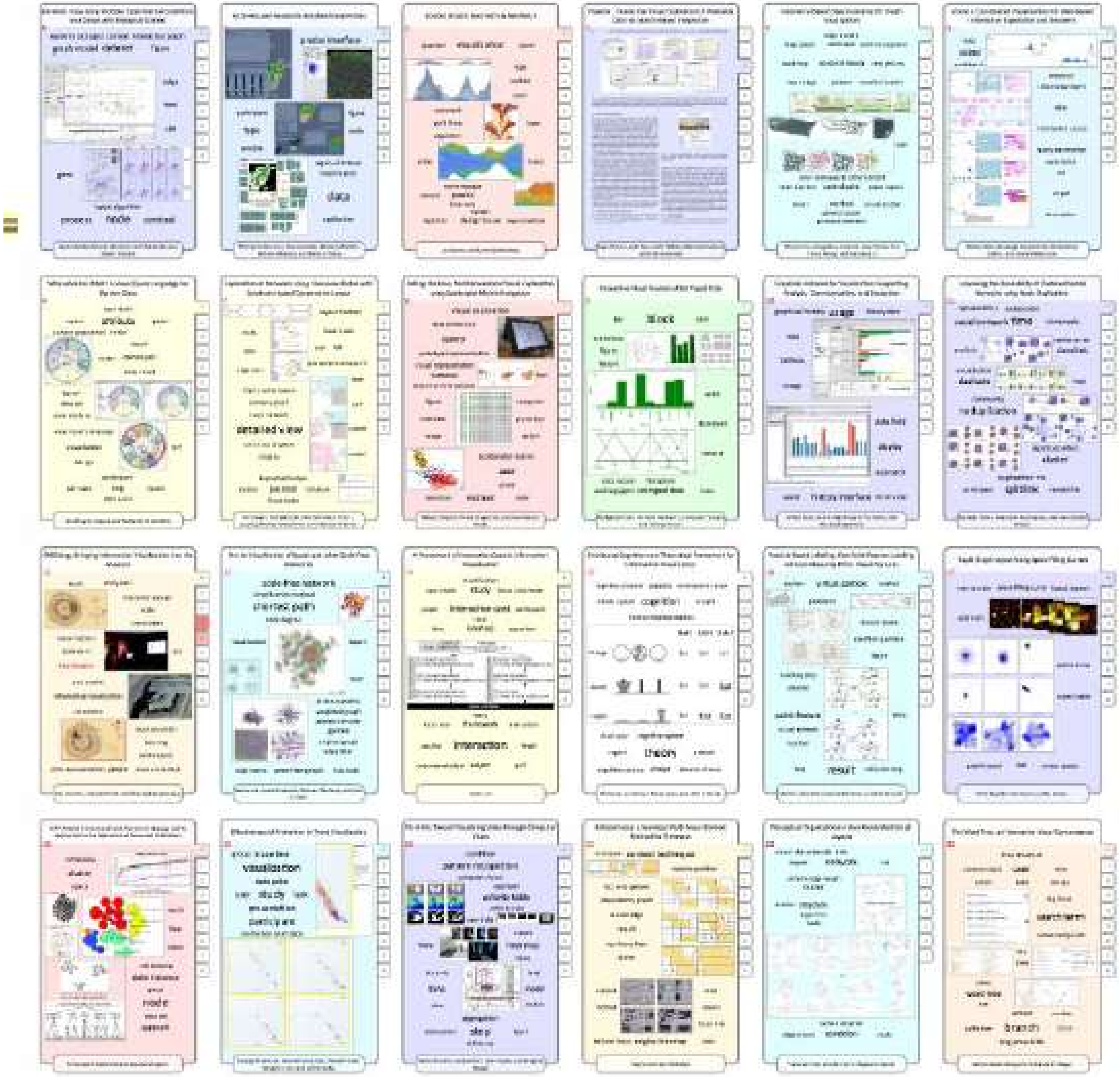


Strobelt et al
TVCG (InfoVis) '09

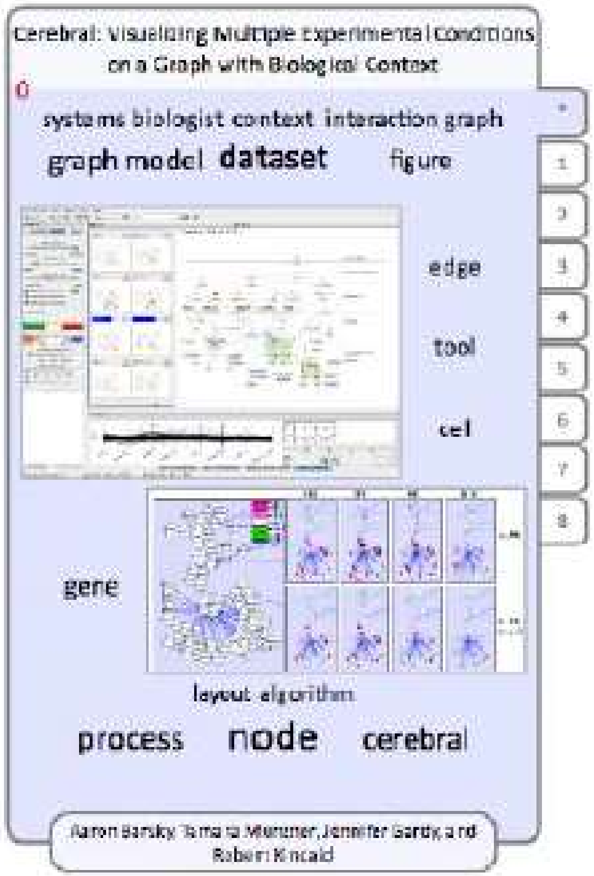
Interaction

- 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

Example

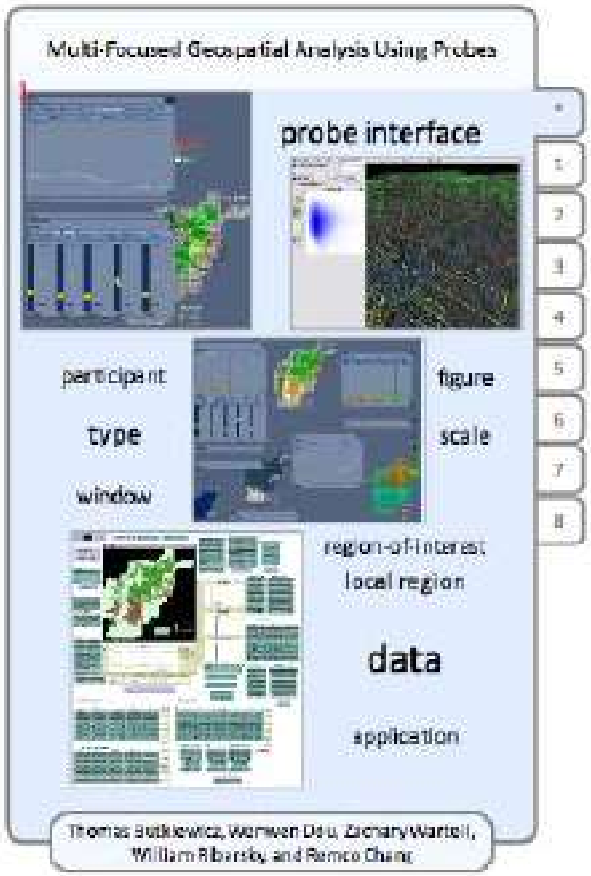


Zooming In



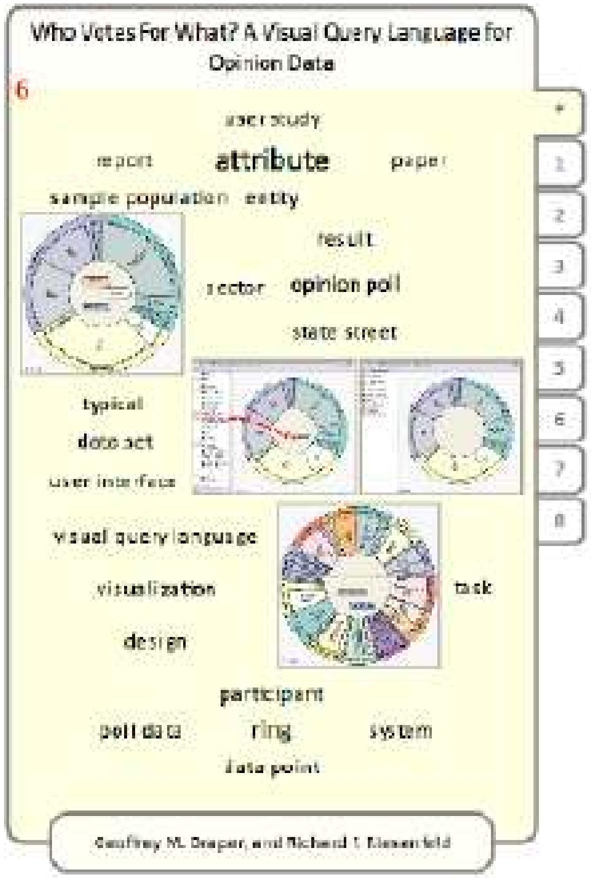
Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context

Aaron Barsky, Tamara Munzner, Jennifer Gardy, and Robert Kincaid



Multi-Focused Geospatial Analysis Using Probes

Thomas Butkiewicz, Wenwen Dou, Zachary Wartell, William Ribarsky, and Remco Chang



Who Votes For What? A Visual Query Language for Opinion Data

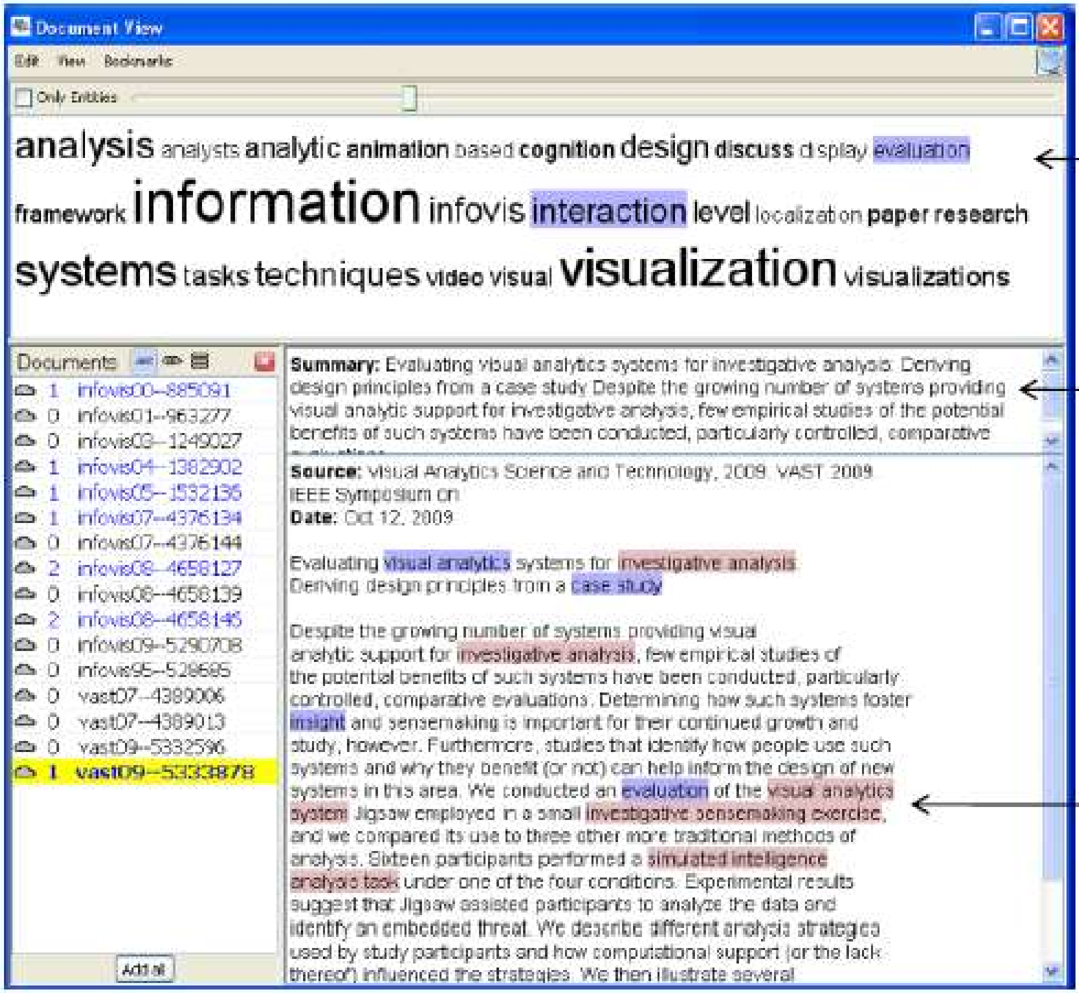
Geoffrey M. Draper, and Richard F. Riesenfeld

Jigsaw

- Jigsaw is an interactive visualization for document exploration and sense-making
- Variety of visualizations ranging from word-specific, to entity connections, to document clusters
- Primary focus is on entity-document and entity-entity connection
- Entities could be people, places, dates, organizations, and so on
- Search capability coupled with interactive exploration

- *Stasko, Görg, & Liu*
- Jigsaw: supporting investigative analysis through interactive visualization
- *InfoVis* 2008

Jisgaw- Document View


Document View
Edit
View
Bookmarks
□ Only Entities
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
Documents
Add all

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

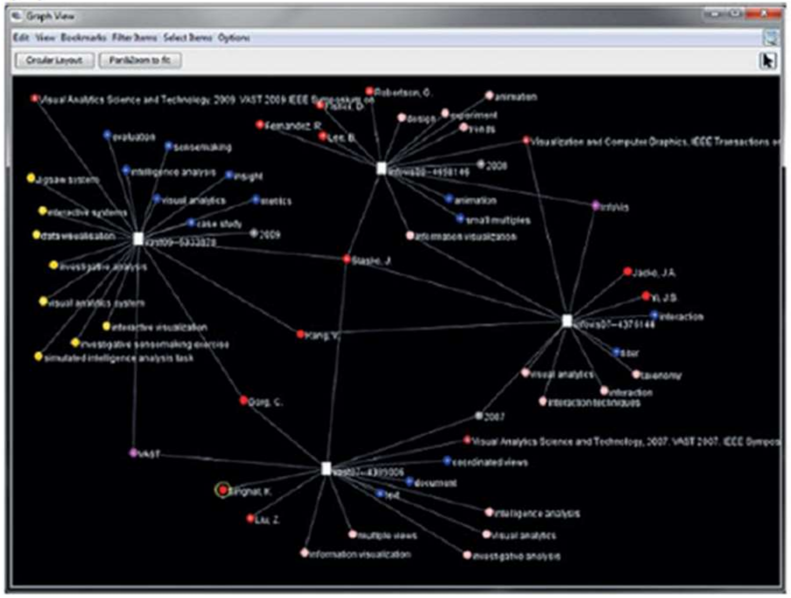
Wordcloud overview

Doc List

Document summary

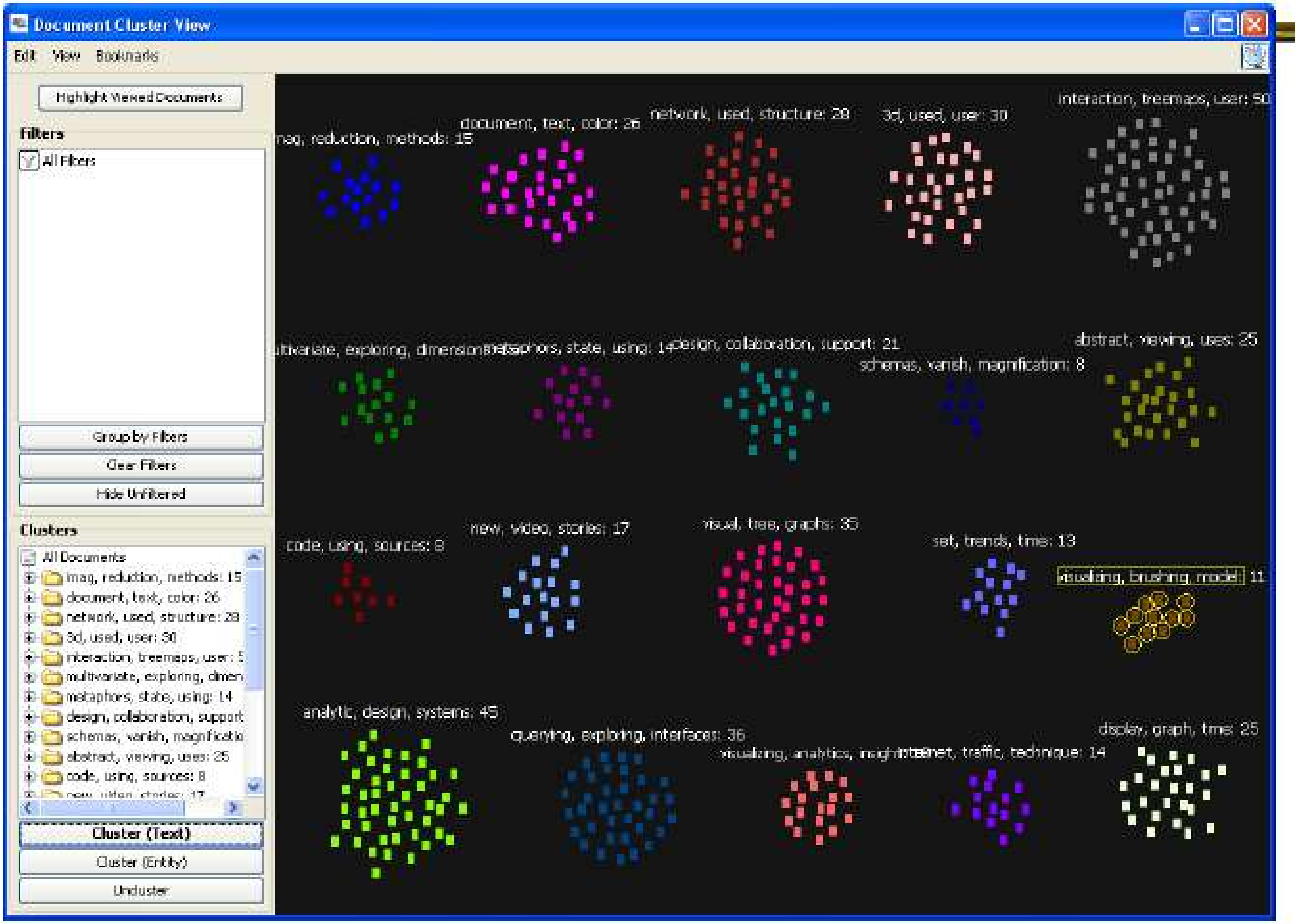
Selected document's text with entities identified

Jisgaw- Graph View



Shows connections between documents and entities in a form of node-link diagram, documents are represented by white rectangles and entities are represented by circles colored differently

Jisgaw - 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, dimension
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
internet, traffic, technique: 14
display, graph, time: 25

Visualization of Document Themes

- Someone may not have time to read them all documents
  - Someone just wants to understand the main topics in them

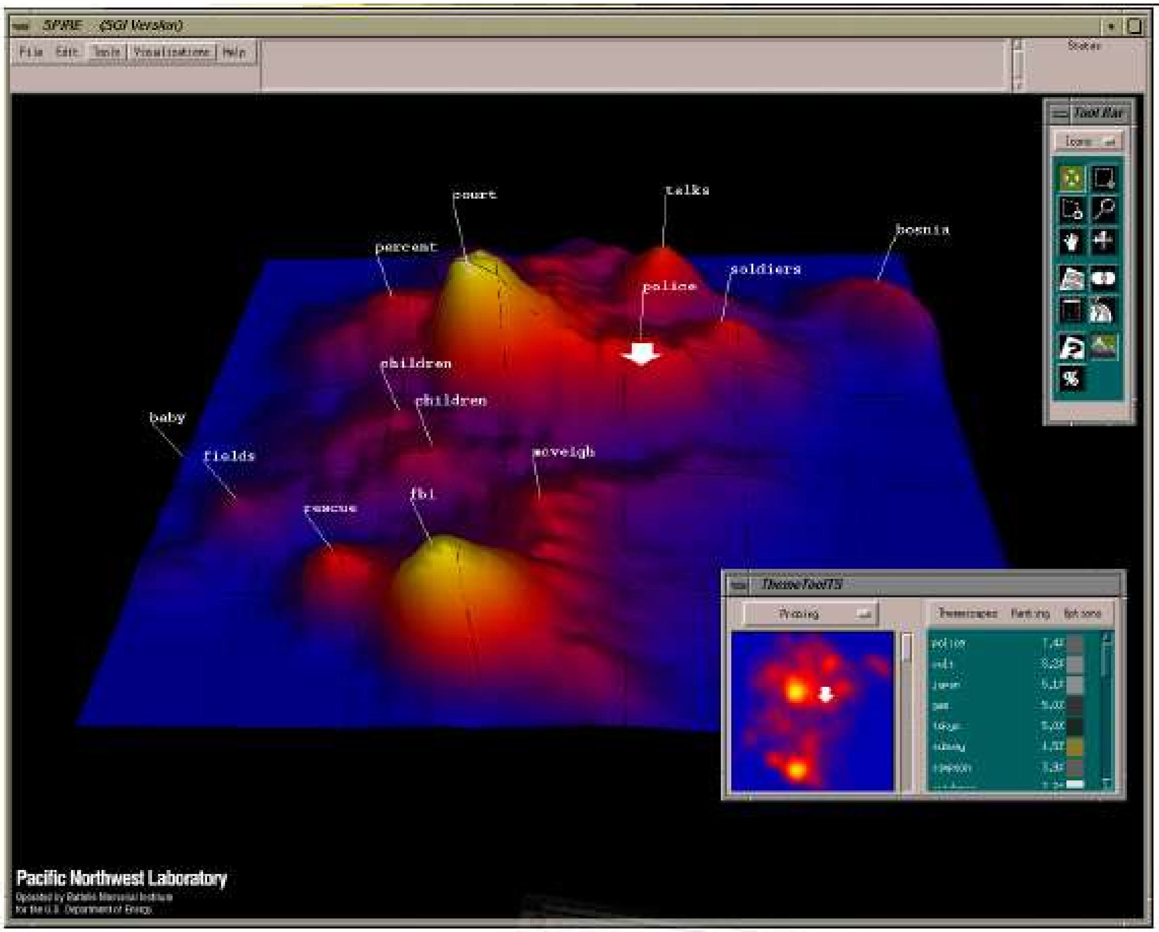
- Look for sets documents that all have common theme
  - Closely related to each other, but different from rest

- The main goal is to discover one or more specific topics and to reflect the relationships among various topics

- Various visualizations developed by *Pacific Northwest National Laboratory*

ThemeView

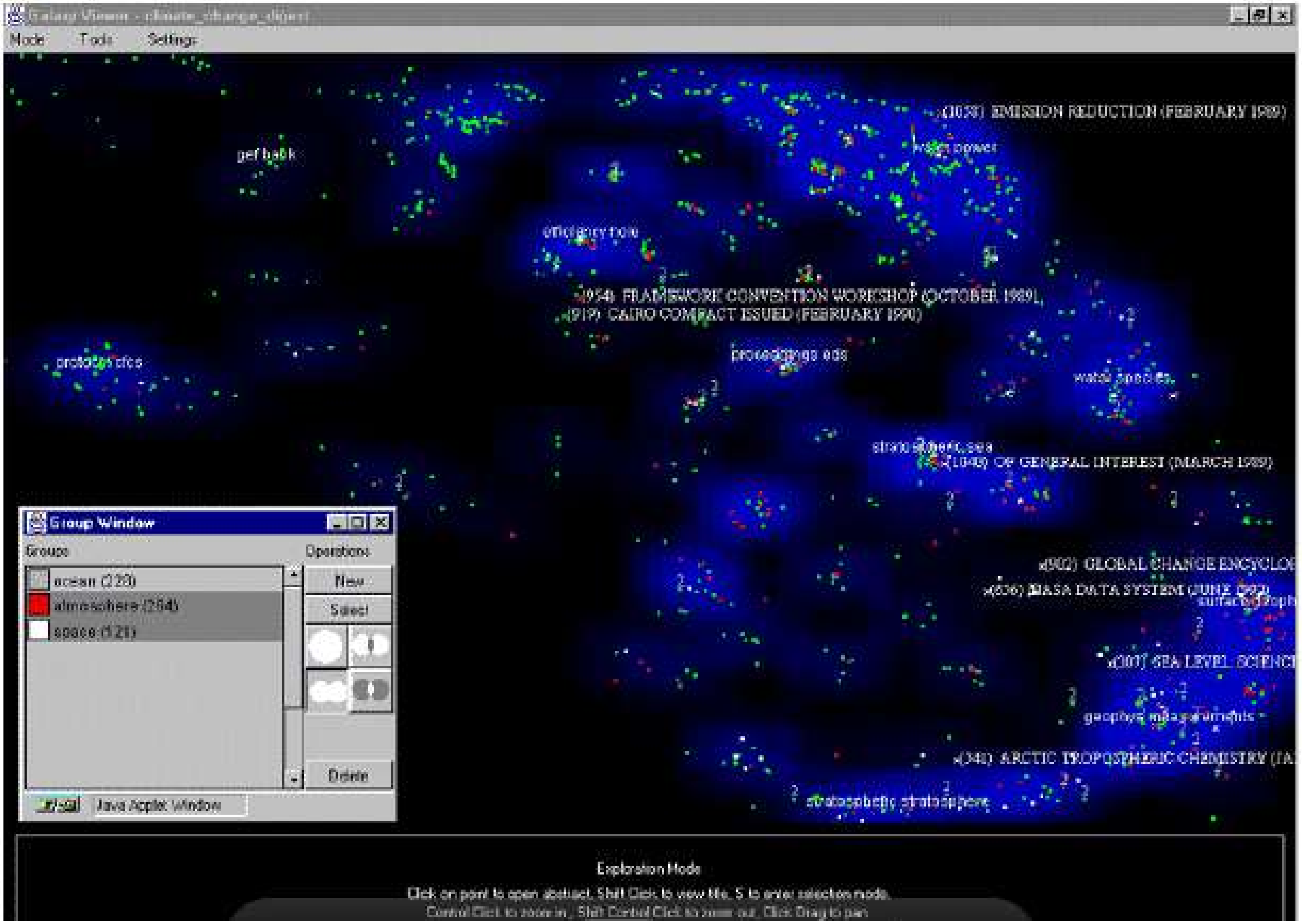
In the ThemeView visualization, the topics or themes within a set of documents are shown as a relief map of natural terrain. The mountains in the ThemeView^{TM} indicate dominant themes. The height of the peaks indicates the relative strengths of the topics in the document set. Similar themes appear close together, while unrelated themes are separated by larger distances. ThemeView provides a visual overview of the major topics contained in a set of documents. Combined with its exploration tools, ThemeView permits the analyst to identify unanticipated relationships and examine changes in topics over time.



https://in-spire.pnnl.gov/

Web Theme

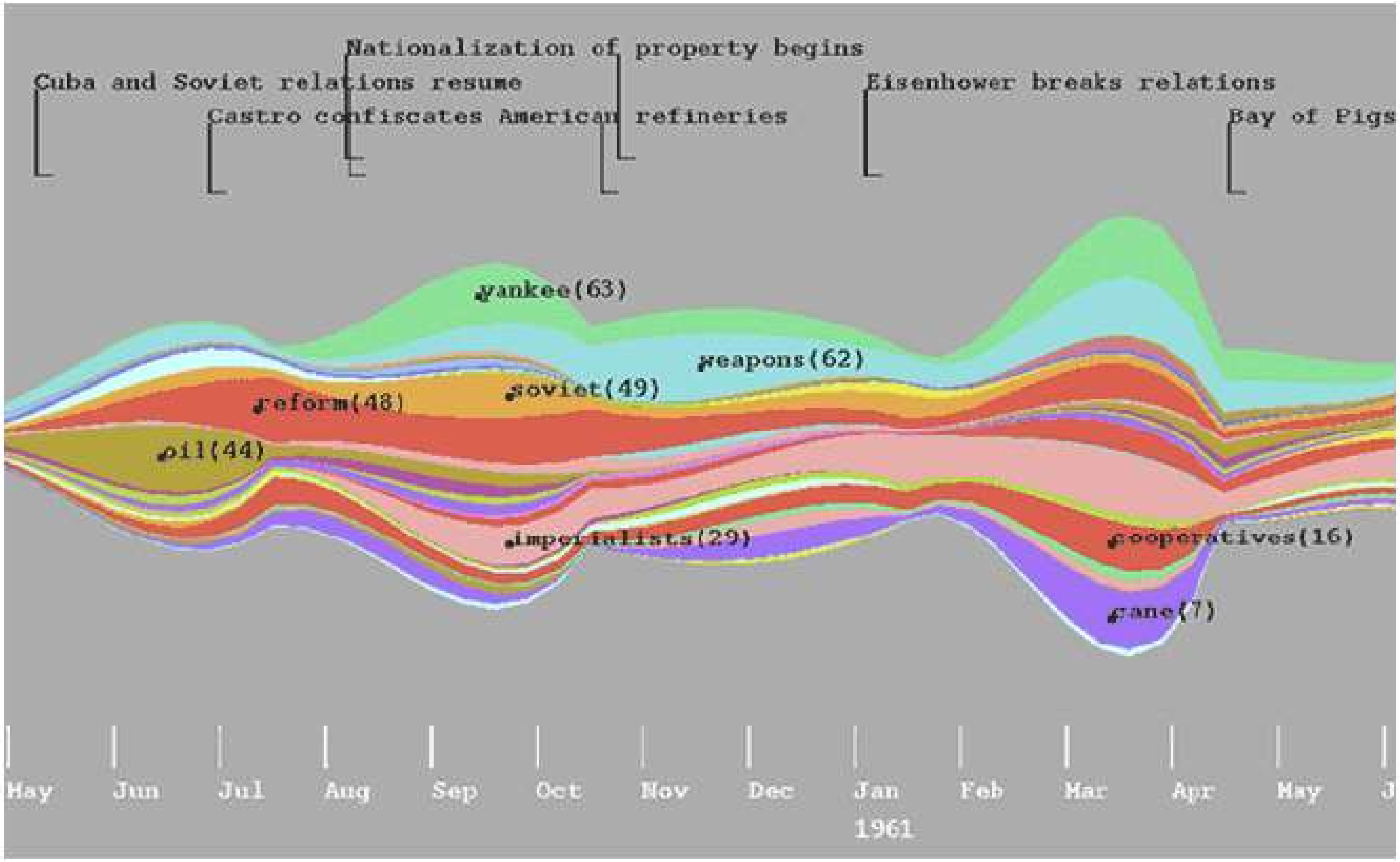
WebTheme provides a way to investigate and understand large volumes of textual information. It has the ability to harvest data from the World Wide Web using search terms, or by following links derived from user specified URLs. Users can rapidly identify themes and concepts found among thousands of pages of text, and then further explore areas of interest.


Galaxy Viewer - climate_change_digest
Mode
Tools
Settings
(1078) EMISSION REDUCTION (FEBRUARY 1989)
water power
gef back
efficiency tole
(954) FRAMEWORK CONVENTION WORKSHOP (OCTOBER 1989)
(919) CAIRO COMPACT ISSUED (FEBRUARY 1990)
proceedings eds
protocols cfcs
water species
stratospheric sea
(1049) OF GENERAL INTEREST (MARCH 1989)
(902) GLOBAL CHANGE ENCYCLO
(606) NASA DATA SYSTEM (JUNE 1992)
surface geoph
(307) SEA LEVEL SCIENC
geophys measurements
(341) ARCTIC TROPOSPHERIC CHEMISTRY (JA
stratospheric stratosphere
Group Window
Groups
Operations
ocean (223)
atmosphere (264)
space (121)
New
Select
Delete
Java Applet Window

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

ThemeRiver

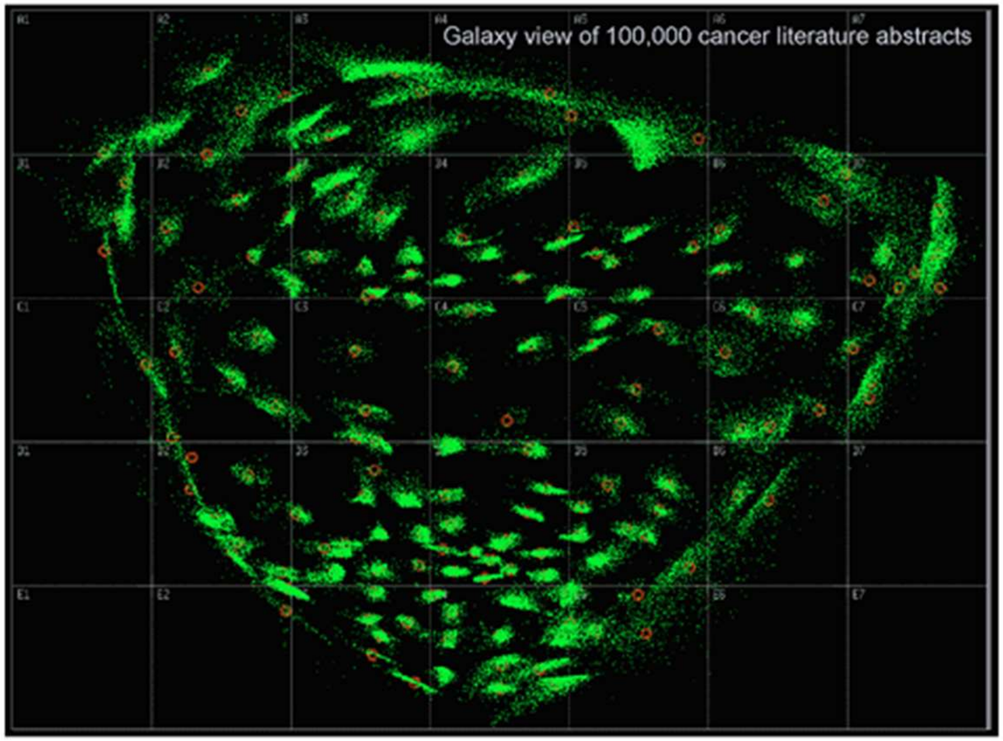


Temporal Variation of Themes

Susan Havre, Elizabeth Hetzler, Paul Whitney, and Lucy Nowell. *ThemeRiver: Visualizing thematic changes in large document collections*. IEEE Transactions on Visualization and Computer Graphics, 2002.

Galaxies

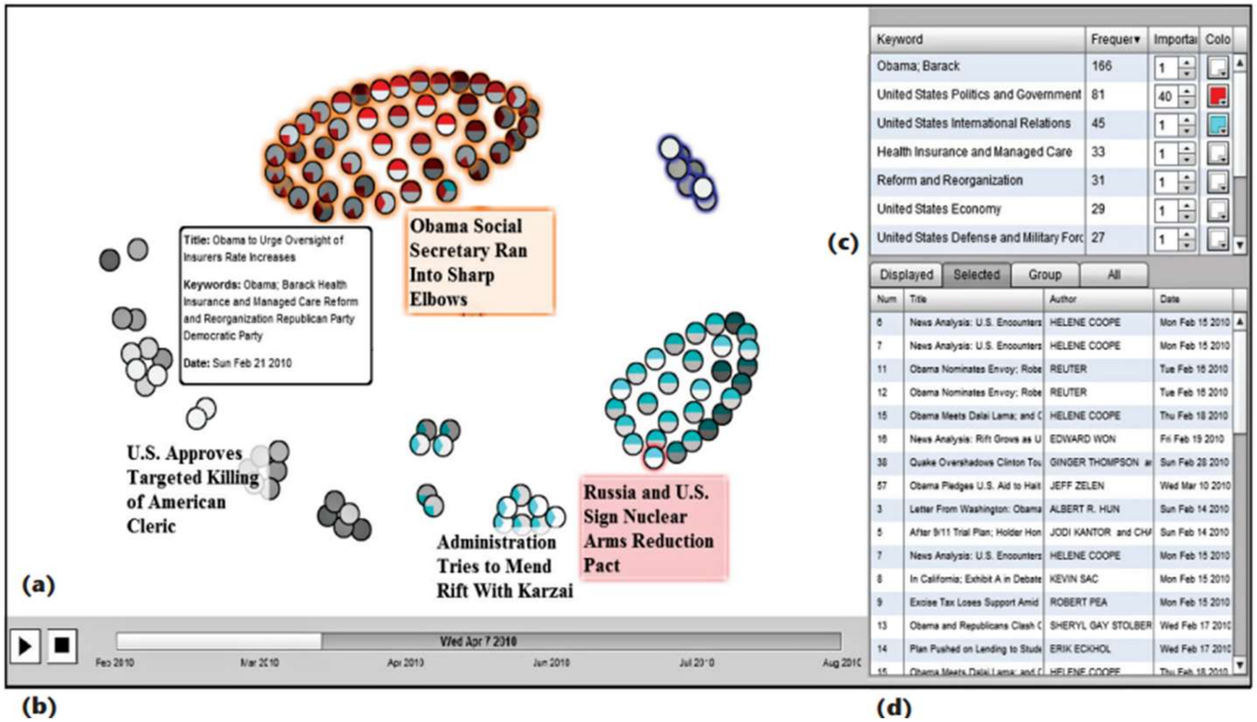
- The Galaxies visualization uses the image of stars in the night sky to represent a set of documents.
- Each document is represented by a single "docustar."
- Closely related documents cluster together while unrelated documents are separated by large distances.
- Several analytical tools are provided with Galaxies to allow users to investigate the document groupings, query the document contents, and investigate time-based trends.



Visualizing Text Streams

- Streamit, a dynamic visualization system for exploring text streams.
- Streamit is based on a dynamic force-directed simulation into which text documents are continuously inserted.
- A dynamic 2D display presents the incoming documents
- as a mass particle moving inside a 2D visualization domain,
- Users can explore documents and document clusters on the basis of keywords or topics.
- They can discover emerging patterns online by monitoring the real-time display.
- They can also examine historical data's temporal evolution through animations that play back past streams.

Streamit


Obama Social Secretary Ran Into Sharp Elbows
U.S. Approves Targeted Killing of American Cleric
Administration Tries to Mend Rift With Karzai
Russia and U.S. Sign Nuclear Arms Reduction Pact
Wed Apr 7 2010
Feb 2010
Mar 2010
Apr 2010
Jun 2010
Jul 2010
Aug 2010
(a)
(b)
(c)
(d)
Displayed
Selected
Group
All

Title: Obama to Urge Oversight of Insurers Rate Increases

Keywords: Obama; Barack Health Insurance and Managed Care Reform and Reorganization Republican Party Democratic Party

Date: Sun Feb 21 2010

Keyword	Frequer▾	Importa	Colo
Obama; Barack	166	1	□
United States Politics and Government	81	40	🟥
United States International Relations	45	1	🟦
Health Insurance and Managed Care	33	1	□
Reform and Reorganization	31	1	□
United States Economy	29	1	□
United States Defense and Military Forc	27	1	□

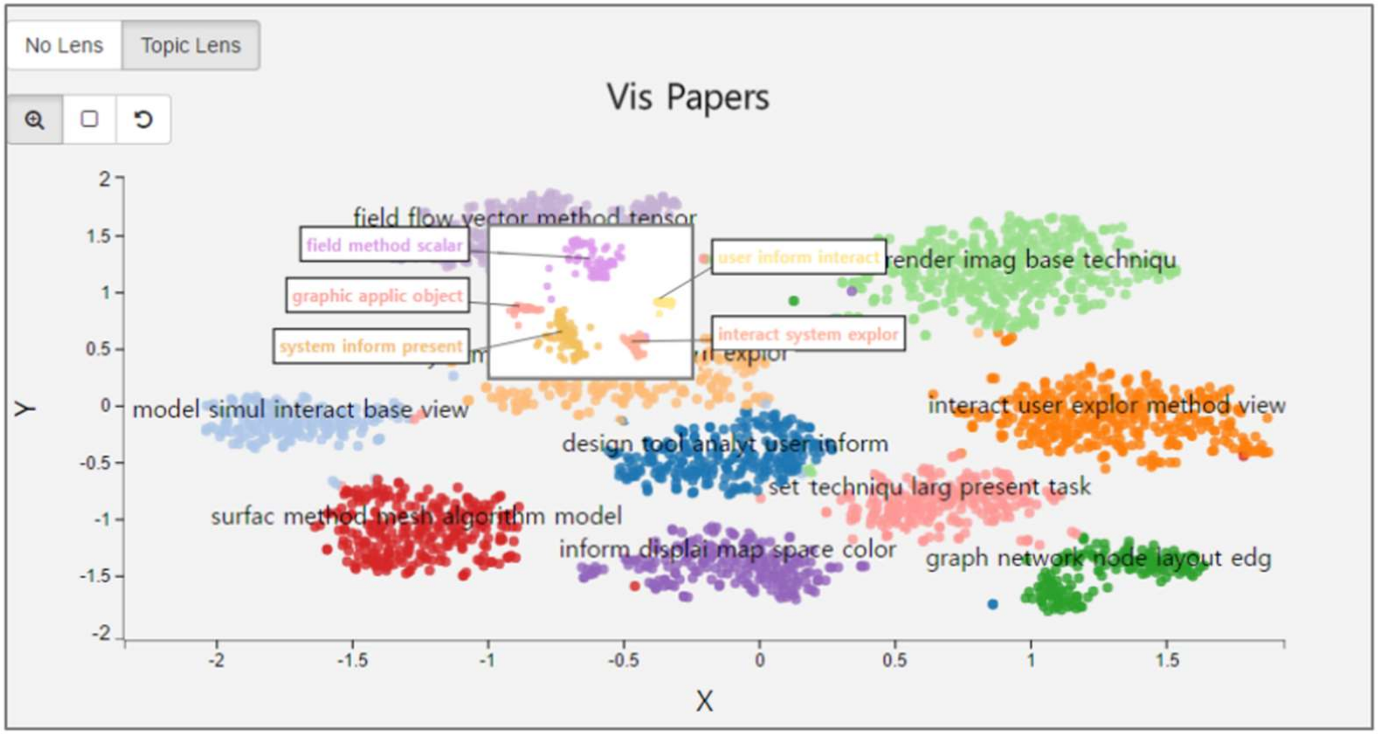
Num	Title	Author	Date
6	News Analysis: U.S. Encounters	HELENE COOPE	Mon Feb 15 2010
7	News Analysis: U.S. Encounters	HELENE COOPE	Mon Feb 15 2010
11	Obama Nominates Envoy; Robe	REUTER	Tue Feb 16 2010
12	Obama Nominates Envoy; Robe	REUTER	Tue Feb 16 2010
15	Obama Meets Dalai Lama; and (	HELENE COOPE	Thu Feb 18 2010
16	News Analysis: Rift Grows as U	EDWARD WON	Fri Feb 19 2010
38	Quake Overshadows Clinton Tou	GINGER THOMPSON ar	Sun Feb 28 2010
57	Obama Pledges U.S. Aid to Hait	JEFF ZELEN	Wed Mar 10 2010
3	Letter From Washington: Obama	ALBERT R. HUN	Sun Feb 14 2010
5	After 9/11 Trial Plan; Holder Hon	JODI KANTOR and CH/	Sun Feb 14 2010
7	News Analysis: U.S. Encounters	HELENE COOPE	Mon Feb 15 2010
8	In California; Exhibit A in Debate	KEVIN SAC	Mon Feb 15 2010
9	Excise Tax Loses Support Amid	ROBERT PEA	Mon Feb 15 2010
13	Obama and Republicans Clash (	SHERYL GAY STOLBER	Wed Feb 17 2010
14	Plan Pushed on Lending to Stude	ERIK ECKHOL	Wed Feb 17 2010
15	Obama Meets Dalai Lama; and (	HELENE COOPE	Thu Feb 18 2010

Streamit's interface. (a) The main window displays the particles' movement. (b) The animation control panel lets users navigate through the simulation. The (c) keyword table and (d) document table are synchronized with the particle stream to maintain an up-to-date list of the keywords and documents

Jamal Alsakran, Yang Chen, Dongning Luo, Ye Zhao, Jing Yang, Wenwen Dou, and Shixia Liu. *Real-Time Visualization of Streaming Text with a Force-Based Dynamic System*. IEEE Computer Graphics and Applications, 2012.

Topic Lens (1/2)

TopicLens is a novel interaction technique that allows a user to dynamically explore data through a lens interface where topic modeling and the corresponding 2D embedding are efficiently computed on the fly.


No Lens
Topic Lens
Vis Papers
field flow vector method tensor
field method scalar
graphic applic object
system inform present
user inform interact
render imag base techniqu
interact system explor
explor
model simul interact base view
interact user explor method view
design tool analyt user inform
set techniqu larg present task
surfac method mesh algorithm model
inform displai map space color
graph network node layout edg
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
Y
-2
-1.5
-1
-0.5
0
0.5
1
1.5
X

Topic Lens (2/2)

- The system initially performs topic modeling and visualizes documents as a scatterplot where the document coordinates are determined by a 2D embedding method
- To generate the 2D scatterplot of documents, a supervised version of t-distributed stochastic neighbor embedding (t-SNE) is used.
- The topic cluster memberships are color-coded.
- The representative keywords are shown in the center of each topic cluster.
- When moving the TopicLens (shown as a small rectangle), we dynamically recompute the topic model and 2D embedding in real time on those documents captured within the lens, revealing their finer-grained topical structure and their visual overview.
- The representative keywords are visualized just outside of the lens pointing to the center of each topic cluster

- Minjeong Kim, Kyeongpil Kang, Deokgun Park, Jaegul Choo, and Niklas Elmqvist. *TopicLens: Efficient Multi-Level Visual Topic Exploration of Large-Scale Document Collections*. IEEE Transactions on Visualization and Computer Graphics, 2017.

Reading

- Search User Interfaces (Marty Hearst)
  - Chapter 10: Information Visualization for Search Interfaces
  - Chapter 11: Information Visualization for Text Analysis
  - http://searchuserinterfaces.com/book/
- Viegas FB, Wattenberg M, Feinberg J. Participatory visualization with wordle IEEE Transactions on Visualization and Computer Graphics 2009
- Carsten Görg, Zhicheng Liu, Jaeyeon Kihm, Jaegul Choo, Haesun Park, and John Stasko. *Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw*. IEEE Transactions on Visualization and Computer Graphics, 2013.