Text and Documents

CS 4460 - Information Visualization

Jim Foley, some material courtesy John Stasko. Some examples from Marti Hearst, *Search User Interfaces*, Cambridge University Press, 2009

Last update: October 2016

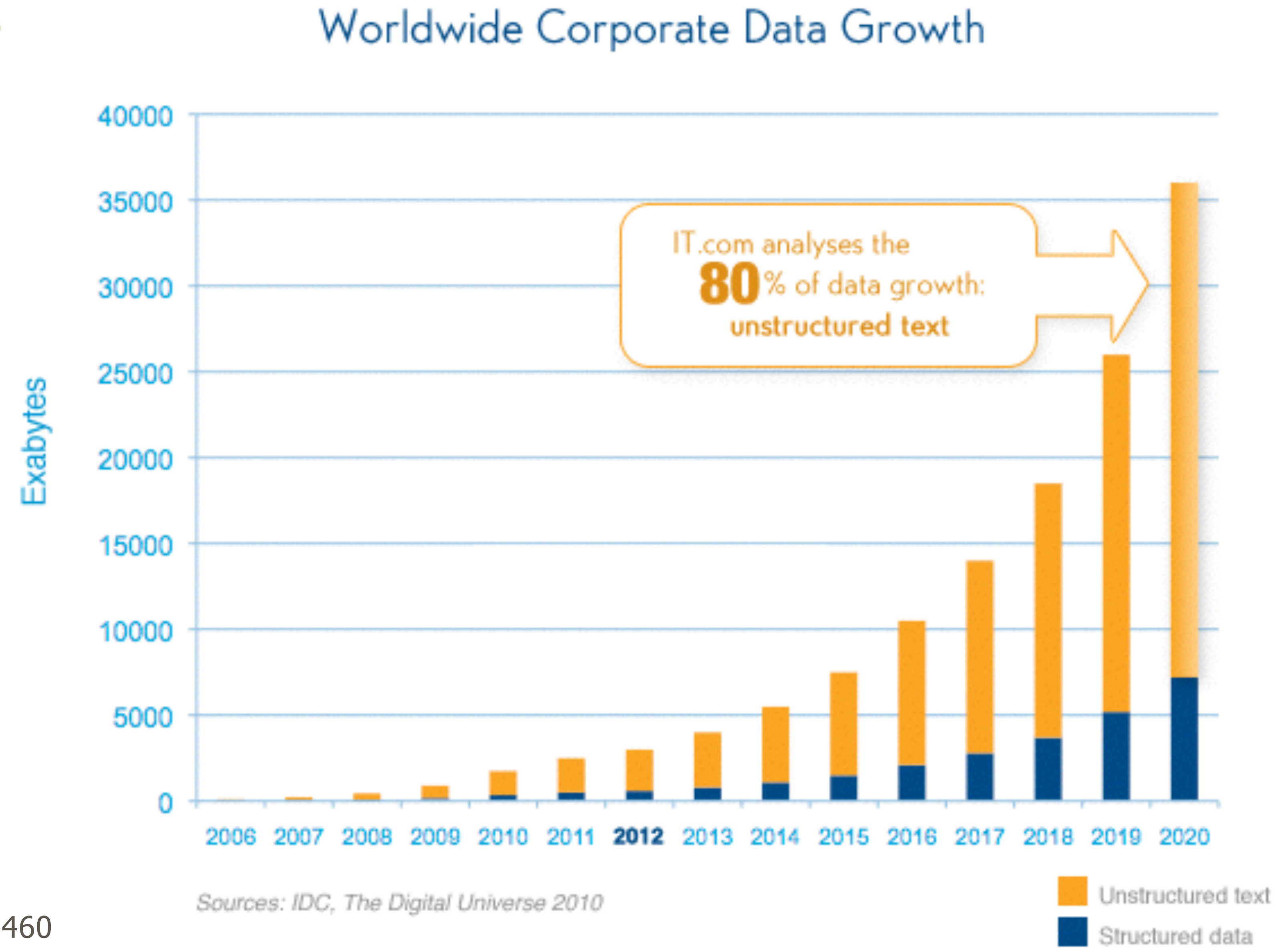
1. You have the following information on 387 articles published in 65 issues of a newsletter for in the months of June, July and August for a given year. Each newsletter contains 5 to 7 articles

- Date published: day/month/year
- Subject category of article: Politics, Sports, Economics, Health, Science, Other
- Author of article: Walt, Mary, John, Beth
- Focus of article: US, global, not known

Sketch a dashboard (several potentially linked visualizations) to convey as much of the above information as possible about the 387 articles. Ideally, you should convey all of the information.

2. There are several programs to detect plagiarism in term papers, by comparing a submitted term paper to a large collection of previously-submitted term papers. If a possible case of plagiarism is found, we need a way to show how the submitted paper is similar to a previously-submitted paper. Sketch a way to do this.

Who Cares about Text?



Worldwide Corporate Data Growth

IT.com analyses the **80**% of data growth: unstructured text

Sources: IDC, The Digital Universe 2010

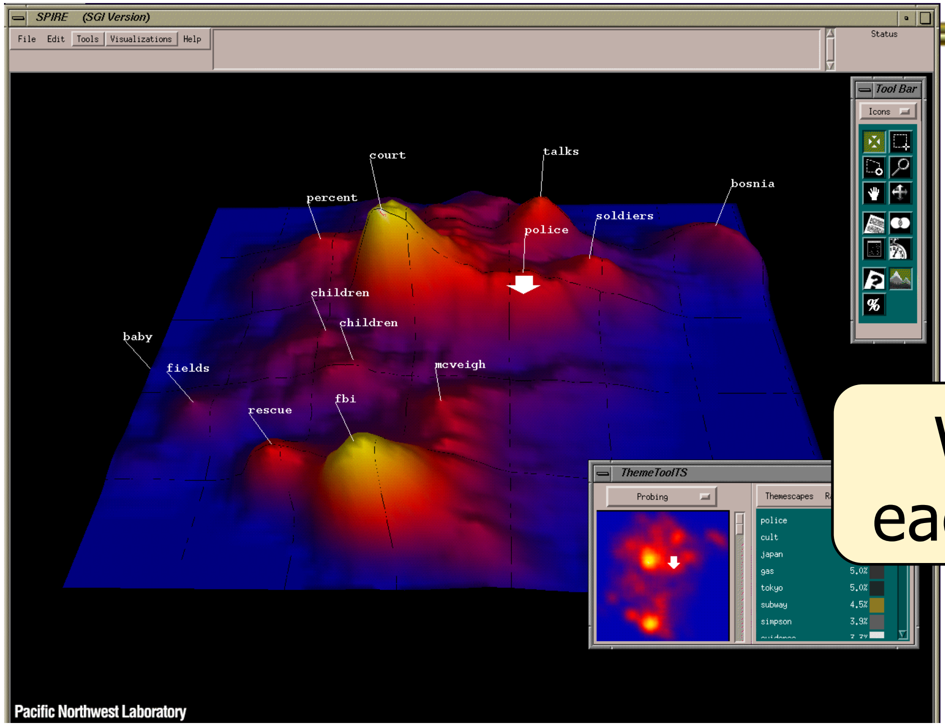
The Key Question for InfoVis

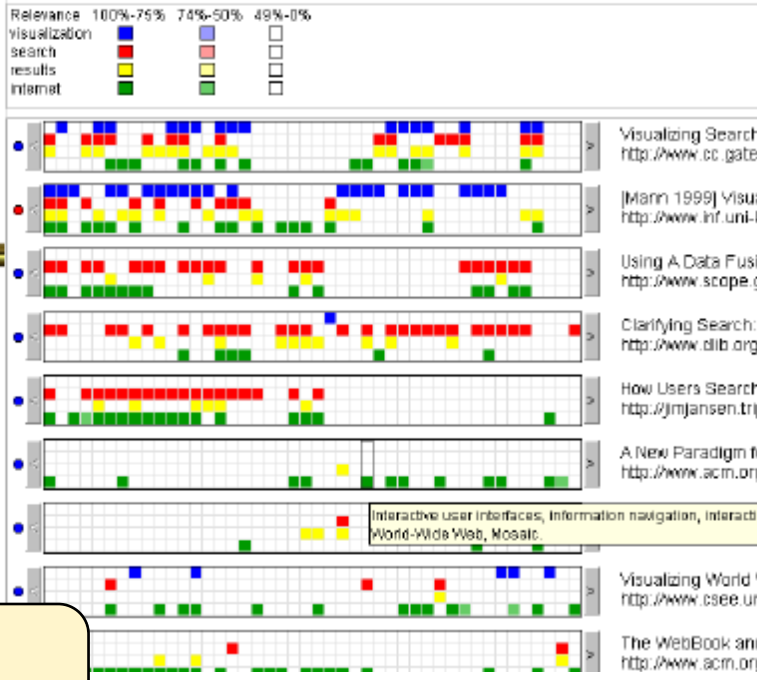
- How can InfoVis help users in gathering, understanding, using information from
  - Document collections (*macro-level*)?

    Everything on the web, library, …
  - Individual (or a few) documents (*micro-level*)?

    Such as a thesaurus, or a book or speech, Shakespeare, Bible, Qur'an, Torah, ….
- Documents include email, all social media, news, books, magazines, …..

Text InfoViz Examples





Relevance	100%-75%	74%-50%	49%-0%
visualization	🟦	🟦	□
search	🟥	🟥	□
results	🟨	🟨	□
internet	🟩	🟩	□

Visualizing Search
http://www.cc.gate

[Mann 1999] Visu
http://www.inf.uni-

Using A Data Fus
http://www.scope.

Clarifying Search:
http://www.dlib.org

How Users Search
http://jimjansen.tri

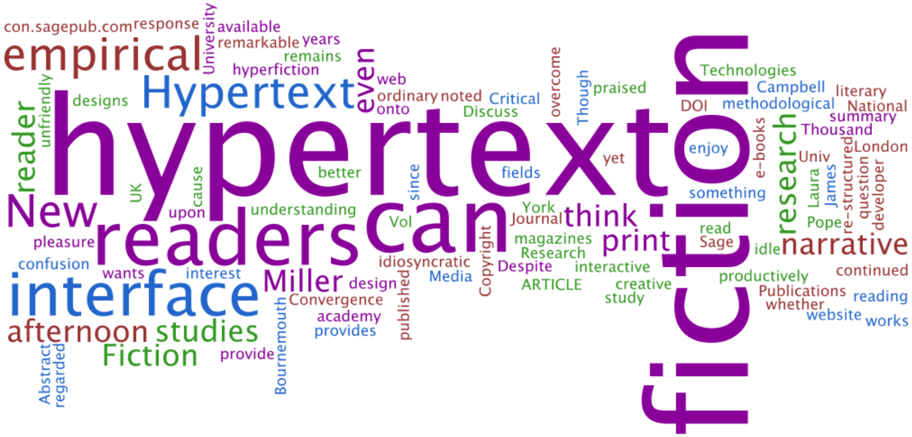
A New Paradigm f
http://www.acm.or

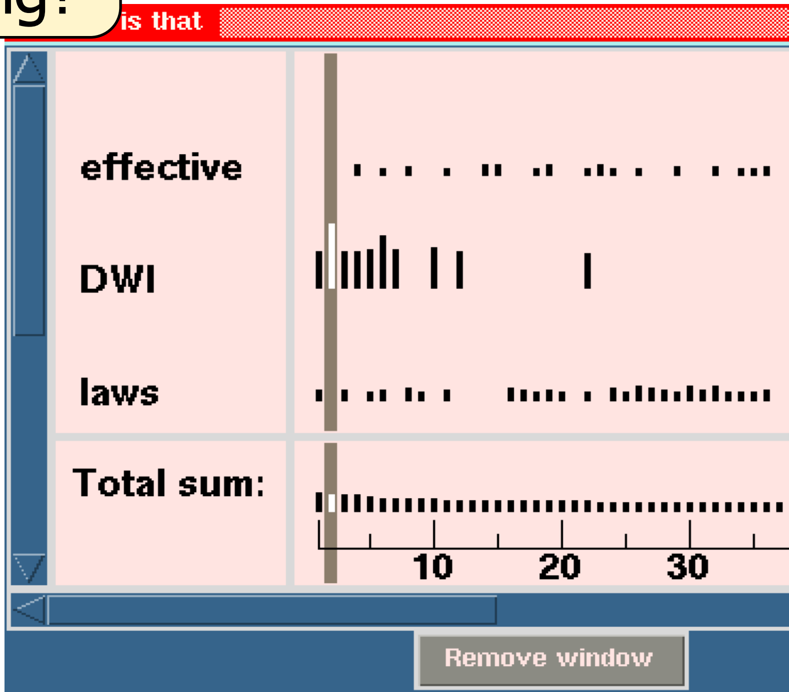
Interactive user interfaces, information navigation, interacti
World-Wide Web, Mosaic

Visualizing World
http://www.csee.u

The WebBook an
http://www.acm.or

What is each doing?





Some Questions about Document(s)

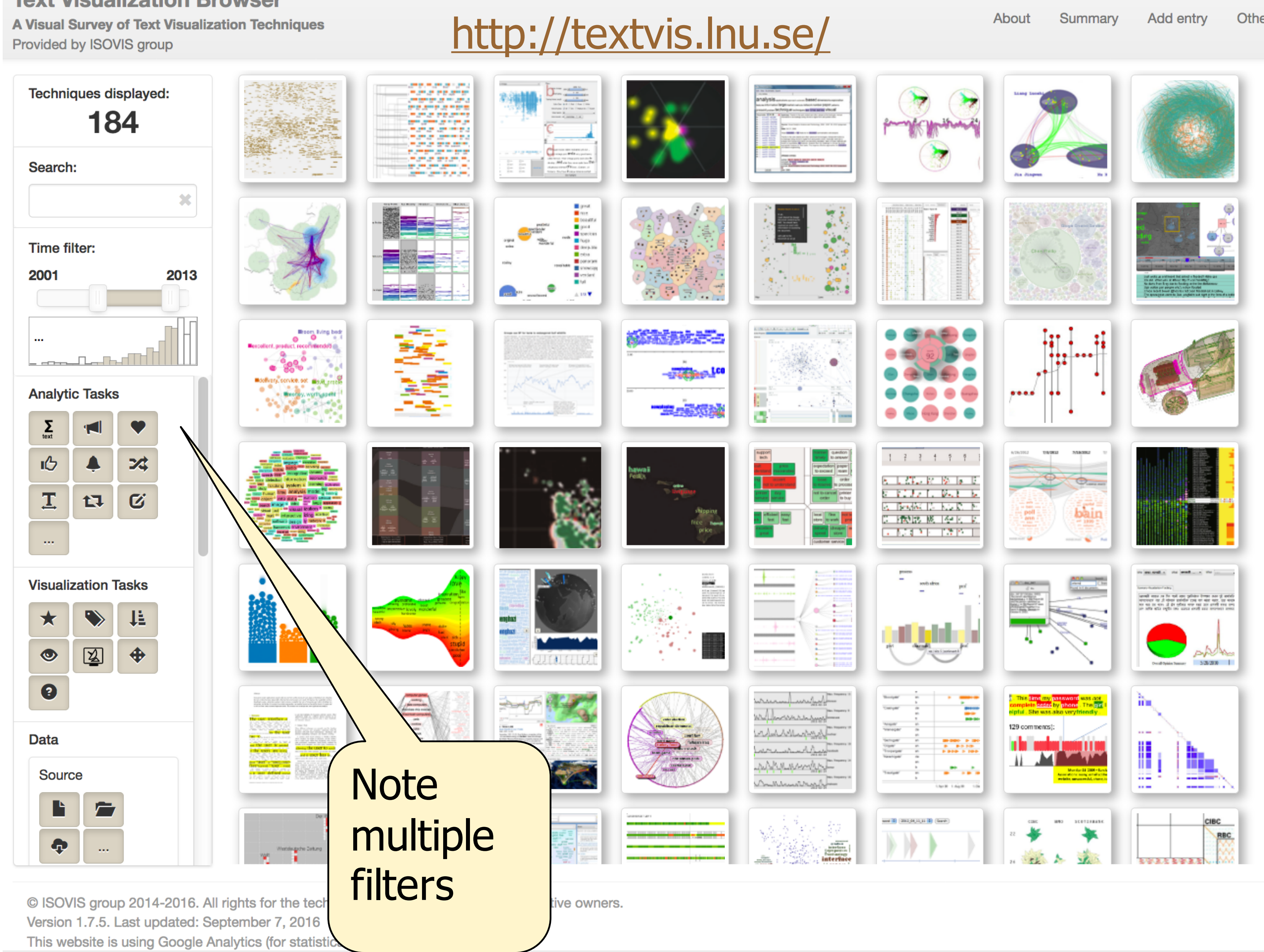
- Summarization: what are the main ideas in document(s)?
- Sentiment analysis: what sentiment(s) document(s) express?
- Trends: how do words/ideas/sentiment change over time in documents?
- How are words/ideas/themes distributed through one or several document?
- Grammar analysis: what grammatical patterns are used in document(s)
- How do two (or more) docs compare?

Some more micro-level, some more macro-level

More Questions about Document(s)

- Search: which docs have certain keywords/topics?
- How similar are two or more docs? In what ways?
- How do documents fit into a larger context?
- How does one document compare to or relate to other documents?
- In what docs (where in docs) is the word “inflation” used close to (next to) the word “spending?”
- Event analysis: what event(s) are reported in document(s)

http://textvis.lnu.se/



Text Visualization Browser

A Visual Survey of Text Visualization Techniques

Provided by ISOVIS group

About Summary Add entry Othe

Techniques displayed:

184

Search:

Time filter:

2001 2013

Analytic Tasks

Visualization Tasks

Data

Source

© ISOVIS group 2014-2016. All rights for the tech tive owners.

Version 1.7.5. Last updated: September 7, 2016

This website is using Google Analytics (for statistic

Note multiple filters

Challenge

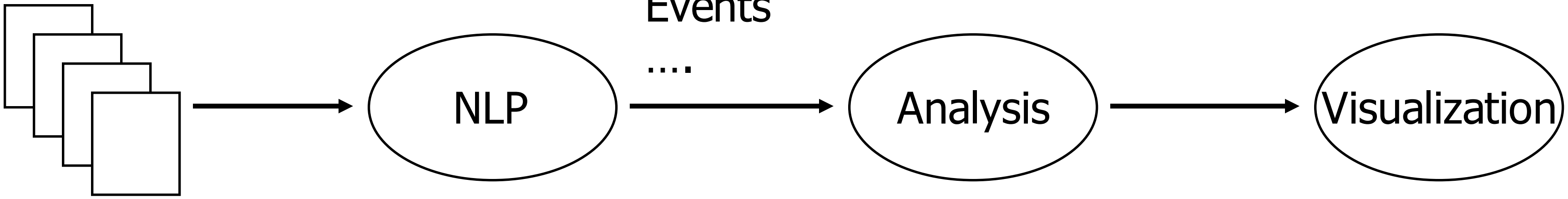
- Text is nominal data with a hugh (infinite) cardinality
  - Does not map to visual encodings as easily as nominal, ordinal & quantitative data
- It is not the text itself that gets encoded, it is information about the text
- Must first extract information from the text
  - Exception – text meta-data (more later)
- Visualizations are of the extracted info.

Pipeline for Text/Doc Visualization

Raw Data
(documents)

Per document
Keyword counts
Sentiments
Events
….

Geometric rep'n
of desired
information



NLP

Analysis

Visualization

Natural Language Processing
Keywords
Key phrases
Grammar
Sentiments
Statistics
Entity extraction
…..

Find relevant
docs
Compare docs

Challenge (Cont'd)

- Meta-data: data about data
- *Unstructured text* does NOT have any explicit meta-data.
  - Just that infinitely big collection of nominal data
  - NLP can extract some meta-data, such as dates
- Contrast to *structured text* of an on-line library with explicit meta-data such as
  - Author name
  - Year of publication
  - Title
  - ISBN number
  - Library of Congress number
  - Publisher name
  - Etc
- Some meta-data also nominal but lower cardinality than free text
  - Simplifies retrieval and visualization process.
- Will see some meta-data driven examples

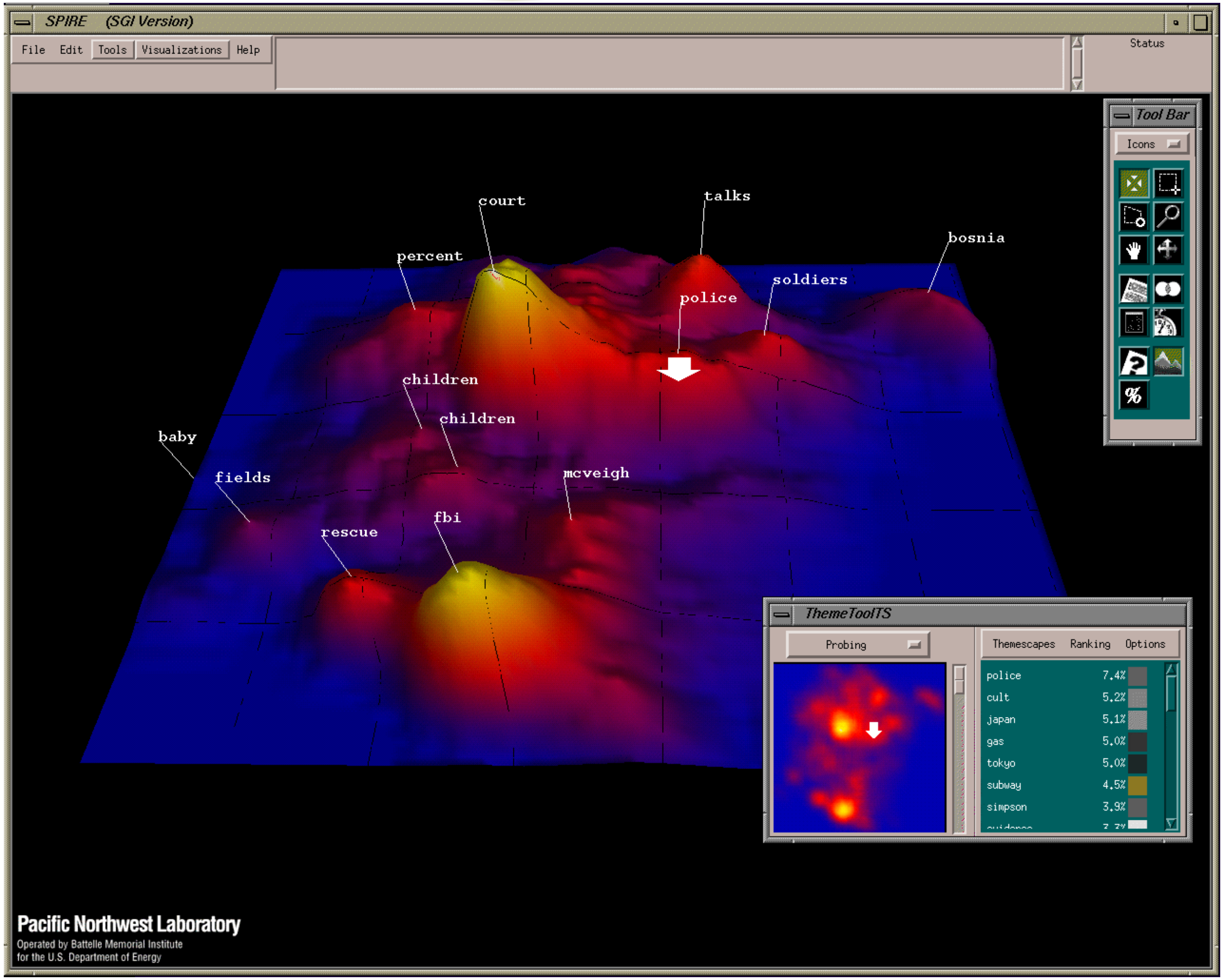
Outline

- Macro-level – searching larger document collections
  - Unstructured – no meta-data
  - Structured – explicit meta-data
- Micro-level
  - Inter-document methods for smaller document collections
    - How do retrieved documents relate to a query?
    - How do retrieved documents relate to one another?
  - Intra-document methods
    - Word usage, grammatical style, …
- With the caveat that some methods can be used in multiple ways

Macro-Level: Large Unstructured

- LARGE does not mean entire WWW!!
- A number of systems endeavor to give a "big picture view" – the "gist" of a large collection of documents
  - Themescape
  - WebThemes
  - Galaxies
  - ThemeRiver
  - Sentiment/emotion over time

Themescape

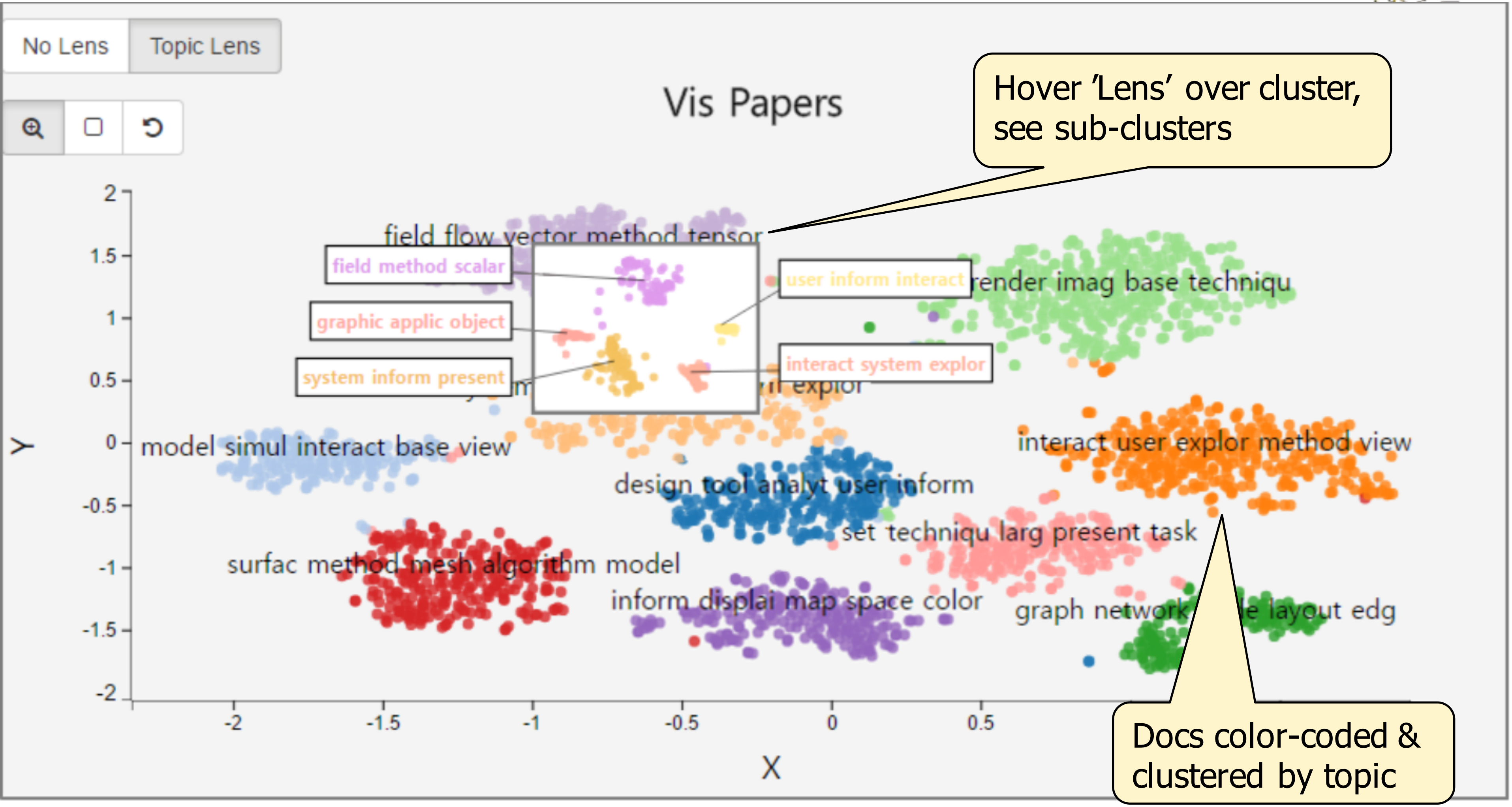


police	7.4%
cult	5.2%
japan	5.1%
gas	5.0%
tokyo	5.0%
subway	4.5%
simpson	3.9%
evidence	3.3%

Height/color encode document density

Received 2016 “test of time” award

TopicLens


No Lens
Topic Lens
Vis Papers
field flow vector method tensor
field method scalar
graphic applic object
system inform present
user inform interact
interact system explor
render imag base techniqu
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
-2
-1.5
-1
-0.5
0
0.5
Y
X

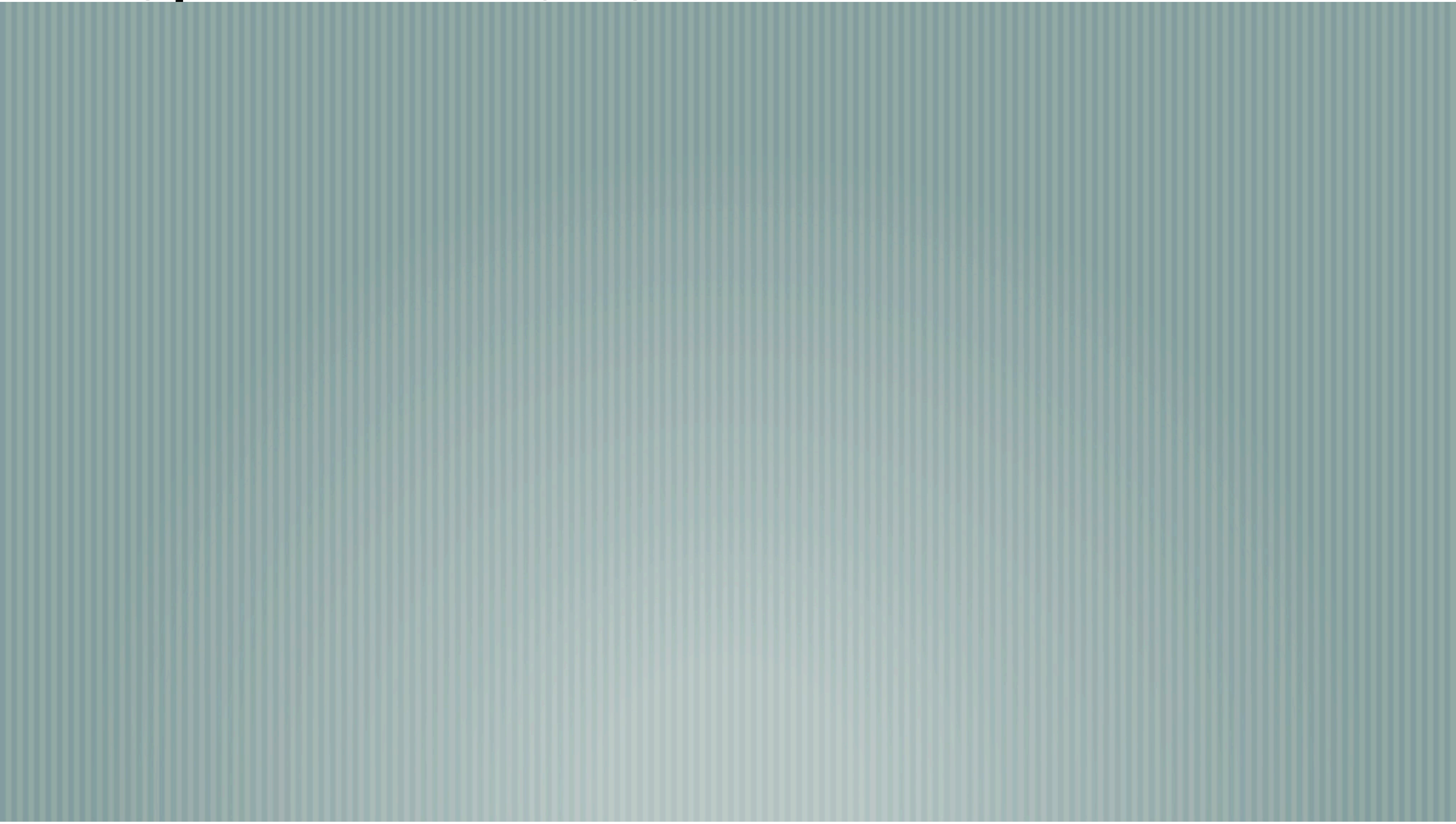
Hover 'Lens' over cluster, see sub-clusters

Docs color-coded & clustered by topic

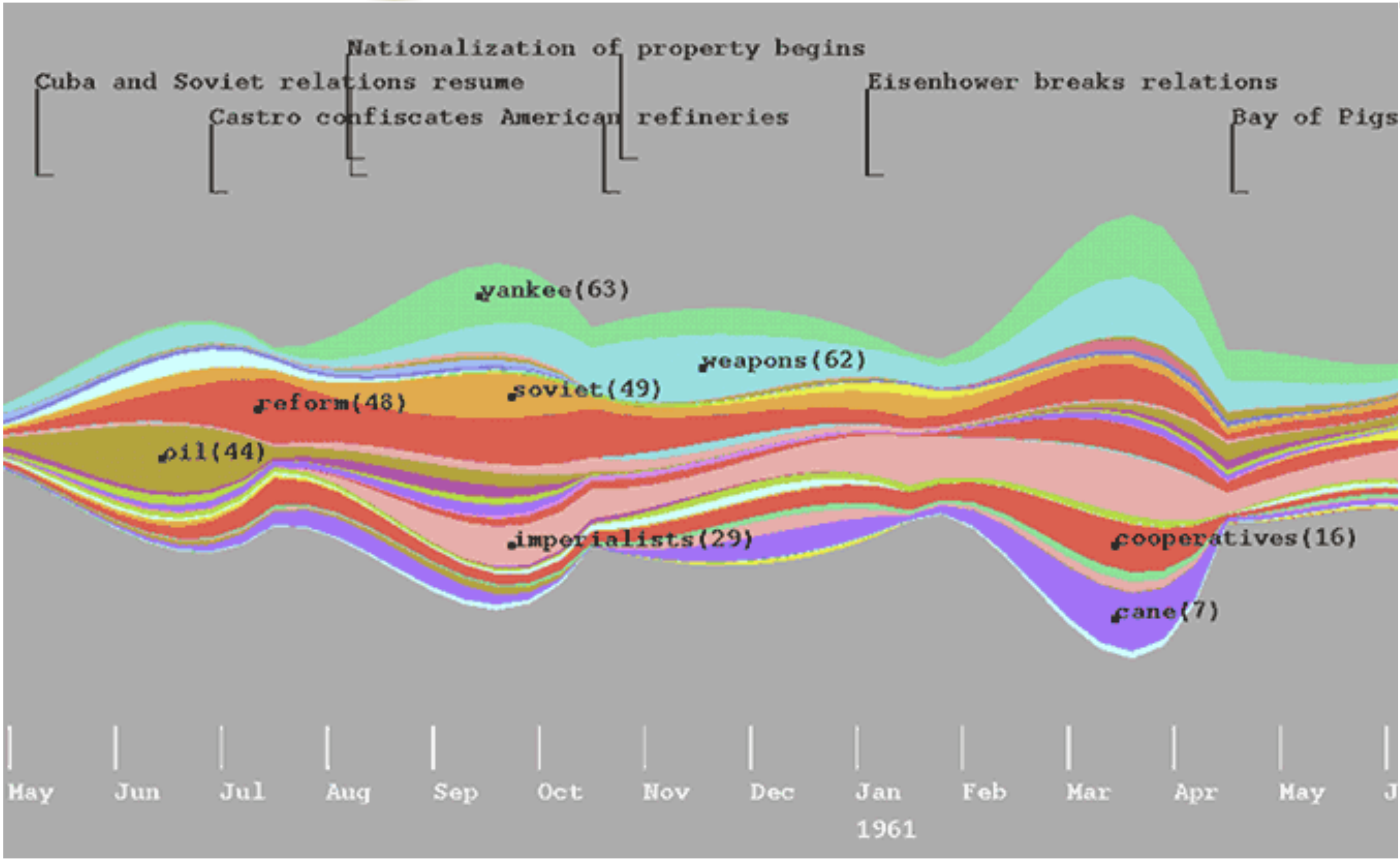
Video Next

Kim et al, TopicLens: Efficient Multi-Level Visual Topic Exploration of Large-Scale Document Collections, IEEE TVCG, Jan 2017

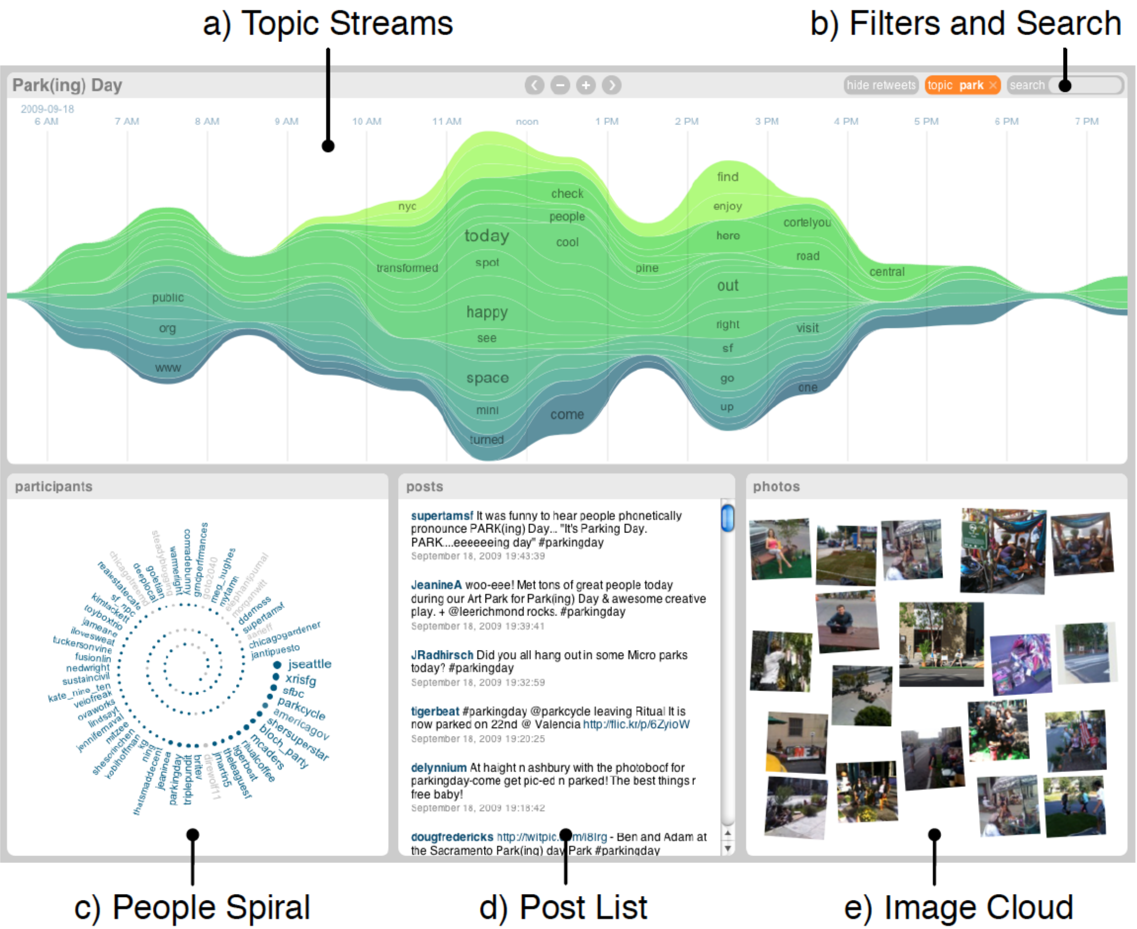
TopicLens Video



ThemeRiver


Nationalization of property begins
Cuba and Soviet relations resume
Castro confiscates American refineries
Eisenhower breaks relations
Bay of Pigs
yankee(63)
weapons(62)
soviet(49)
reform(48)
oil(44)
imperialists(29)
cooperatives(16)
cane(7)
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
J
1961

Visual Backchannel for Large-Scale Events



supertamsf It was funny to hear people phonetically pronounce PARK(ing) Day... "It's Parking Day. PARK....eeeeeeing day" #parkingday
September 18, 2009 19:43:39

JeanineA woo-eee! Met tons of great people today during our Art Park for Park(ing) Day & awesome creative play. + @leerichmond rocks. #parkingday
September 18, 2009 19:39:41

JRadhirsch Did you all hang out in some Micro parks today? #parkingday
September 18, 2009 19:32:59

tigerbeat #parkingday @parkcycle leaving Ritual It is now parked on 22nd @ Valencia http://flic.kr/p/6ZyioW
September 18, 2009 19:20:25

delynnium At haight n ashbury with the photoboof for parkingday-come get pic-ed n parked! The best things r free baby!
September 18, 2009 19:18:42

dougfredericks http://twitpic.com/i8lrg - Ben and Adam at the Sacramento Park(ing) day Park #parkingday

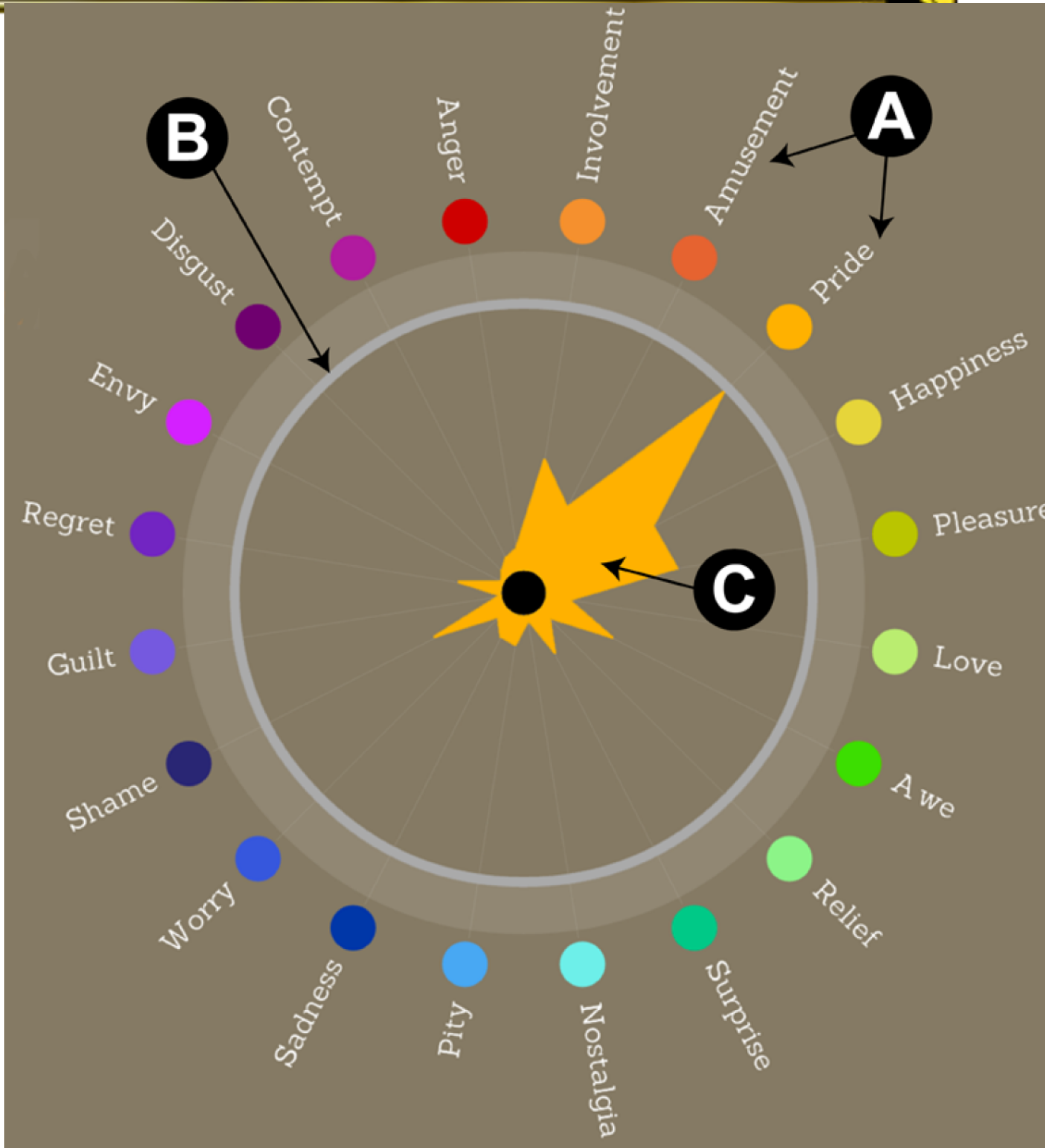
Fig. 1. The Visual Backchannel—here shown for Twitter posts about the event Park(ing) Day—consists of a) Topic Streams: a visualization representing topical development, b) controls for filtering and searching, c) People Spiral indicating the activity of participants, d) chronologically ordered list of posts, and e) Image Cloud displaying shared photos.

Emotions/trends/sentiments over Time

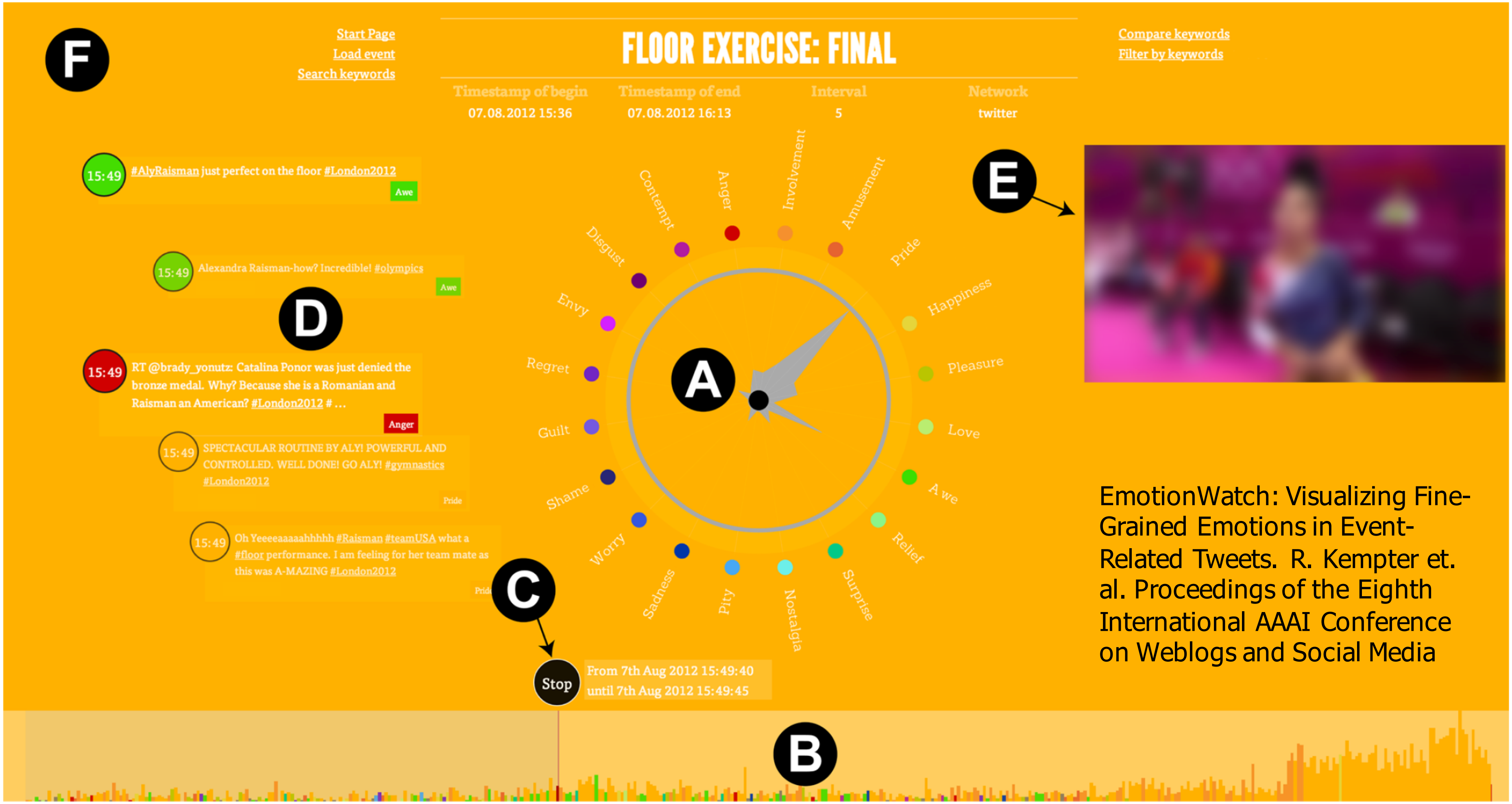
- Theme River style
- Many small multiples
- Animation
- Sparkclouds
- Parallel wordclouds

Emotions from tweets

- A: Twenty emotions
- B: Thickness = number of tweets analyzed
- C: Star (radar) plot of emotions



Visualizing Emotions in Event-Related Tweets



Start Page
Load event
Search keywords

FLOOR EXERCISE: FINAL

Timestamp of begin	Timestamp of end	Interval	Network
07.08.2012 15:36	07.08.2012 16:13	5	twitter

Compare keywords
Filter by keywords

15:49 #AlyRaisman just perfect on the floor #London2012 — Awe

15:49 Alexandra Raisman-how? Incredible! #olympics — Awe

15:49 RT @brady_yonutz: Catalina Ponor was just denied the bronze medal. Why? Because she is a Romanian and Raisman an American? #London2012 # … — Anger

15:49 SPECTACULAR ROUTINE BY ALY! POWERFUL AND CONTROLLED. WELL DONE! GO ALY! #gymnastics #London2012 — Pride

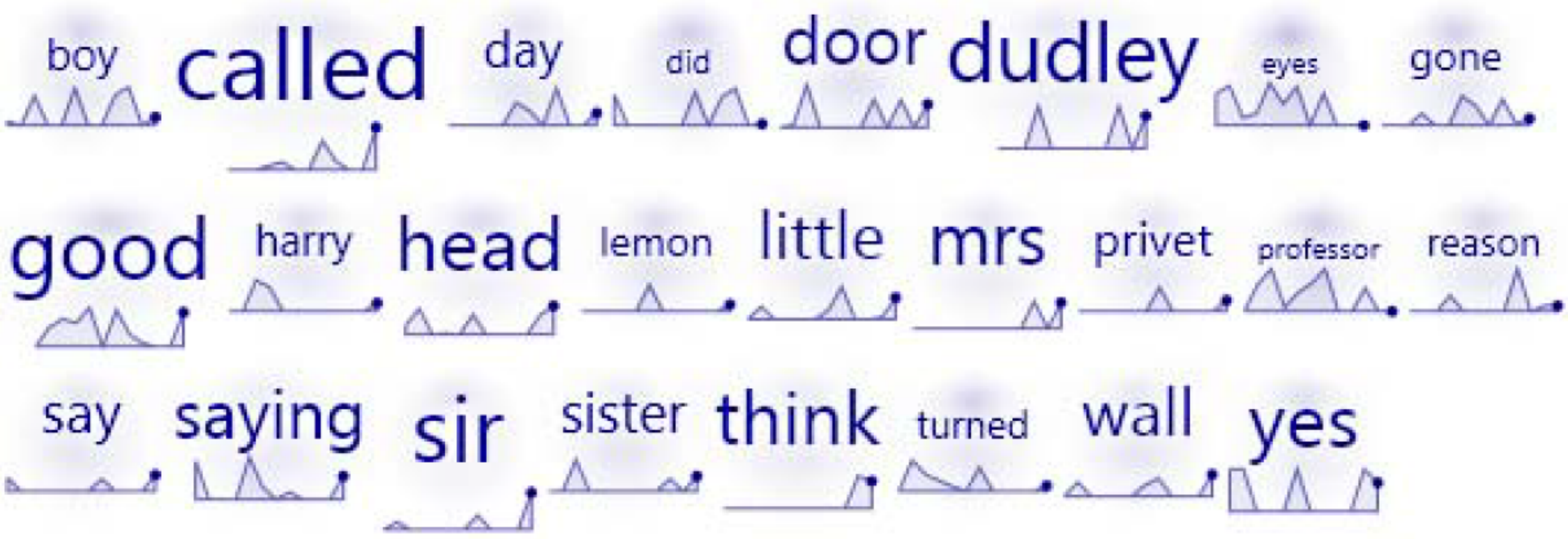
15:49 Oh Yeeeeaaaaahhhhh #Raisman #teamUSA what a #floor performance. I am feeling for her team mate as this was A-MAZING #London2012 — Pride

From 7th Aug 2012 15:49:40
until 7th Aug 2012 15:49:45

EmotionWatch: Visualizing Fine-Grained Emotions in Event-Related Tweets. R. Kempter et. al. Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media

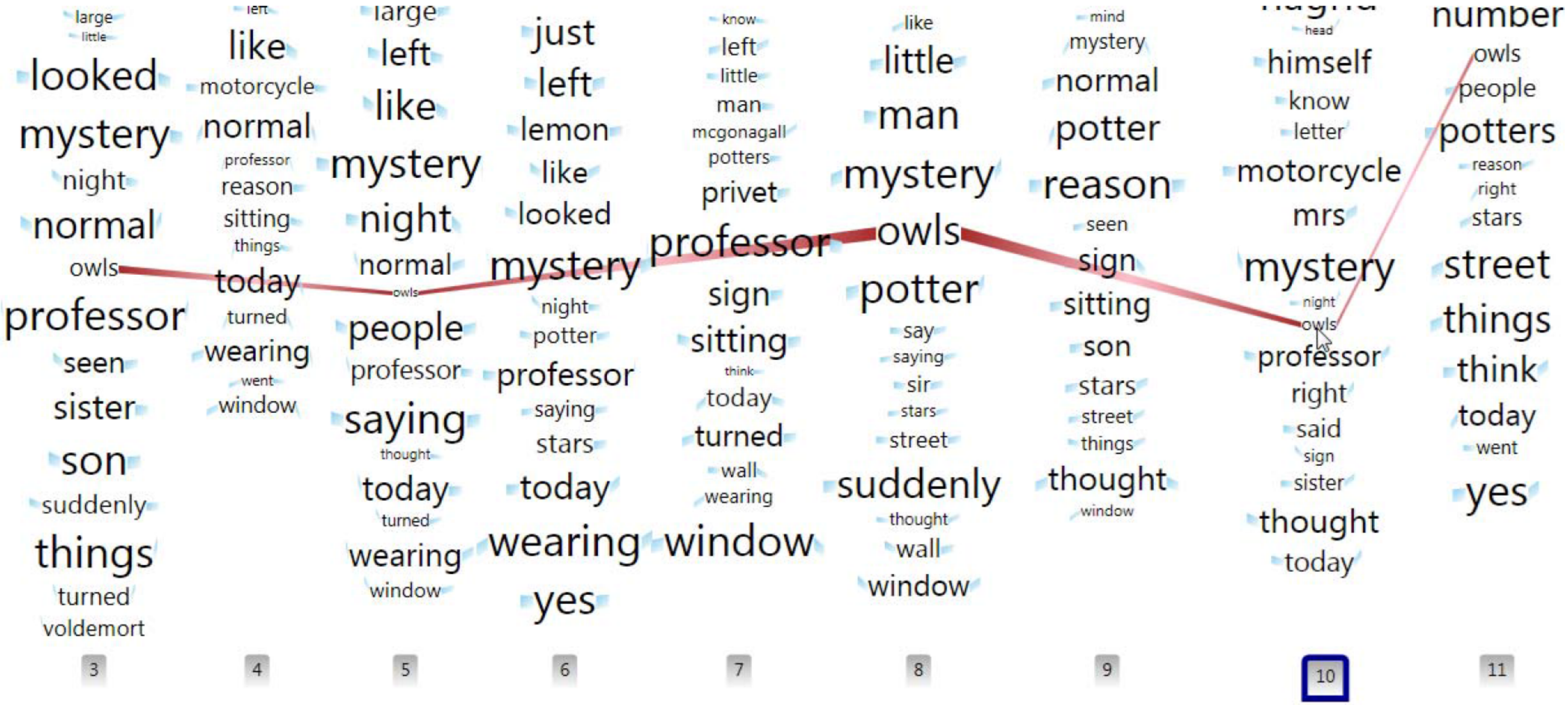
Figure 2: The detailed view showing the Women's Gymnastics Floor Exercise Final. A - Emotion wheel showing the emotion profile of the current time interval; B - Timeline visualizing the emotion flow; C - Button to stop/resume the animation; D - Tweets of the current time interval; E - Video; F - Background with color of the dominant emotion.

Trends via "Sparkclouds"



Sparkclouds: visualizing trends in tag clouds. Lee et al, IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, Nov/Dec 2010

Trends via Parallel Wordclouds



Pros/Cons

- Theme River style
- Many small multiples
- Animation
- Sparkclouds
- Parallel wordclouds

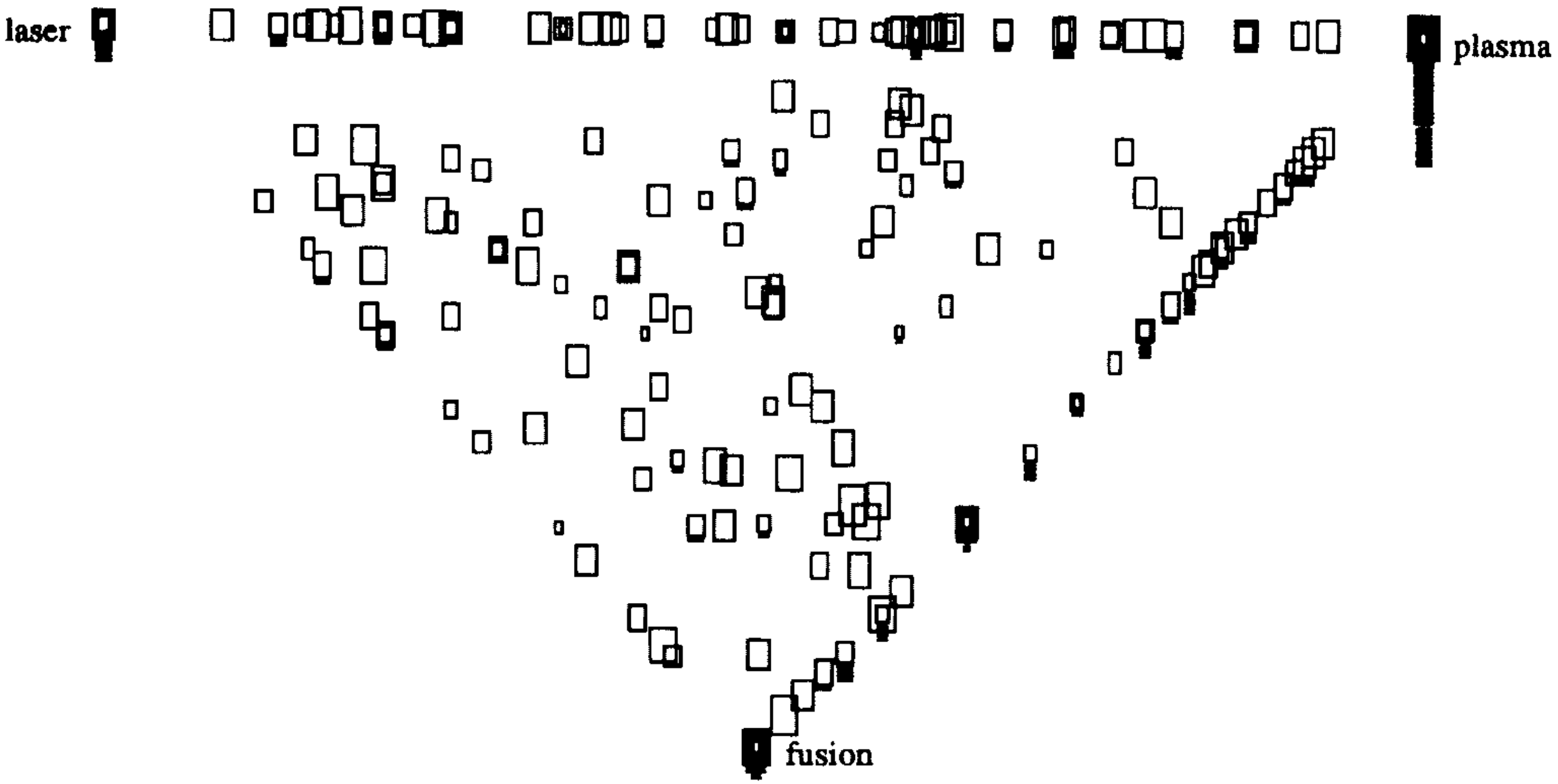
- Other ways to do it???

Understanding Relevance - Macro-level – Dozens or Hundreds

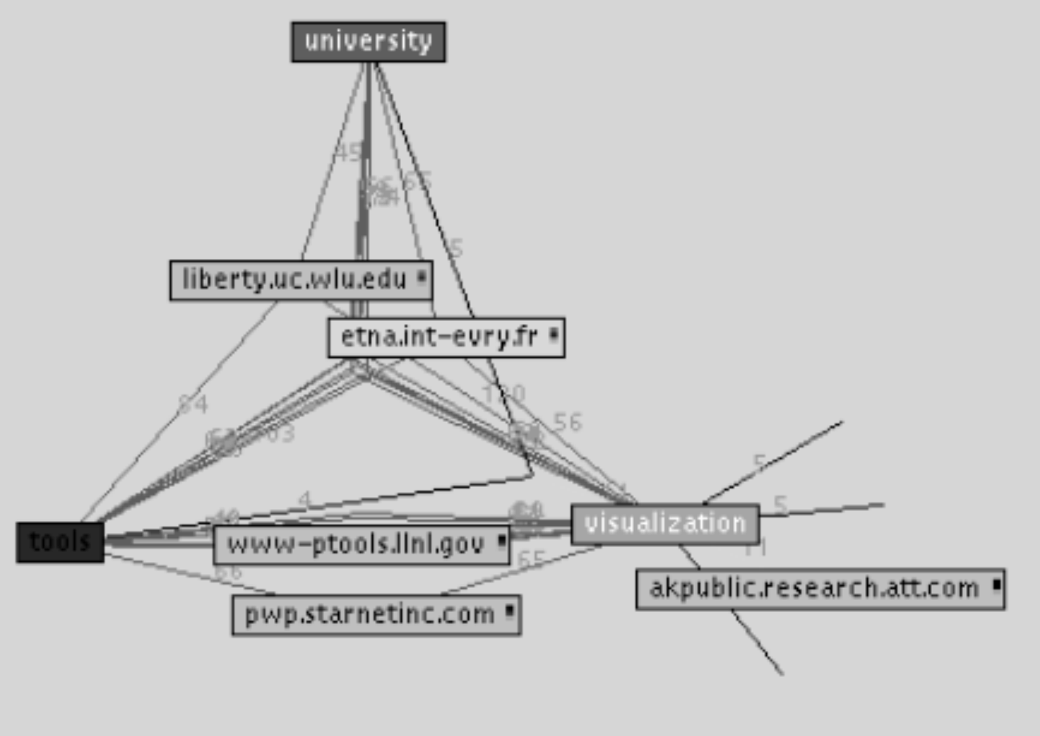
- Visualize Keywords and Doc’s
- Show relation of each Doc to Keywords
- “Similar” Doc’s cluster together

- VIBE
- SQUID
- JIGSAW
- Google
- Veerasamy’s work
- TileBars (dozen or less)

A VIBE Visualization



SQWID: Search Query Weighted Info Display (VIBE-like)

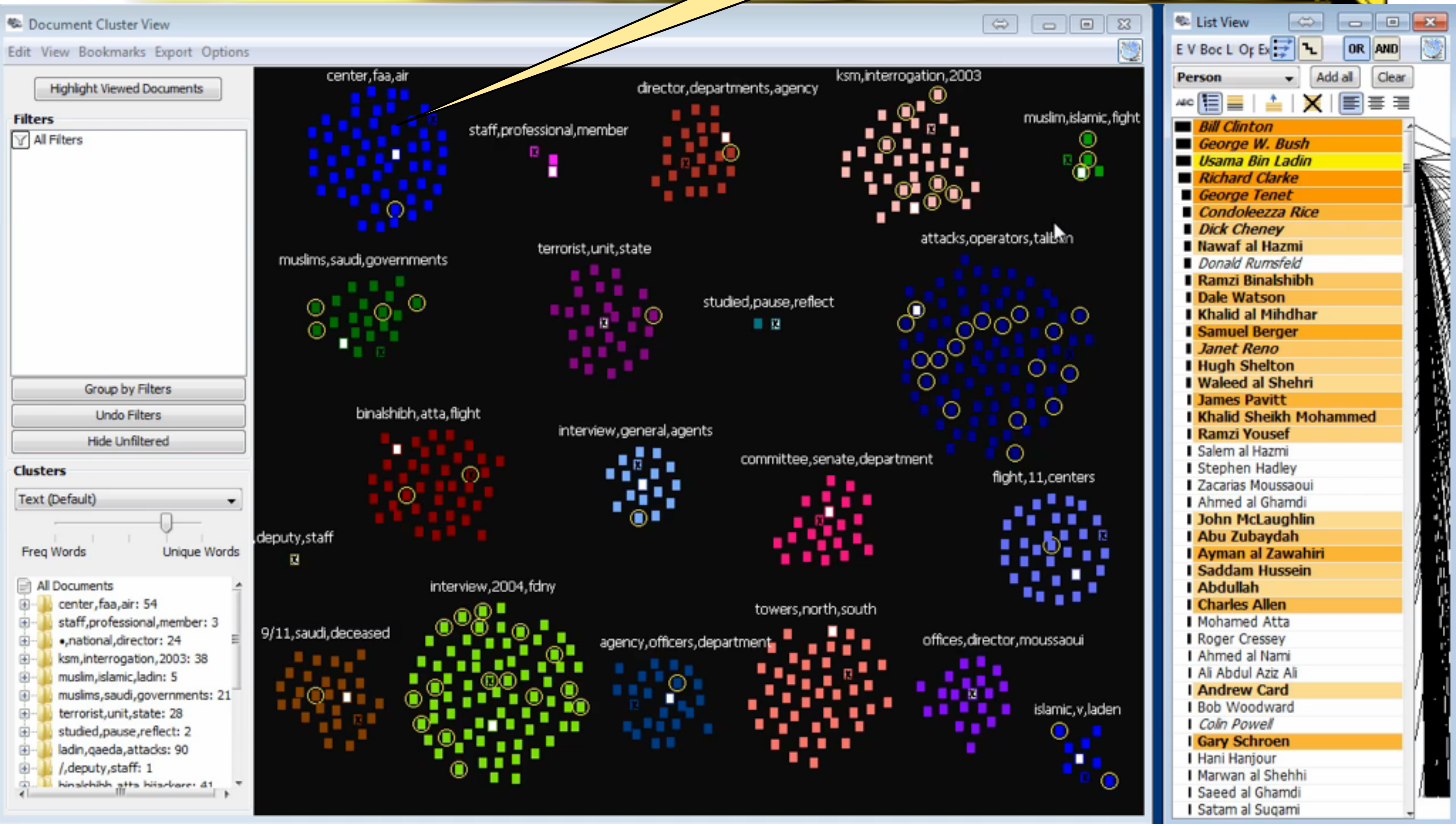


- Keywords “pull” Doc’s
  - (University, Visualization, Tools)
- Doc’s can go outside convex hull of keywords (unlike some other approaches)

McCrickard and Kehoe, Visualizing Search Results using SQWID, Poster paper in Proceedings of the 6th World Wide Web Conference (WWW6), Santa Clara CA, April 1997

Jigsaw

Related docs clustered together

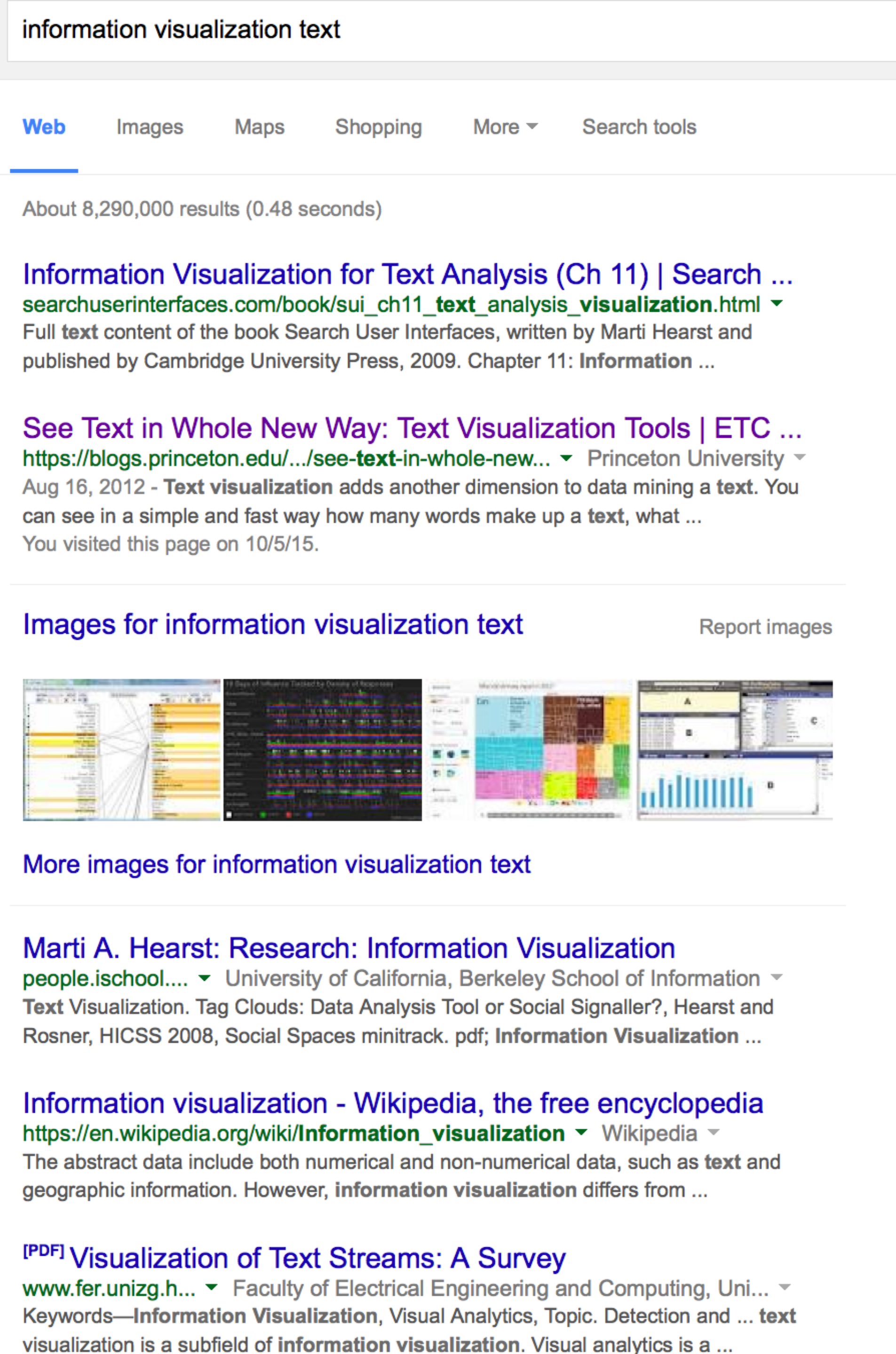


All Documents
- center,faa,air: 54
- staff,professional,member: 3
- •,national,director: 24
- ksm,interrogation,2003: 38
- muslim,islamic,ladin: 5
- muslims,saudi,governments: 21
- terrorist,unit,state: 28
- studied,pause,reflect: 2
- ladin,qaeda,attacks: 90
- /,deputy,staff: 1
- binalshibh,atta,hijackers: 41

Bill Clinton
George W. Bush
Usama Bin Ladin
Richard Clarke
George Tenet
Condoleezza Rice
Dick Cheney
Nawaf al Hazmi
Donald Rumsfeld
Ramzi Binalshibh
Dale Watson
Khalid al Mihdhar
Samuel Berger
Janet Reno
Hugh Shelton
Waleed al Shehri
James Pavitt
Khalid Sheikh Mohammed
Ramzi Yousef
Salem al Hazmi
Stephen Hadley
Zacarias Moussaoui
Ahmed al Ghamdi
John McLaughlin
Abu Zubaydah
Ayman al Zawahiri
Saddam Hussein
Abdullah
Charles Allen
Mohamed Atta
Roger Cressey
Ahmed al Nami
Ali Abdul Aziz Ali
Andrew Card
Bob Woodward
Colin Powell
Gary Schroen
Hani Hanjour
Marwan al Shehhi
Saeed al Ghamdi
Satam al Suqami

Problems with Google

- Query results given as text
- Difficult to understand:
  - Relationships between documents with the same relevance value
  - Relative relevance/irrelevance to the query
    - Only by position in list of search results
  - Relationships of matching documents to components of the query statement
  - Relationships between results of multiple queries



information visualization text

About 8,290,000 results (0.48 seconds)

Information Visualization for Text Analysis (Ch 11) | Search ...
searchuserinterfaces.com/book/sui_ch11_**text**_analysis_**visualization**.html
Full **text** content of the book Search User Interfaces, written by Marti Hearst and published by Cambridge University Press, 2009. Chapter 11: **Information** ...

See Text in Whole New Way: Text Visualization Tools | ETC ...
https://blogs.princeton.edu/.../see-**text**-in-whole-new... Princeton University
Aug 16, 2012 - **Text visualization** adds another dimension to data mining a **text**. You can see in a simple and fast way how many words make up a **text**, what ...
You visited this page on 10/5/15.

Images for information visualization text

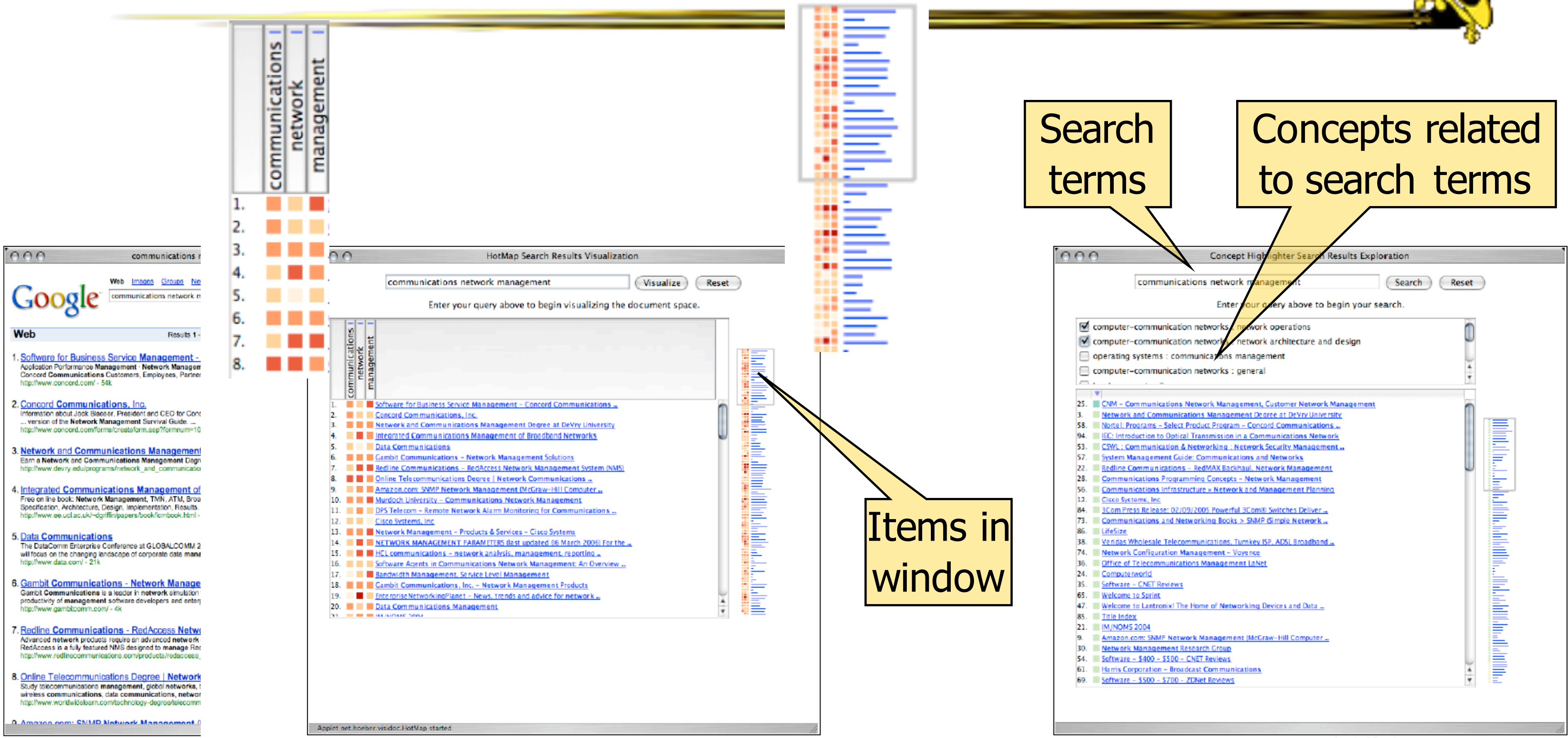
More images for information visualization text

Marti A. Hearst: Research: Information Visualization
people.ischool.... University of California, Berkeley School of Information
Text Visualization. Tag Clouds: Data Analysis Tool or Social Signaller?, Hearst and Rosner, HICSS 2008, Social Spaces minitrack. pdf; **Information Visualization** ...

Information visualization - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/**Information_visualization** Wikipedia
The abstract data include both numerical and non-numerical data, such as **text** and geographic information. However, **information visualization** differs from ...

[PDF] Visualization of Text Streams: A Survey
www.fer.unizg.h... Faculty of Electrical Engineering and Computing, Uni...
Keywords—**Information Visualization**, Visual Analytics, Topic. Detection and ... **text** visualization is a subfield of **information visualization**. Visual analytics is a ...

Adding InfoViz to Google



communications r

Web Images Groups Ne

communications network m

Web Results 1 -

1. Software for Business Service Management -
Application Performance Management · Network Managem
Concord Communications Customers, Employees, Partner
http://www.concord.com/ - 54k

2. Concord Communications, Inc.
Information about Jack Blasser, President and CEO for Conc
... version of the Network Management Survival Guide. ...
http://www.concord.com/forms/createform.asp?formnum=10

3. Network and Communications Management
Earn a Network and Communications Management Degr
http://www.devry.edu/programs/network_and_communicatio

4. Integrated Communications Management of
Free on line book: Network Management, TMN, ATM, Broa
Specification, Architecture, Design, Implementation, Results.
http://www.ee.ucl.ac.uk/~dgriffin/papers/book/icmbook.html -

5. Data Communications
The DataComm Enterprise Conference at GLOBALCOMM 2
will focus on the changing landscape of corporate data mana
http://www.data.com/ - 21k

6. Gambit Communications - Network Manage
Gambit Communications is a leader in network simulation
productivity of management software developers and enter
http://www.gambitcomm.com/ - 4k

7. Redline Communications - RedAccess Netw
Advanced network products require an advanced network
RedAccess is a fully featured NMS designed to manage Red
http://www.redlinecommunications.com/products/redaccess_

8. Online Telecommunications Degree | Network
Study telecommunications management, global networks,
wireless communications, data communications, networ
http://www.worldwidelearn.com/technology-degree/telecomm

9. Amazon.com: SNMP Network Management (

(

HotMap Search Results Visualization

communications network management

Visualize Reset

Enter your query above to begin visualizing the document space.

communications network management

1. Software for Business Service Management – Concord Communications ...
2. Concord Communications, Inc.
3. Network and Communications Management Degree at DeVry University
4. Integrated Communications Management of Broadband Networks
5. Data Communications
6. Gambit Communications – Network Management Solutions
7. Redline Communications – RedAccess Network Management System (NMS)
8. Online Telecommunications Degree | Network Communications ...
9. Amazon.com: SNMP Network Management (McGraw-Hill Computer ...
10. Murdoch University – Communications Network Management
11. DPS Telecom – Remote Network Alarm Monitoring for Communications ...
12. Cisco Systems, Inc
13. Network Management – Products & Services – Cisco Systems
14. NETWORK MANAGEMENT PARAMETERS (last updated 06 March 2006) For the ...
15. HCL communications – network analysis, management, reporting ...
16. Software Agents in Communications Network Management: An Overview ...
17. Bandwidth Management, Service Level Management
18. Gambit Communications, Inc. – Network Management Products
19. EnterpriseNetworkingPlanet – News, trends and advice for network ...
20. Data Communications Management

Applet net.hoeber.visidoc.HotMap started

(b) HotMap

Concept Highlighter Search Results Exploration

communications network management

Search Reset

Enter your query above to begin your search.

☑ computer-communication networks : network operations
☑ computer-communication networks : network architecture and design
☐ operating systems : communications management
☐ computer-communication networks : general

25. CNM – Communications Network Management, Customer Network Management
3. Network and Communications Management Degree at DeVry University
58. Nortel: Programs – Select Product Program – Concord Communications ...
94. IEC: Introduction to Optical Transmission in a Communications Network
53. CSWL : Communication & Networking : Network Security Management ...
57. System Management Guide: Communications and Networks
22. Redline Communications – RedMAX Backhaul, Network Management
28. Communications Programming Concepts – Network Management
56. Communications Infrastructure » Network and Management Planning
13. Cisco Systems, Inc
84. 3Com Press Release: 02/09/2005 Powerful 3Com® Switches Deliver ...
73. Communications and Networking Books > SNMP (Simple Network ...
86. LifeSize
38. Veridas Wholesale Telecommunications, Turnkey ISP, ADSL Broadband ...
74. Network Configuration Management – Voyence
36. Office of Telecommunications Management LaNet
24. Computerworld
35. Software – CNET Reviews
65. Welcome to Sprint
47. Welcome to Lantronix! The Home of Networking Devices and Data ...
85. Title Index
21. IM/NOMS 2004
9. Amazon.com: SNMP Network Management (McGraw-Hill Computer ...
30. Network Management Research Group
54. Software – $400 – $500 – CNET Reviews
61. Harris Corporation – Broadcast Communications
69. Software – $500 – $700 – ZDNet Reviews

(c) Concept Highlighter

Search terms

Concepts related to search terms

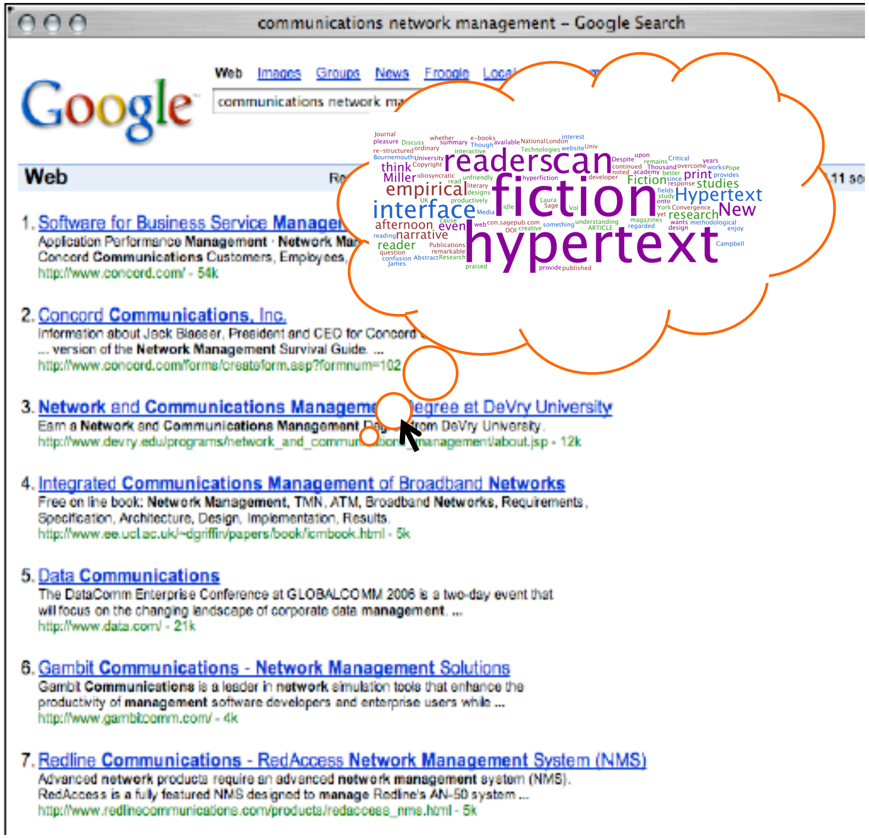
Items in window

HotMap and Concept Highlighter tested somewhat better. See paper for details.

Hoeber & Yang, Comparative Study of Web Search Interfaces, 2006 Conference on Web Intelligence (ACM Digital Library)

Adding InfoViz to Google

- Another idea (due to JDF)
- Pop-up (tool-tip) word clouds over each retrieved document ☺


communications network management – Google Search
Web Images Groups News Froogle Local
communications network ma
Web
Re
11 se
Journal pleasure Discuss whether summary e-books Though available National London interest Technologies website Univ re-structured ordinary interactive Bournemouth University readerscan Despite upon remains Critical years continued Thousand overcome works Pope think Copyright noted academy better print provides Miller idiosyncratic unfriendly hyperfiction developer Fiction since response studies read empirical literary designs fiction fields study Hypertext onto UK productively Laura Sage idle Vol interface Media York Convergence New yet research afternoon cause even web con.sagepub.com something understanding magazines wants methodological DOI creative ARTICLE regarded design enjoy reading narrative reader Publications hypertext Campbell question remarkable confusion Abstract Research James praised provide published

1. Software for Business Service Manage
Application Performance Management · Network Ma
Concord Communications Customers, Employees,
http://www.concord.com/ - 54k

2. Concord Communications, Inc.
Information about Jack Blaeser, President and CEO for Concord
... version of the Network Management Survival Guide. ...
http://www.concord.com/forms/createform.asp?formnum=102

3. Network and Communications Manageme egree at DeVry University
Earn a Network and Communications Management Deg from DeVry University.
http://www.devry.edu/programs/network_and_commun ation managementabout.jsp - 12k

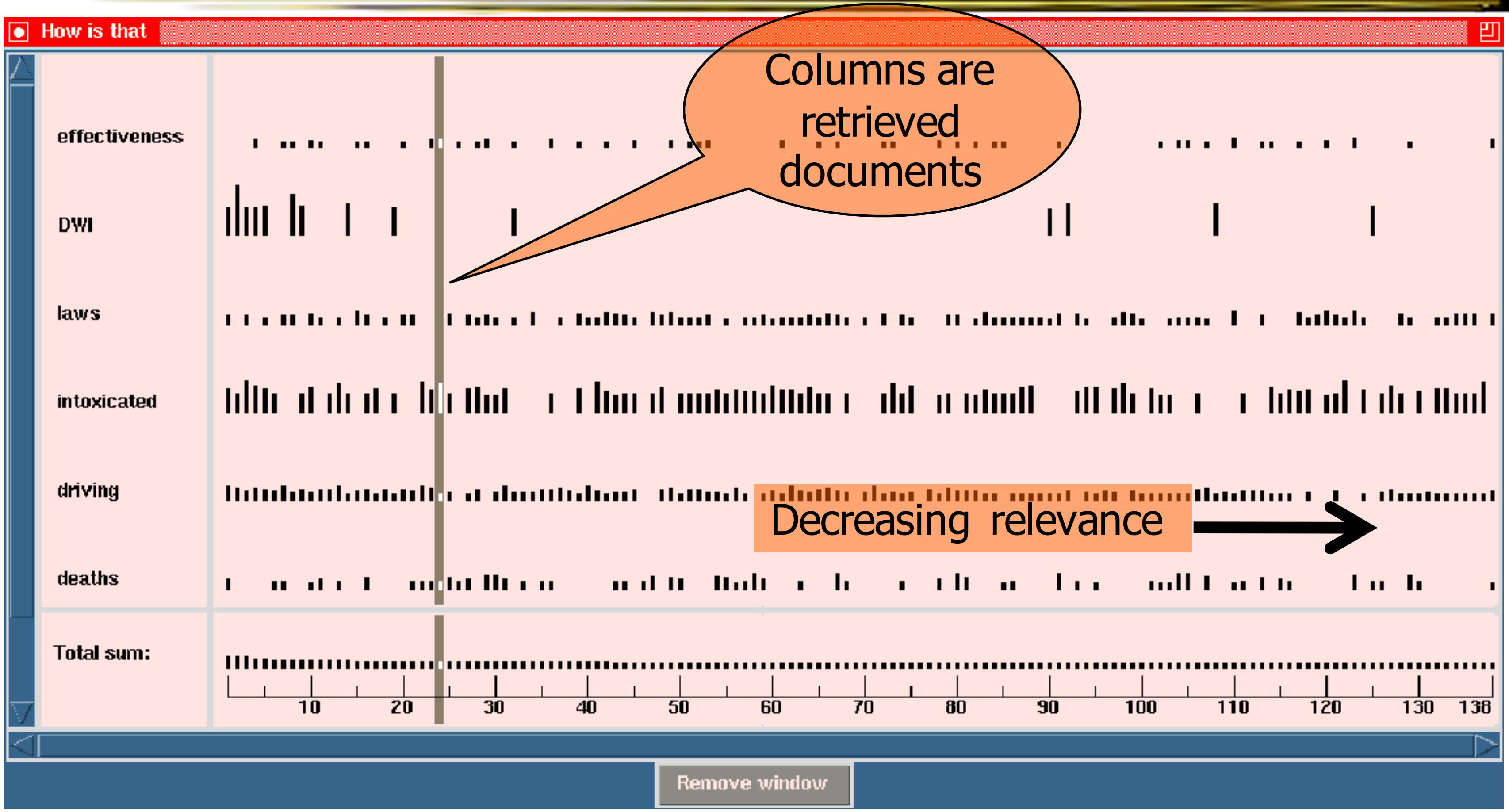
4. Integrated Communications Management of Broadband Networks
Free on line book: Network Management, TMN, ATM, Broadband Networks, Requirements, Specification, Architecture, Design, Implementation, Results.
http://www.ee.ucl.ac.uk/~dgriffin/papers/book/icmbook.html - 5k

5. Data Communications
The DataComm Enterprise Conference at GLOBALCOMM 2006 is a two-day event that will focus on the changing landscape of corporate data management. ...
http://www.data.com/ - 21k

6. Gambit Communications - Network Management Solutions
Gambit Communications is a leader in network simulation tools that enhance the productivity of management software developers and enterprise users while ...
http://www.gambitcomm.com/ - 4k

7. Redline Communications - RedAccess Network Management System (NMS)
Advanced network products require an advanced network management system (NMS). RedAccess is a fully featured NMS designed to manage Redline's AN-50 system ...
http://www.redlinecommunications.com/products/redaccess_nms.html - 5k

Understanding Relevance - Veerasamy

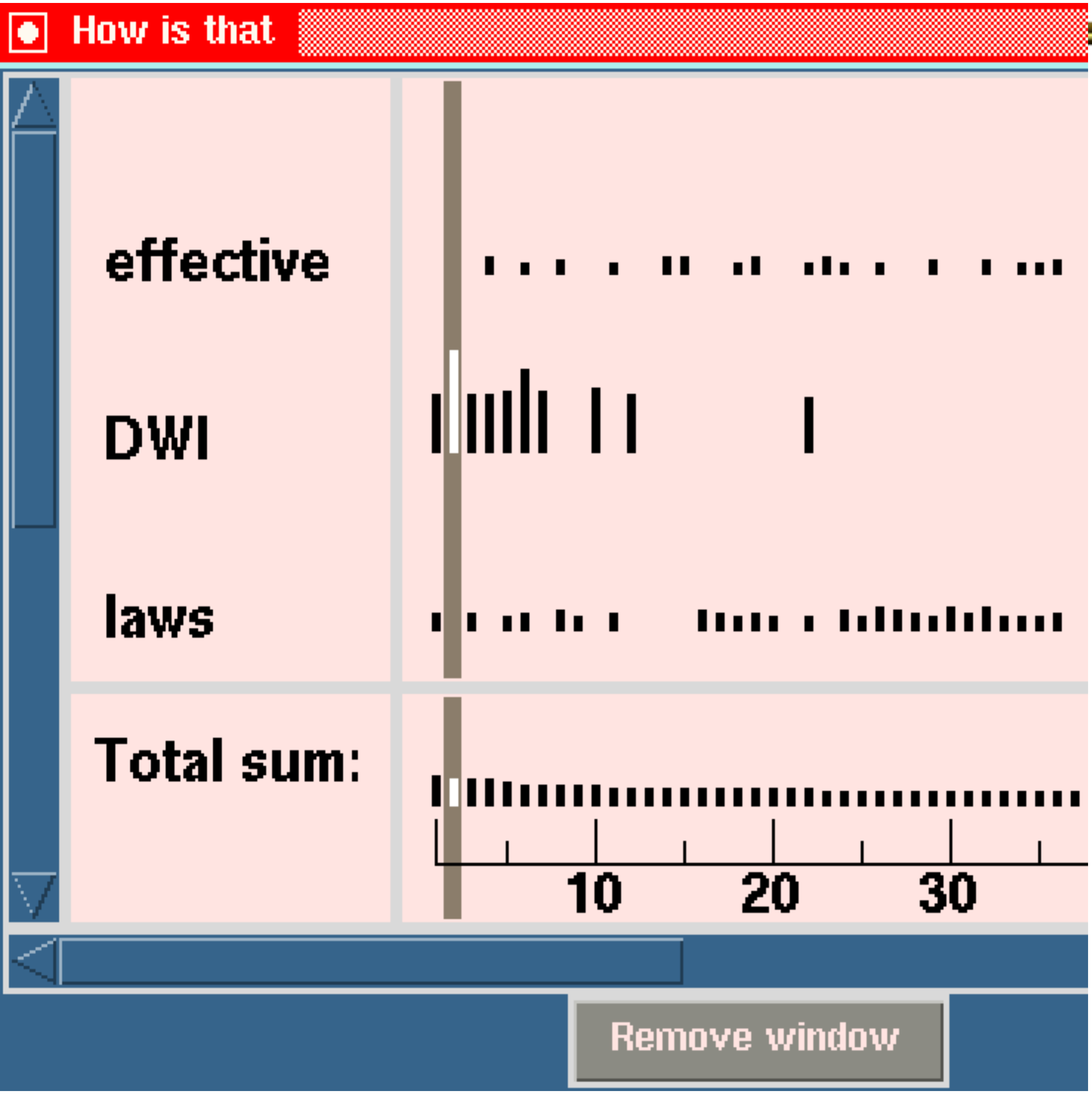


Columns are retrieved documents

Decreasing relevance

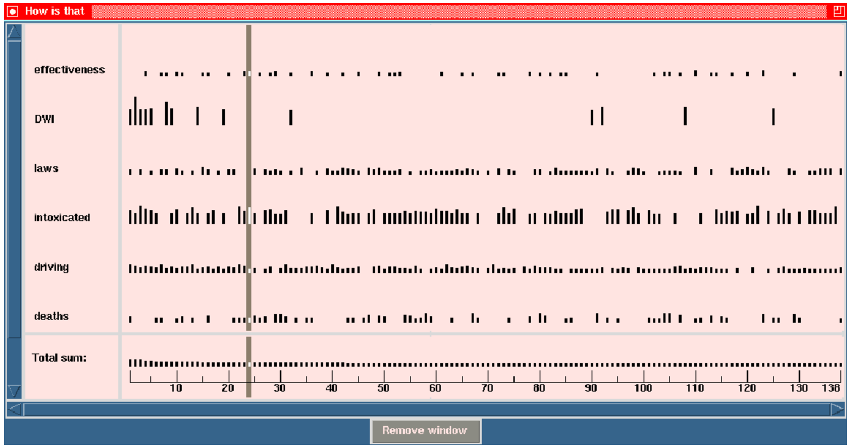
Veerasamy & Belkin, *Evaluation of a tool for visualization of information retrieval results, ACM Conference on Research and Development in Information Retrieval*, Pro. 19th annual international ACM SIGIR conference (ACM Digital Library)

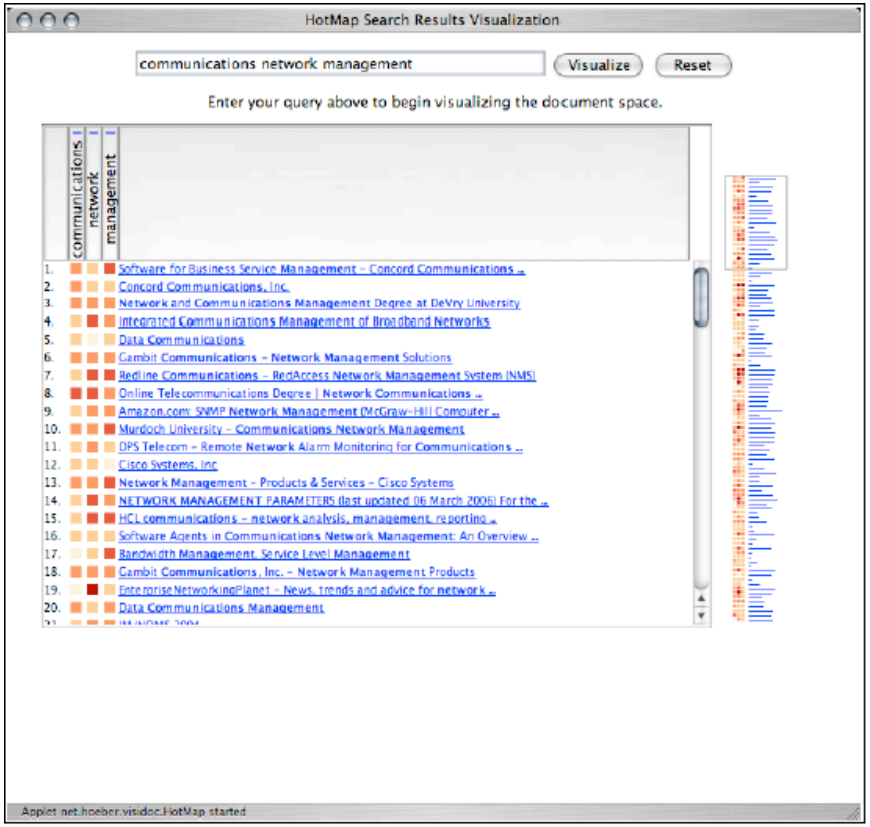
Understanding Relevance



- Keyword bars show relevance for document
- "Total sum" gives overall relevance
- Experimental comparison against no InfoVis
  - No difference :-(

Pros and Cons??




HotMap Search Results Visualization
communications network management
Visualize
Reset
communications
network
management

Enter your query above to begin visualizing the document space.

1. Software for Business Service Management - Concord Communications ...
2. Concord Communications, Inc.
3. Network and Communications Management Degree at DeVry University
4. Integrated Communications Management of Broadband Networks
5. Data Communications
6. Gambit Communications - Network Management Solutions
7. Redline Communications - RedAccess Network Management System (NMS)
8. Online Telecommunications Degree | Network Communications ...
9. Amazon.com: SNMP Network Management (McGraw-Hill Computer ...
10. Murdoch University - Communications Network Management
11. DPS Telecom - Remote Network Alarm Monitoring for Communications ...
12. Cisco Systems, Inc
13. Network Management - Products & Services - Cisco Systems
14. NETWORK MANAGEMENT PARAMETERS (last updated 06 March 2006) For the ...
15. HCL communications - network analysis, management, reporting ...
16. Software Agents in Communications Network Management: An Overview ...
17. Bandwidth Management, Service Level Management
18. Gambit Communications, Inc. - Network Management Products
19. EnterpriseNetworkingPlanet - News, trends and advice for network ...
20. Data Communications Management

Applet net.hoeber.visidoc.HotMap started

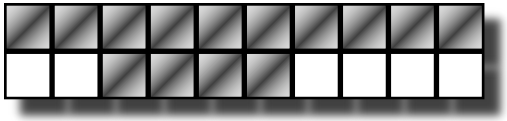
(b) HotMap

Understanding Relevance - TileBars

- 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

Hearst
CHI ‘95

TileBars

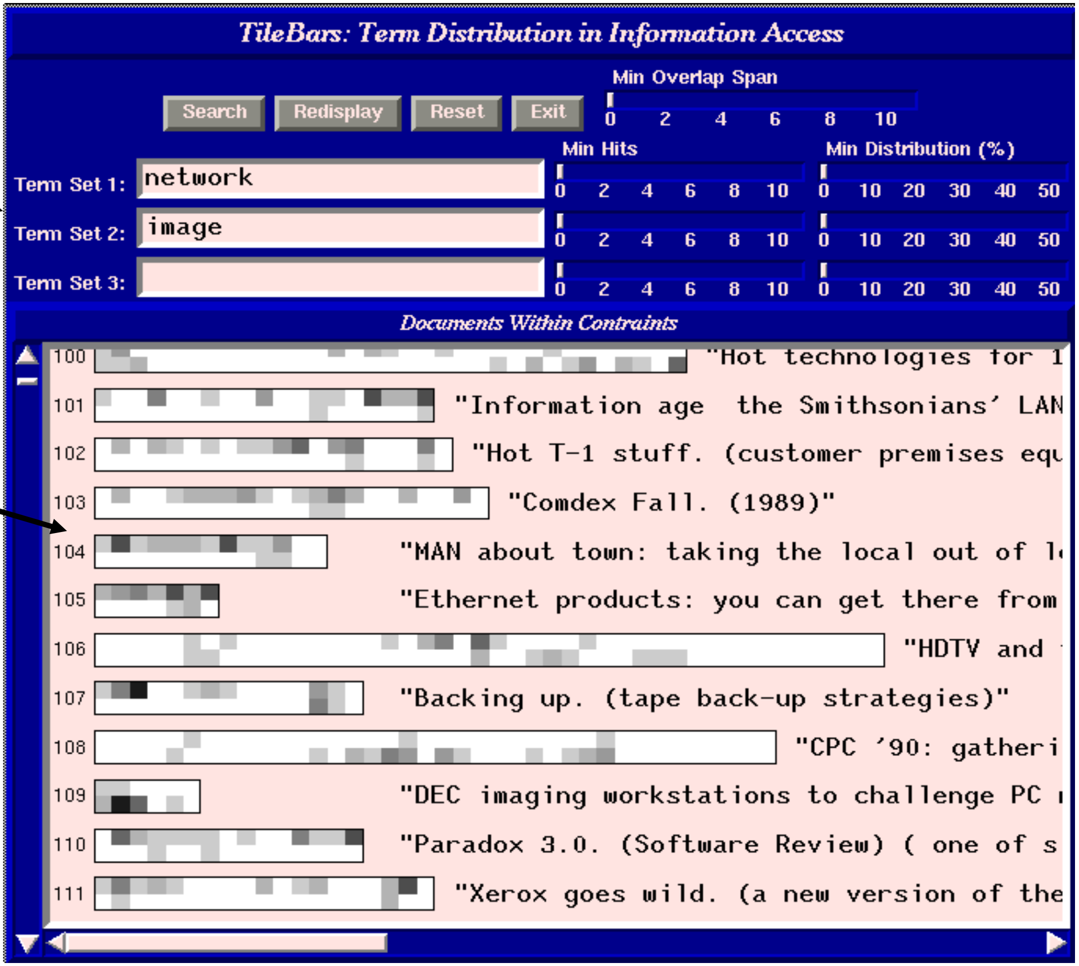


- Graphical representation of term distribution and overlap
- Simultaneously indicate:
  - Relative document length
  - Frequency of term sets in document
  - Distribution of term sets with respect to the document and each other

TileBars Interface

Search terms

Presentation



Term Set 1: network

Term Set 2: image

Term Set 3:

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

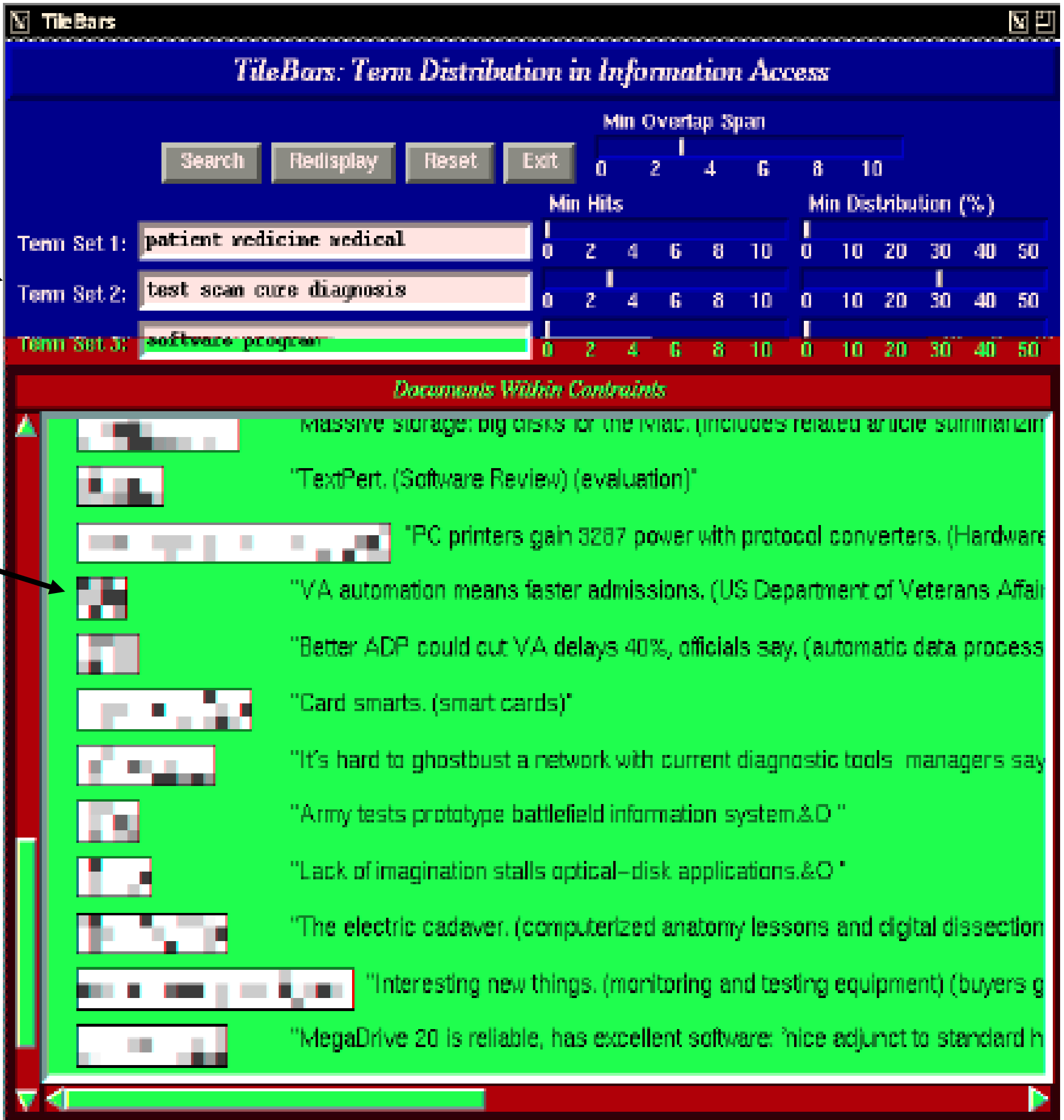
110 "Paradox 3.0. (Software Review) (one of s

111 "Xerox goes wild. (a new version of the

Additional Terms => More Rows

Three search terms

Three rows per tile bar



TileBars: Term Distribution in Information Access

Term Set 1: patient medicine medical

Term Set 2: test scan cure diagnosis

Term Set 3: software program

Documents Within Contraints

Massive storage: big disks for the Mac. (includes related article summarizin

"TextPert. (Software Review) (evaluation)"

"PC printers gain 3287 power with protocol converters. (Hardware

"VA automation means faster admissions. (US Department of Veterans Affai

"Better ADP could cut VA delays 40%, officials say. (automatic data process

"Card smarts. (smart cards)"

"It's hard to ghostbust a network with current diagnostic tools managers say

"Army tests prototype battlefield information system.&O "

"Lack of imagination stalls optical-disk applications.&O "

"The electric cadaver. (computerized anatomy lessons and digital dissection

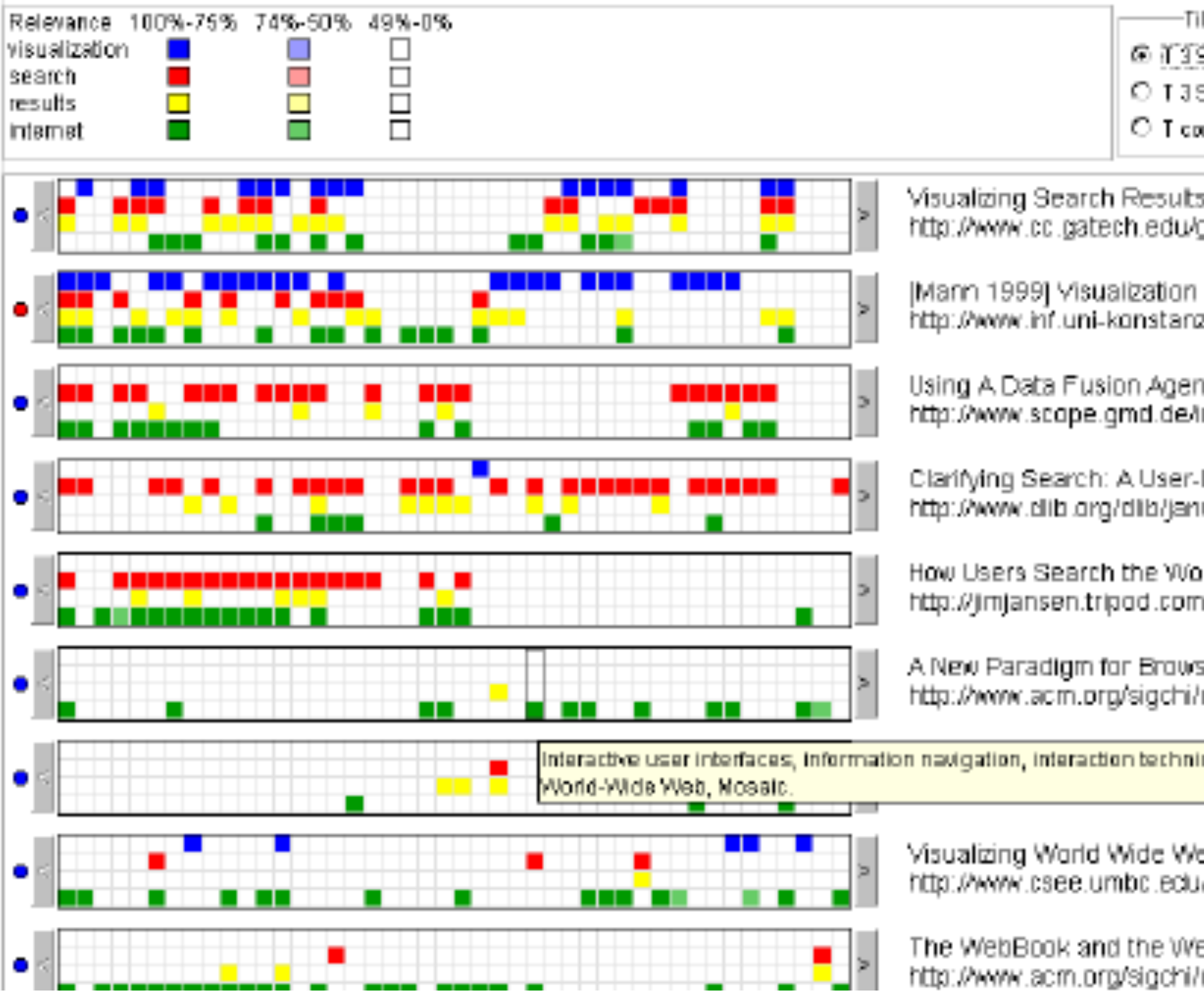
"Interesting new things. (monitoring and testing equipment) (buyers g

"MegaDrive 20 is reliable, has excellent software: 'nice adjunct to standard h

TileBars Interface

Color coding shows degree of relevance

Notice how doc length is treated



Relevance	100%-75%	74%-50%	49%-0%
visualization	🟦	▨	□
search	🟥	▨	□
results	🟨	▨	□
internet	🟩	▨	□

Ti

◉ T 39

○ T 39

○ T co

Visualizing Search Results
http://www.cc.gatech.edu/g

[Mann 1999] Visualization
http://www.inf.uni-konstanz

Using A Data Fusion Agen
http://www.scope.gmd.de/i

Clarifying Search: A User-I
http://www.dlib.org/dlib/jan

How Users Search the Wo
http://jimjansen.tripod.com

A New Paradigm for Brows
http://www.acm.org/sigchi/

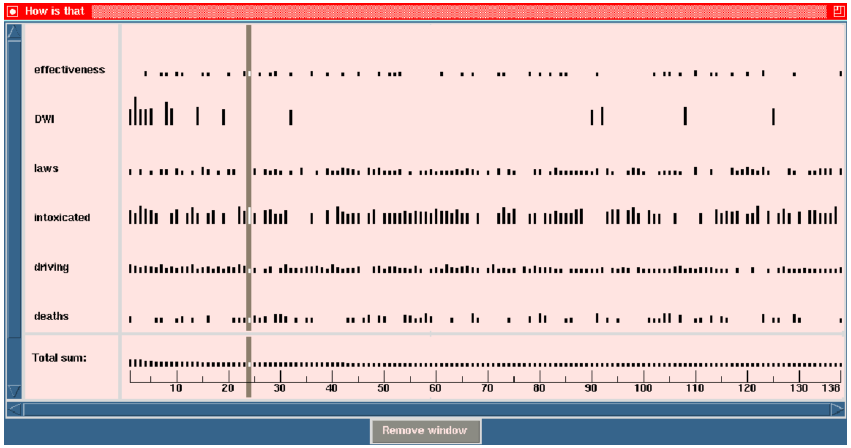
Interactive user interfaces, information navigation, interaction techni
World-Wide Web, Mosaic.

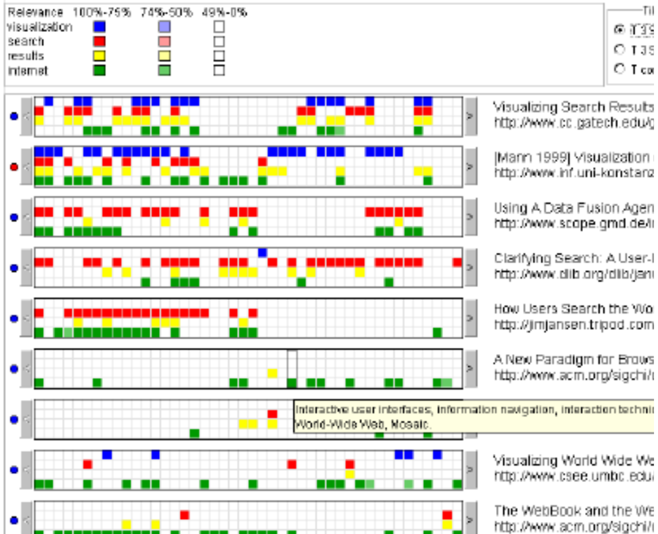
Visualizing World Wide We
http://www.csee.umbc.edu/

The WebBook and the We
http://www.acm.org/sigchi/

TileBar Issues

- Other Issues?
  - How about scaling up to more documents and search terms
- But what does TileBars give us that Veerasamy method does not? Pros/cons?





Relevance	100%-75%	74%-50%	49%-0%
visualization	🟦	🟦	□
search	🟥	🟥	□
results	🟨	🟨	□
internet	🟩	🟩	□

Visualizing Search Results
http://www.cc.gatech.edu/g

[Mann 1999] Visualization
http://www.inf.uni-konstanz

Using A Data Fusion Agen
http://www.scope.gmd.de/i

Clarifying Search: A User-I
http://www.dlib.org/dlib/jan

How Users Search the Wo
http://jimjansen.tripod.com

A New Paradigm for Brows
http://www.acm.org/sigchi/

Interactive user interfaces, information navigation, interaction techni
World-Wide Web, Mosaic.

Visualizing World Wide We
http://www.csee.umbc.edu/

The WebBook and the We
http://www.acm.org/sigchi/

End of Understanding Relevance

- VIBE
- SQUID
- JIGSAW
- Google (and extensions)
- Veerasamy’s work
- TileBars (dozen or less)

- Pros/cons?
- How does each scale with
  - Number of documents?
  - Number of keywords?

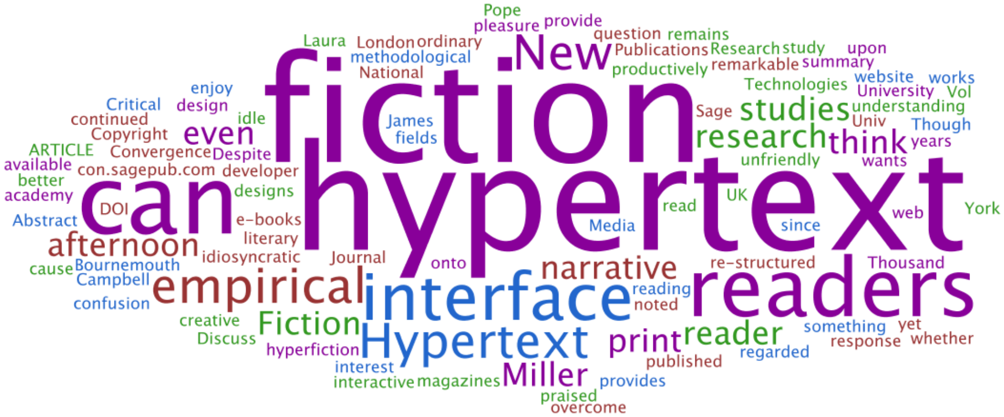
Making Sense of Smaller Collections

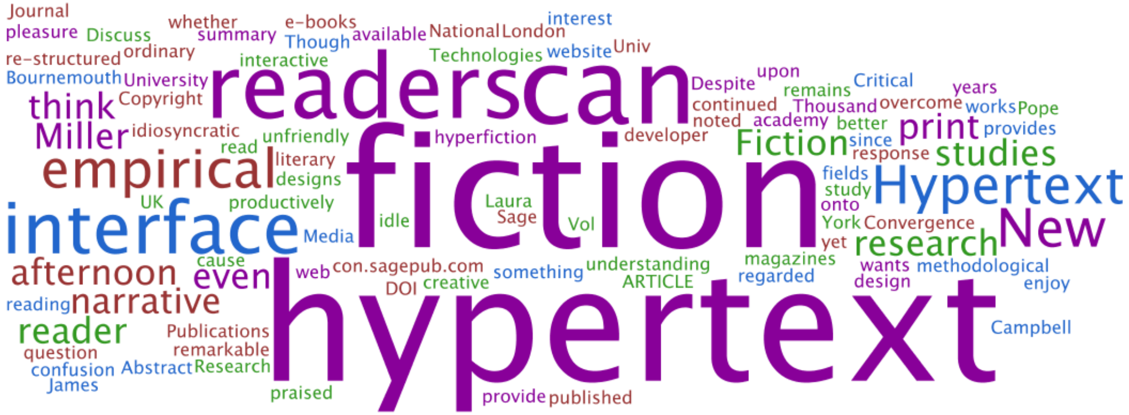
- Getting an overall sense
  - Word/Tag clouds
- Comparing multiple docs
  - NYT State of Union example
  - Washington Post State of Union example
  - New York Times Political Convention speeches
  - Plagiarism

Tag Clouds and Word Clouds

- **Tag** Clouds represent explicit meta-data about a document or web site or picture.
  - Typically user-assigned – “crowd-sourced keywords”
- **Word** Clouds, in contrast, are generated from the words of a document or document collection or web site.
  - In some sense they are automatically generated meta-data.
  - We could call them implicit meta-data
- Ways to use are same, but we will always refer to Word Clouds, not Tag Clouds

Alpha order / prominent in center / etc

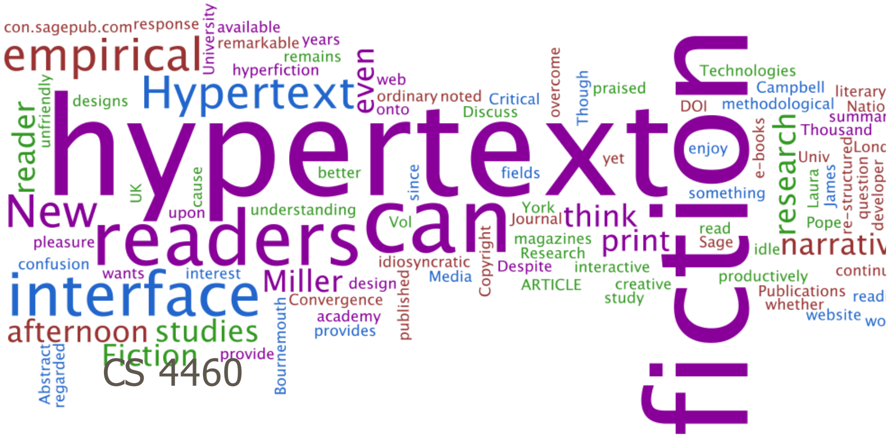




Same / Different Orientation

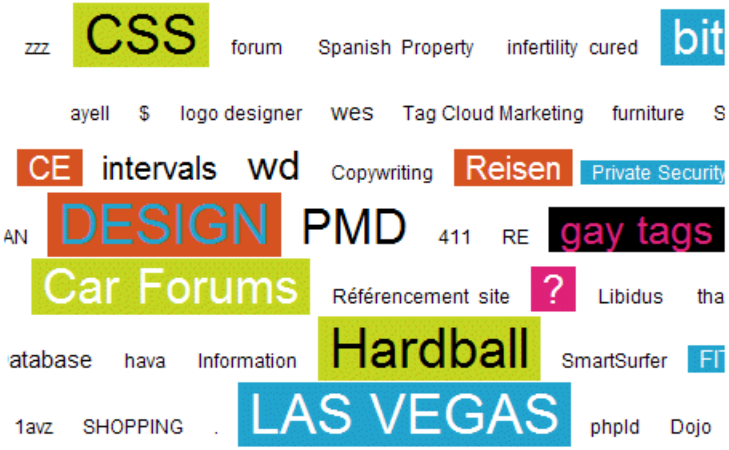
hypertext
fiction
readers
can
interface
empirical
Hypertext
New
research
Fiction
print
studies
narrative
reader
afternoon
even
think
Miller

hypertext
fiction
readers
can
interface
empirical
Hypertext
New
print
studies
research
Fiction
Miller
afternoon
narrative
reader
think
even



text
hypertext
reading
texts
author
Nelson
Vannevar
University
different
reader
will
Bush
writing
new
order
life
Downloaded
Singapore
neo-structuralists
work
knowledge
one
September
can
National
dio.sagepub.com

Lots of Other Design Options



zzz CSS forum Spanish Property infertility cured bit

ayell $ logo designer wes Tag Cloud Marketing furniture S

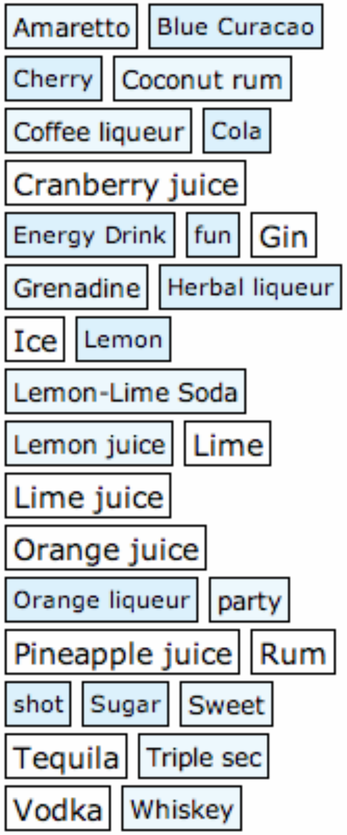
CE intervals wd Copywriting Reisen Private Security

AN DESIGN PMD 411 RE gay tags

Car Forums Référencement site ? Libidus tha

atabase hava Information Hardball SmartSurfer FI

1avz SHOPPING . LAS VEGAS phpld Dojo



Amaretto Blue Curacao

Cherry Coconut rum

Coffee liqueur Cola

Cranberry juice

Energy Drink fun Gin

Grenadine Herbal liqueur

Ice Lemon

Lemon-Lime Soda

Lemon juice Lime

Lime juice

Orange juice

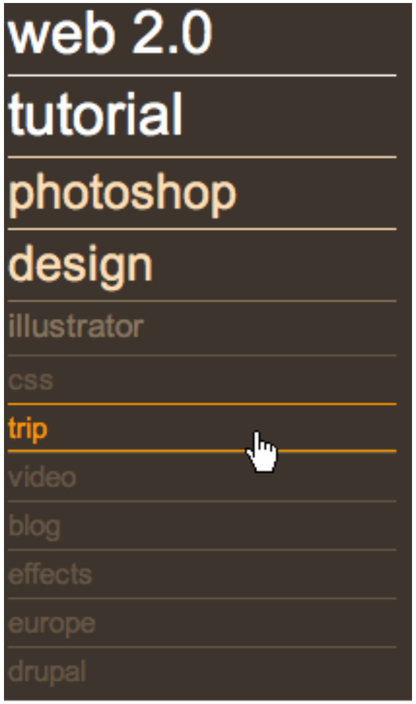
Orange liqueur party

Pineapple juice Rum

shot Sugar Sweet

Tequila Triple sec

Vodka Whiskey



web 2.0

tutorial

photoshop

design

illustrator

css

trip

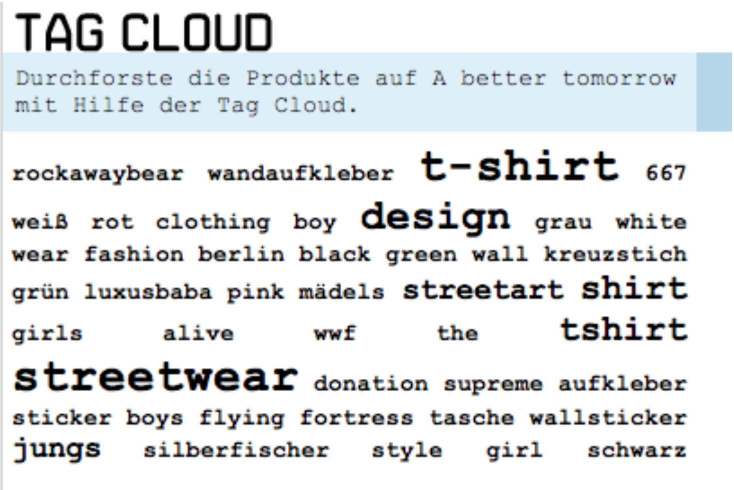
video

blog

effects

europe

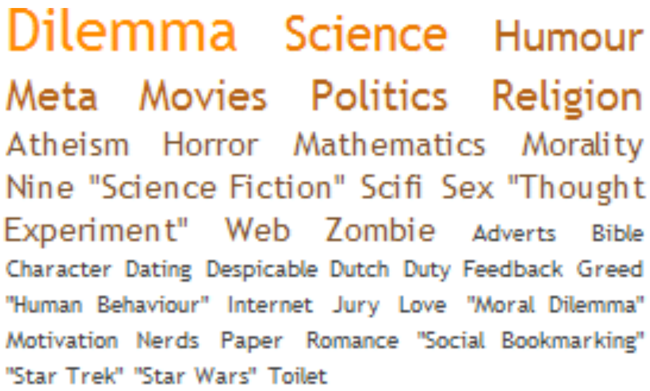
drupal



TAG CLOUD

Durchforste die Produkte auf A better tomorrow mit Hilfe der Tag Cloud.

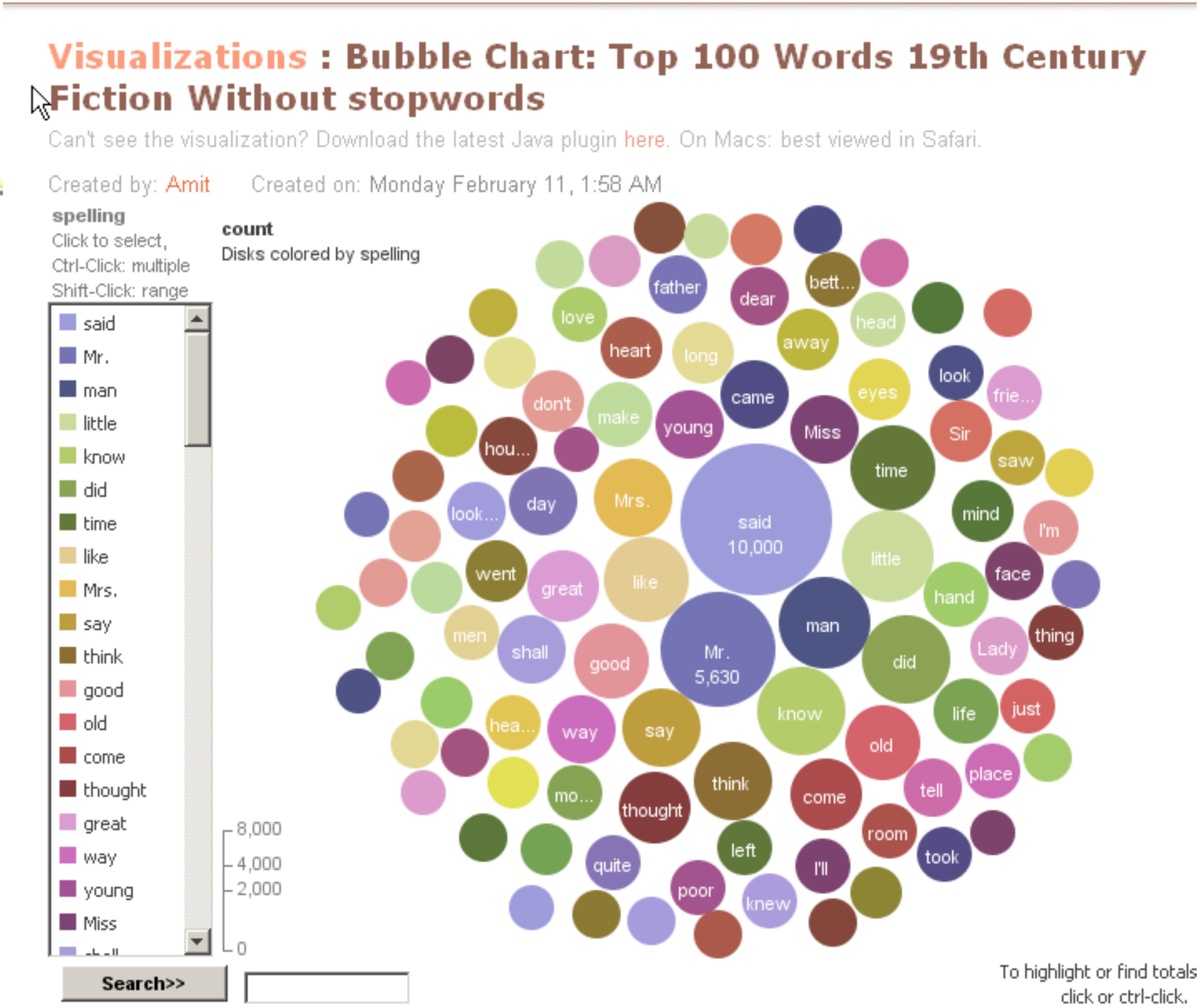
rockawaybear wandaufkleber t-shirt 667 weiß rot clothing boy design grau white wear fashion berlin black green wall kreuzstich grün luxusbaba pink mädels streetart shirt girls alive wwf the tshirt streetwear donation supreme aufkleber sticker boys flying fortress tasche wallsticker jungs silberfischer style girl schwarz



Dilemma Science Humour Meta Movies Politics Religion Atheism Horror Mathematics Morality Nine "Science Fiction" Scifi Sex "Thought Experiment" Web Zombie Adverts Bible Character Dating Despicable Dutch Duty Feedback Greed "Human Behaviour" Internet Jury Love "Moral Dilemma" Motivation Nerds Paper Romance "Social Bookmarking" "Star Trek" "Star Wars" Toilet

Bubble Chart

Pros/
Cons?



Visualizations : Bubble Chart: Top 100 Words 19th Century Fiction Without stopwords

Can't see the visualization? Download the latest Java plugin here. On Macs: best viewed in Safari.

Created by: Amit   Created on: Monday February 11, 1:58 AM

spelling
Click to select,
Ctrl-Click: multiple
Shift-Click: range

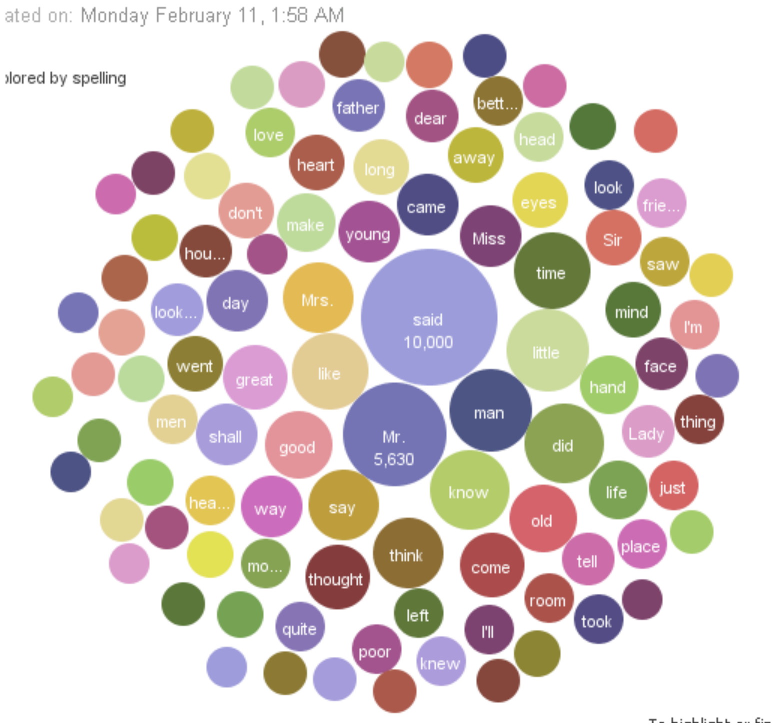
count
Disks colored by spelling

To highlight or find totals
click or ctrl-click.

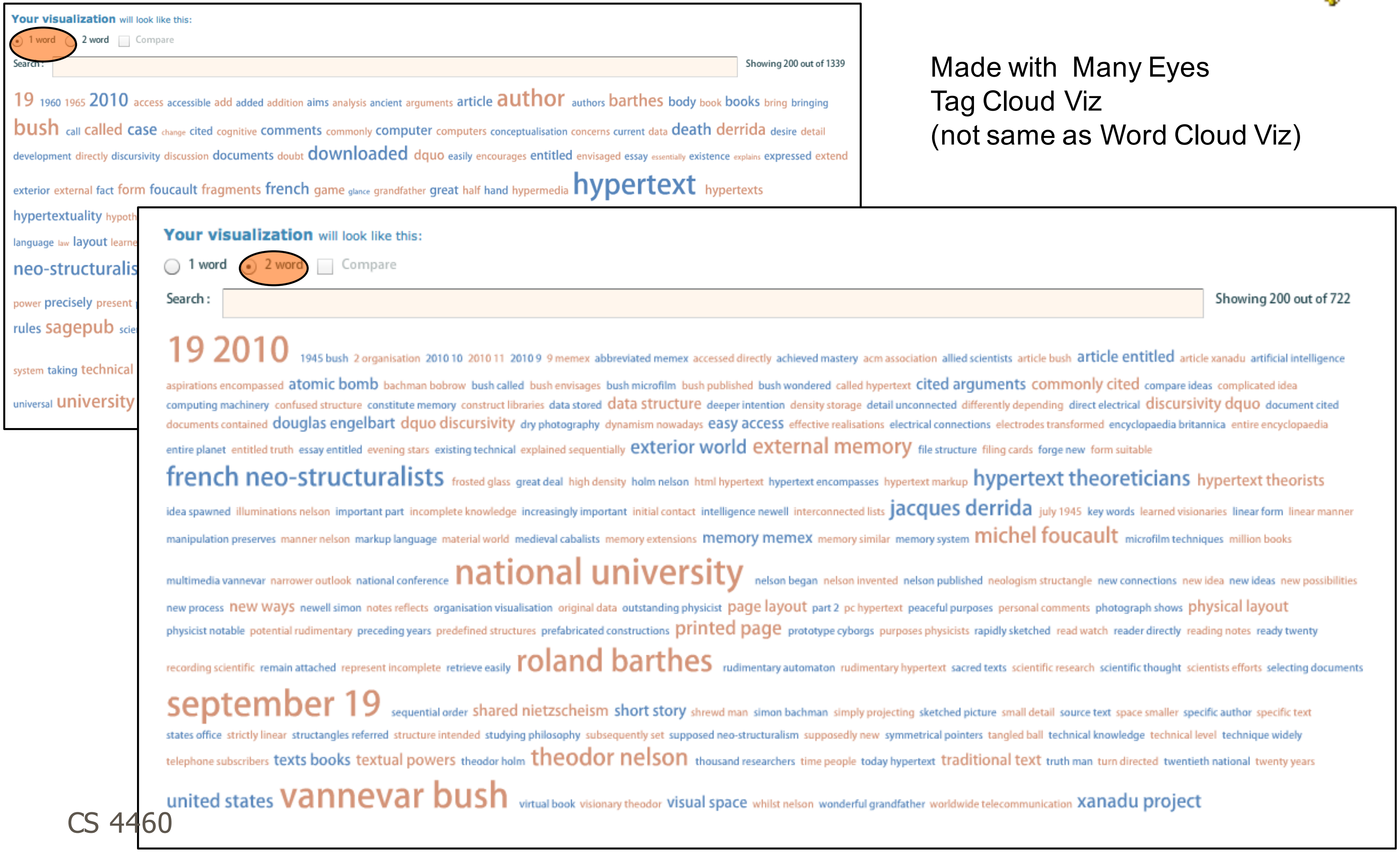
Bubble Charts (implemented by Wattenberg)

Other Uses for Word Bubble Chart?

- How could we use this to find differences between
  - Authors
  - Centuries
  - Versions
- Group time ☺



Word Pairs more Useful



Your visualization will look like this:

● 1 word ○ 2 word □ Compare

Search :

Showing 200 out of 1339

19 1960 1965 2010 access accessible add added addition aims analysis ancient arguments article author authors barthes body book books bring bringing
bush call called case change cited cognitive comments commonly computer computers conceptualisation concerns current data death derrida desire detail
development directly discursivity discussion documents doubt downloaded dquo easily encourages entitled envisaged essay essentially existence explains expressed extend
exterior external fact form foucault fragments french game glance grandfather great half hand hypermedia hypertext hypertexts
hypertextuality hypoth
language law layout learne
neo-structuralis
power precisely present
rules sagepub scie
system taking technical
universal university

Your visualization will look like this:

○ 1 word ● 2 word □ Compare

Search :

Showing 200 out of 722

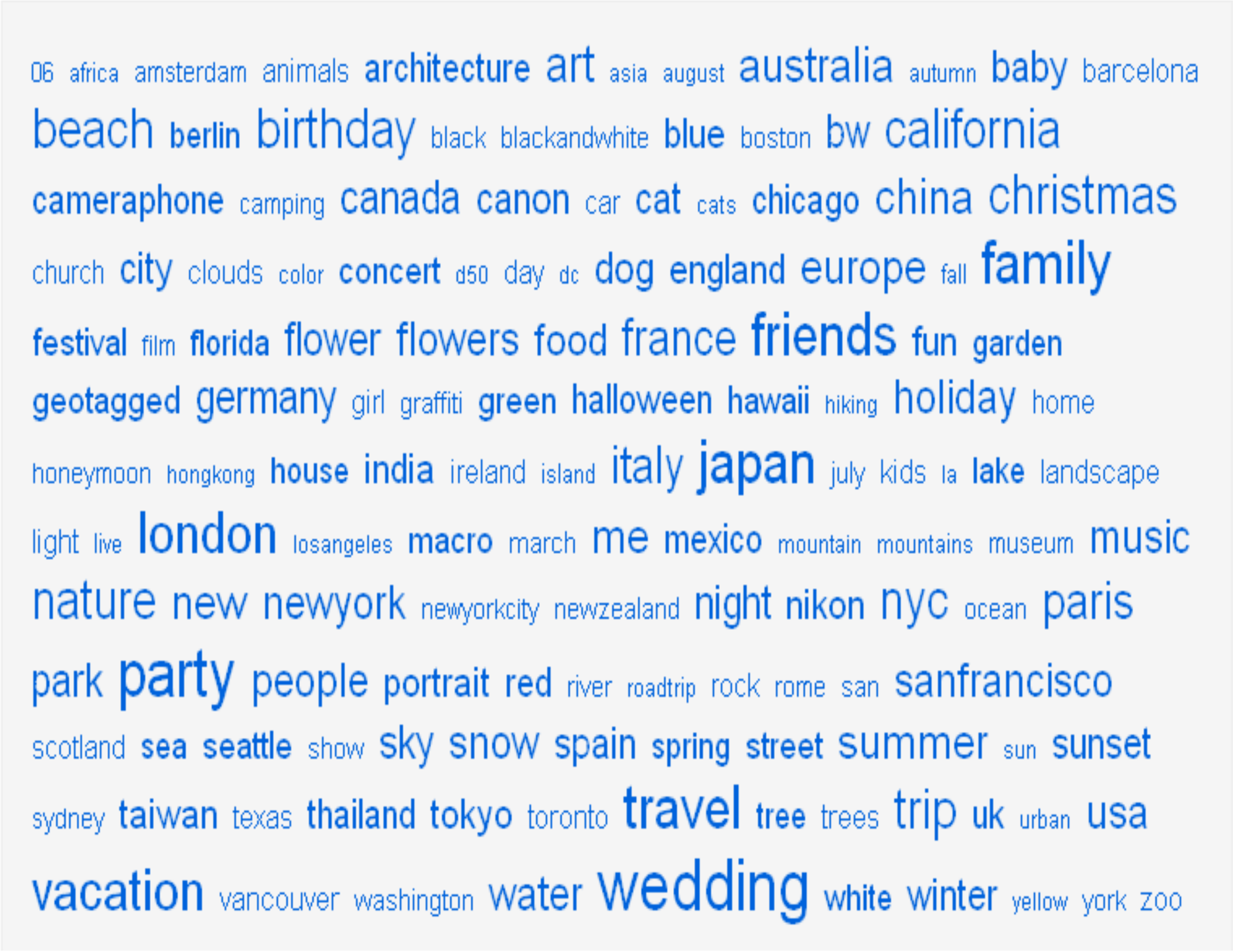
19 2010 1945 bush 2 organisation 2010 10 2010 11 2010 9 9 memex abbreviated memex accessed directly achieved mastery acm association allied scientists article bush article entitled article xanadu artificial intelligence
aspirations encompassed atomic bomb bachman bobrow bush called bush envisages bush microfilm bush published bush wondered called hypertext cited arguments commonly cited compare ideas complicated idea
computing machinery confused structure constitute memory construct libraries data stored data structure deeper intention density storage detail unconnected differently depending direct electrical discursivity dquo document cited
documents contained douglas engelbart dquo discursivity dry photography dynamism nowadays easy access effective realisations electrical connections electrodes transformed encyclopaedia britannica entire encyclopaedia
entire planet entitled truth essay entitled evening stars existing technical explained sequentially exterior world external memory file structure filing cards forge new form suitable
french neo-structuralists frosted glass great deal high density holm nelson html hypertext hypertext encompasses hypertext markup hypertext theoreticians hypertext theorists
idea spawned illuminations nelson important part incomplete knowledge increasingly important initial contact intelligence newell interconnected lists jacques derrida july 1945 key words learned visionaries linear form linear manner
manipulation preserves manner nelson markup language material world medieval cabalists memory extensions memory memex memory similar memory system michel foucault microfilm techniques million books
multimedia vannevar narrower outlook national conference national university nelson began nelson invented nelson published neologism structangle new connections new idea new ideas new possibilities
new process new ways newell simon notes reflects organisation visualisation original data outstanding physicist page layout part 2 pc hypertext peaceful purposes personal comments photograph shows physical layout
physicist notable potential rudimentary preceding years predefined structures prefabricated constructions printed page prototype cyborgs purposes physicists rapidly sketched read watch reader directly reading notes ready twenty
recording scientific remain attached represent incomplete retrieve easily roland barthes rudimentary automaton rudimentary hypertext sacred texts scientific research scientific thought scientists efforts selecting documents
september 19 sequential order shared nietzscheism short story shrewd man simon bachman simply projecting sketched picture small detail source text space smaller specific author specific text
states office strictly linear structangles referred structure intended studying philosophy subsequently set supposed neo-structuralism supposedly new symmetrical pointers tangled ball technical knowledge technical level technique widely
telephone subscribers texts books textual powers theodor holm theodor nelson thousand researchers time people today hypertext traditional text truth man turn directed twentieth national twenty years
united states vannevar bush virtual book visionary theodor visual space whilst nelson wonderful grandfather worldwide telecommunication xanadu project

Made with Many Eyes
Tag Cloud Viz
(not same as Word Cloud Viz)

Meaningful Associations Confused

Find the country names in this cloud

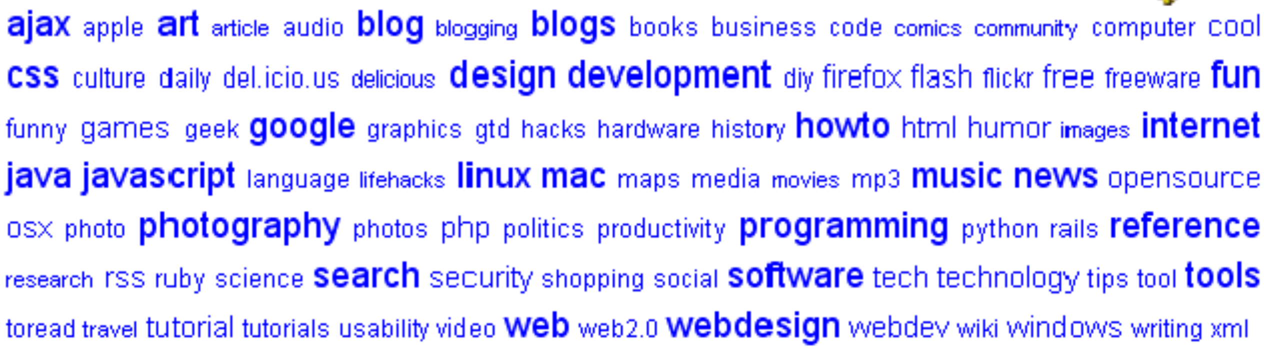
FAST!!



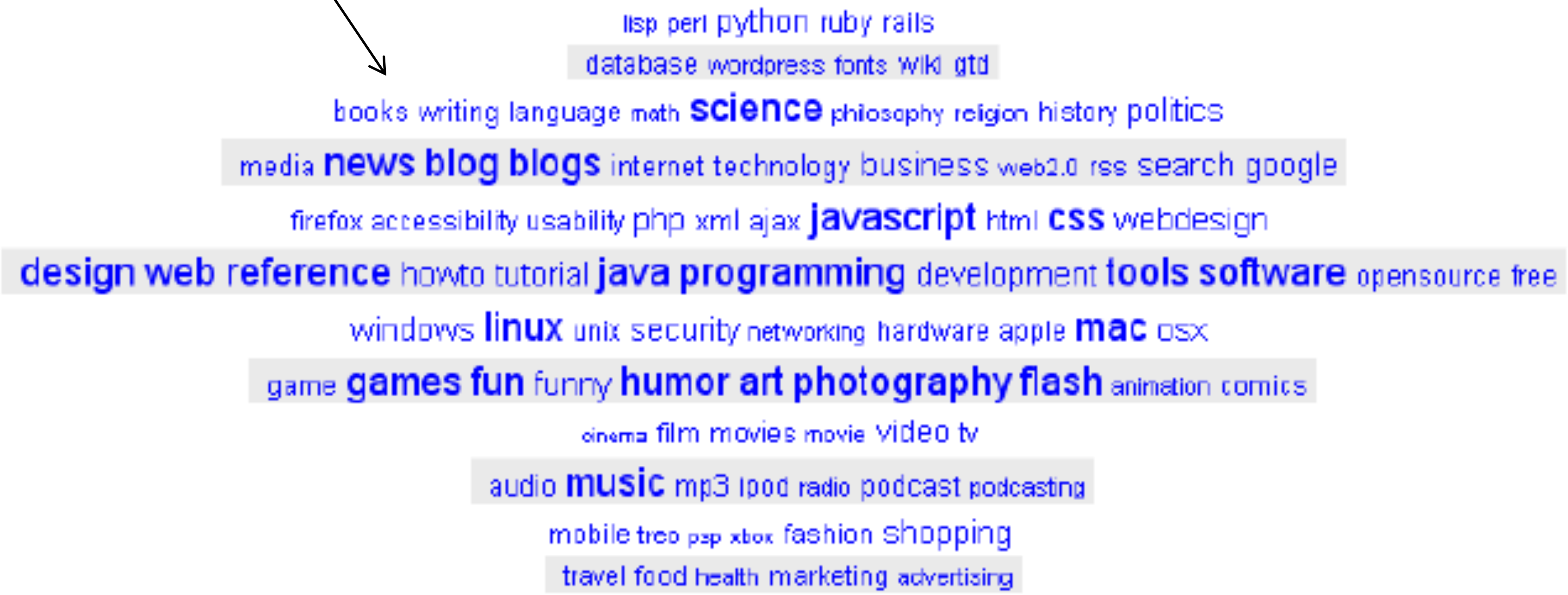
06 africa amsterdam animals architecture art asia august australia autumn baby barcelona beach berlin birthday black blackandwhite blue boston bw california cameraphone camping canada canon car cat cats chicago china christmas church city clouds color concert d50 day dc dog england europe fall family festival film florida flower flowers food france friends fun garden geotagged germany girl graffiti green halloween hawaii hiking holiday home honeymoon hongkong house india ireland island italy japan july kids la lake landscape light live london losangeles macro march me mexico mountain mountains museum music nature new newyork newyorkcity newzealand night nikon nyc ocean paris park party people portrait red river roadtrip rock rome san sanfrancisco scotland sea seattle show sky snow spain spring street summer sun sunset sydney taiwan texas thailand tokyo toronto travel tree trees trip uk urban usa vacation vancouver washington water wedding white winter yellow york zoo

Alternative: "Semantic" Layout

Tags are grouped based on clustering and co-occurrence analysis – words that co-occur close to one another in the text are placed together in the cloud



ajax apple art article audio blog blogging blogs books business code comics community computer cool
css culture daily del.icio.us delicious design development diy firefox flash flickr free freeware fun
funny games geek google graphics gtd hacks hardware history howto html humor images internet
java javascript language lifehacks linux mac maps media movies mp3 music news opensource
osx photo photography photos php politics productivity programming python rails reference
research rss ruby science search security shopping social software tech technology tips tool tools
toread travel tutorial tutorials usability video web web2.0 webdesign webdev wiki windows writing xml

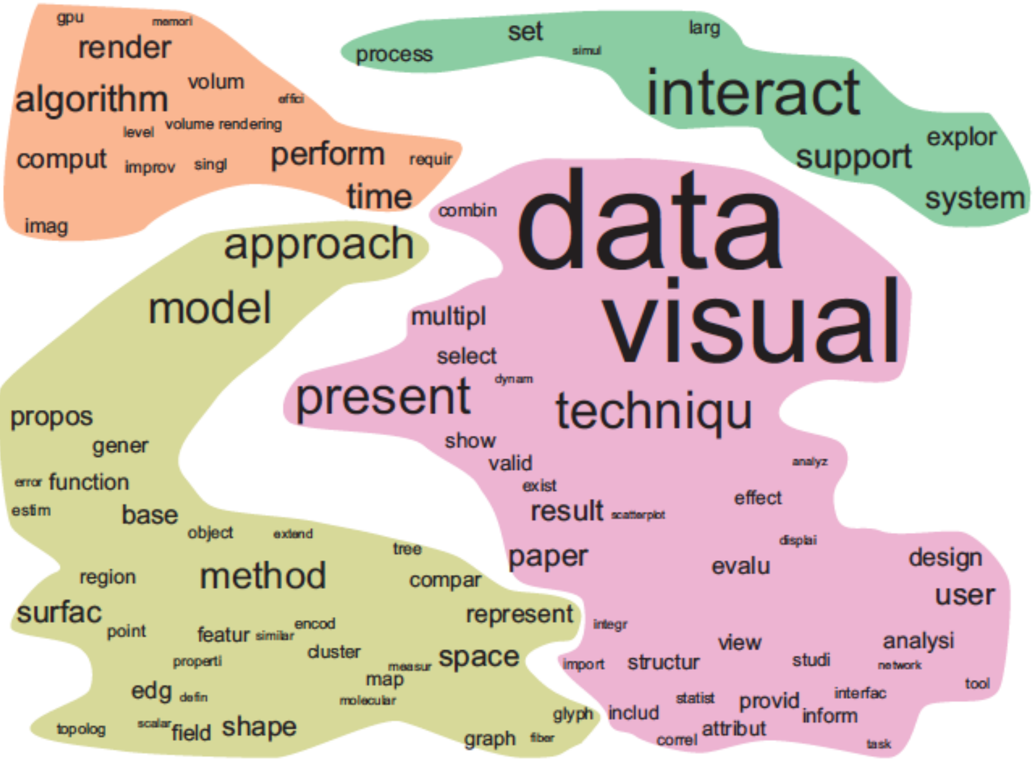


lisp perl python ruby rails
database wordpress fonts wiki gtd
books writing language math science philosophy religion history politics
media news blog blogs internet technology business web2.0 rss search google
firefox accessibility usability php xml ajax javascript html css webdesign
design web reference howto tutorial java programming development tools software opensource free
windows linux unix security networking hardware apple mac osx
game games fun funny humor art photography flash animation comics
cinema film movies movie video tv
audio music mp3 ipod radio podcast podcasting
mobile treo psp xbox fashion shopping
travel food health marketing advertising

Hassan-Monteroa & Herrero-Solana, *Improving Tag-Clouds as Visual Information Retrieval Interfaces*, InSciT2006

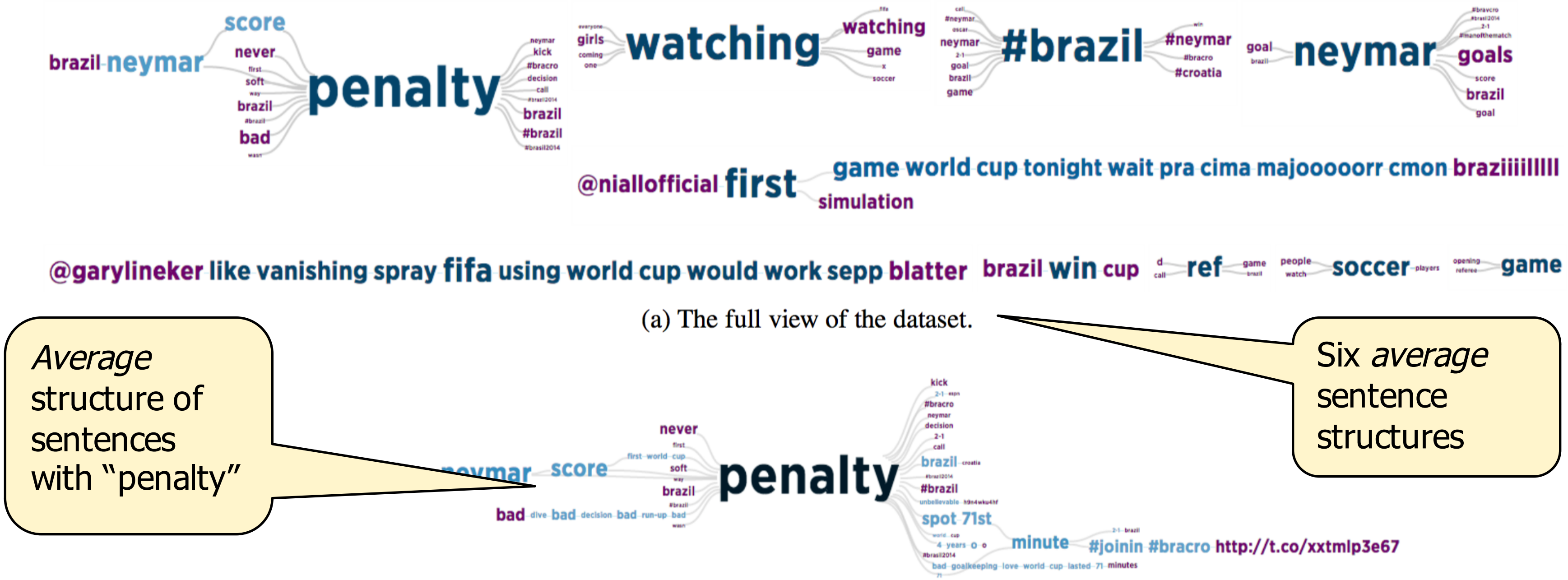
Semantics-preserving Word Clouds

1. Determine co-occurrence word count within sentence.
2. Separate words into clusters based on high co-occurrence count.



Semantic-Preserving Word Clouds by Seam Carving, Wu et. al, EuroVis 20111

SentenTree



(a) The full view of the dataset.

(b) A zoomed-in view focused on *penalty* after the viewer has clicked on that word.

Fig. 8: A SentenTree visualization of tweets commenting on the third goal of the opening game of the World Cup. This dataset contains 132,599 tweets (75,930 unique tweets).

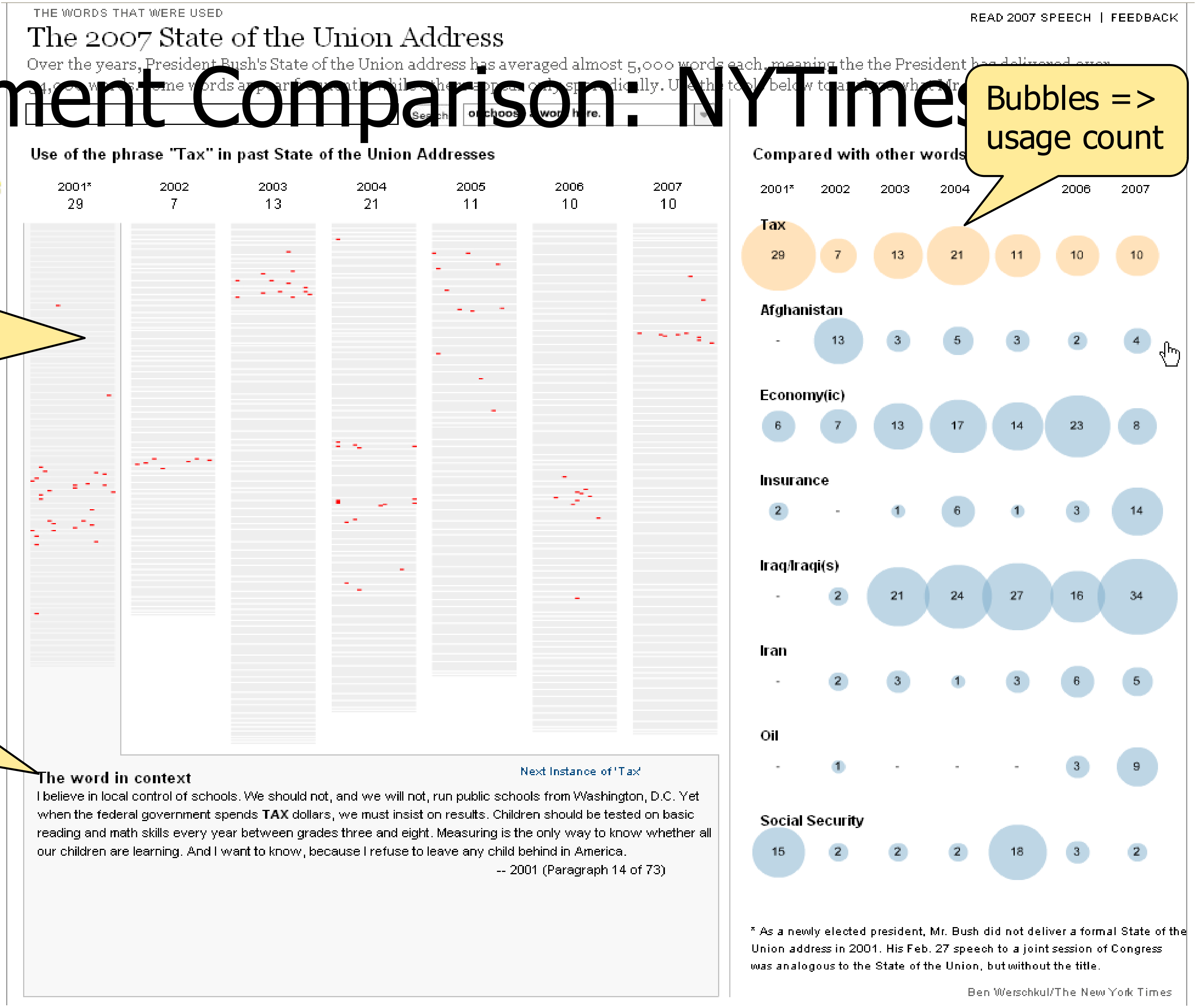
Average structure of sentences with "penalty"

Six *average* sentence structures

Think of this as a really smart word cloud!

Hu, Wongsuphasawat and Stasko, Visualizing Social Media Content with SentenTree, IEEE Transactions on Visualization and Computer Graphics, January 2017.

Document Comparison: NYTimes



THE WORDS THAT WERE USED

READ 2007 SPEECH | FEEDBACK

The 2007 State of the Union Address

Over the years, President Bush's State of the Union address has averaged almost 5,000 words each, meaning the the President has delivered over 34,000 words. Some words appear frequently while others appear only sporadically. Use the tool below to analyze what Mr.

Search or choose a word here.

Use of the phrase "Tax" in past State of the Union Addresses

2001*	2002	2003	2004	2005	2006	2007
29	7	13	21	11	10	10

The word in context

Next Instance of 'Tax'

I believe in local control of schools. We should not, and we will not, run public schools from Washington, D.C. Yet when the federal government spends **TAX** dollars, we must insist on results. Children should be tested on basic reading and math skills every year between grades three and eight. Measuring is the only way to know whether all our children are learning. And I want to know, because I refuse to leave any child behind in America.

-- 2001 (Paragraph 14 of 73)

Compared with other words

	2001*	2002	2003	2004		2006	2007
Tax	29	7	13	21	11	10	10
Afghanistan	-	13	3	5	3	2	4
Economy(ic)	6	7	13	17	14	23	8
Insurance	2	-	1	6	1	3	14
Iraq/Iraqi(s)	-	2	21	24	27	16	34
Iran	-	2	3	1	3	6	5
Oil	-	1	-	-	-	3	9
Social Security	15	2	2	2	18	3	2

* As a newly elected president, Mr. Bush did not deliver a formal State of the Union address in 2001. His Feb. 27 speech to a joint session of Congress was analogous to the State of the Union, but without the title.

Ben Werschkul/The New York Times

Bubbles => usage count

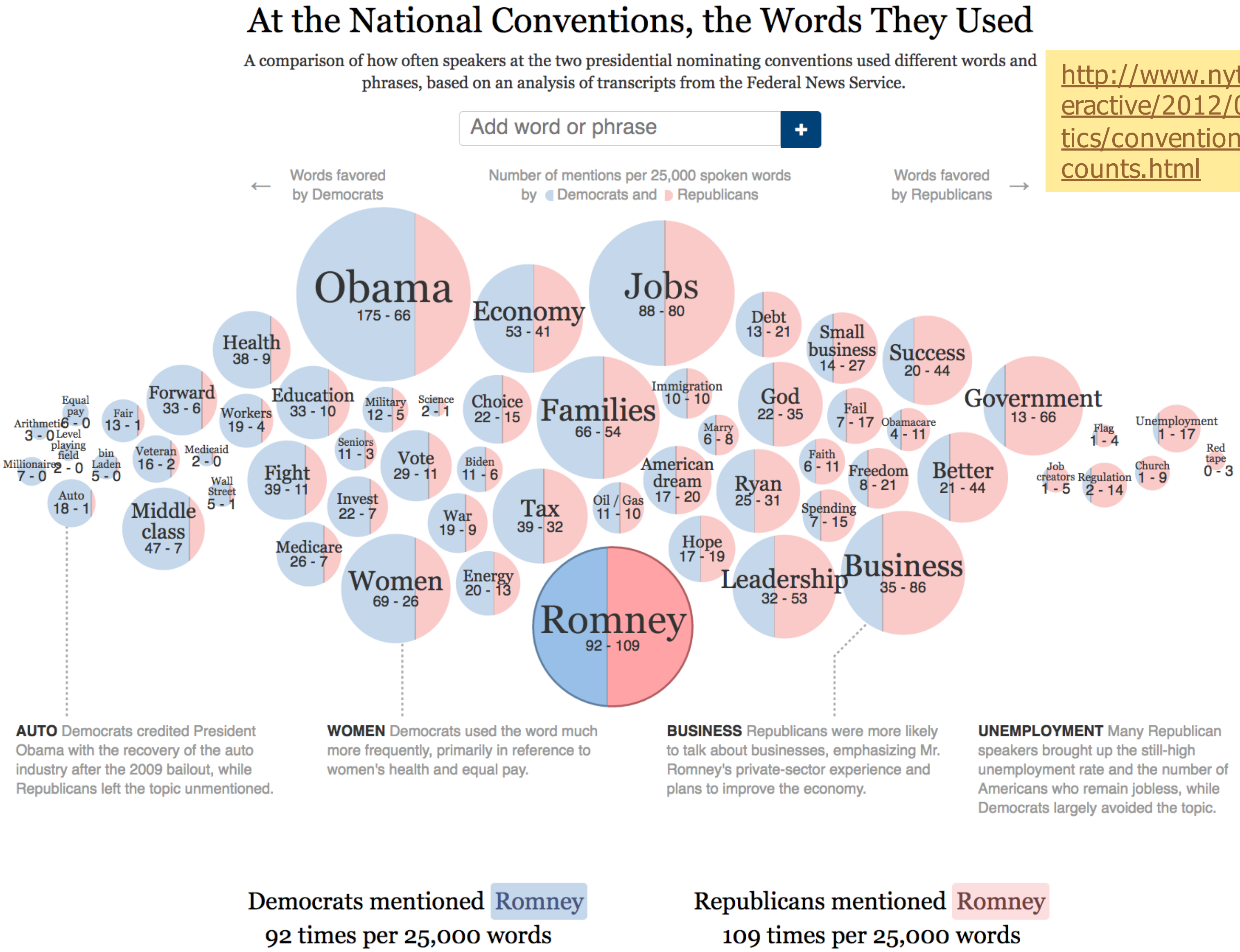
Each row is line of text. Each use of 'tax' is highlighted

Hover over a use of 'tax' for context

Putting it together: Ben Werschkul of the *New York Times (DEAD LINK ☺)*
http://www.nytimes.com/ref/washington/20070123_STATEOFUNION.html?initialWord=iraq

Word Counts – Republicans, Democrats

http://www.nytimes.com/interactive/2012/09/06/us/politics/convention-word-counts.html



At the National Conventions, the Words They Used

A comparison of how often speakers at the two presidential nominating conventions used different words and phrases, based on an analysis of transcripts from the Federal News Service.

Add word or phrase +

AUTO Democrats credited President Obama with the recovery of the auto industry after the 2009 bailout, while Republicans left the topic unmentioned.

WOMEN Democrats used the word much more frequently, primarily in reference to women's health and equal pay.

BUSINESS Republicans were more likely to talk about businesses, emphasizing Mr. Romney's private-sector experience and plans to improve the economy.

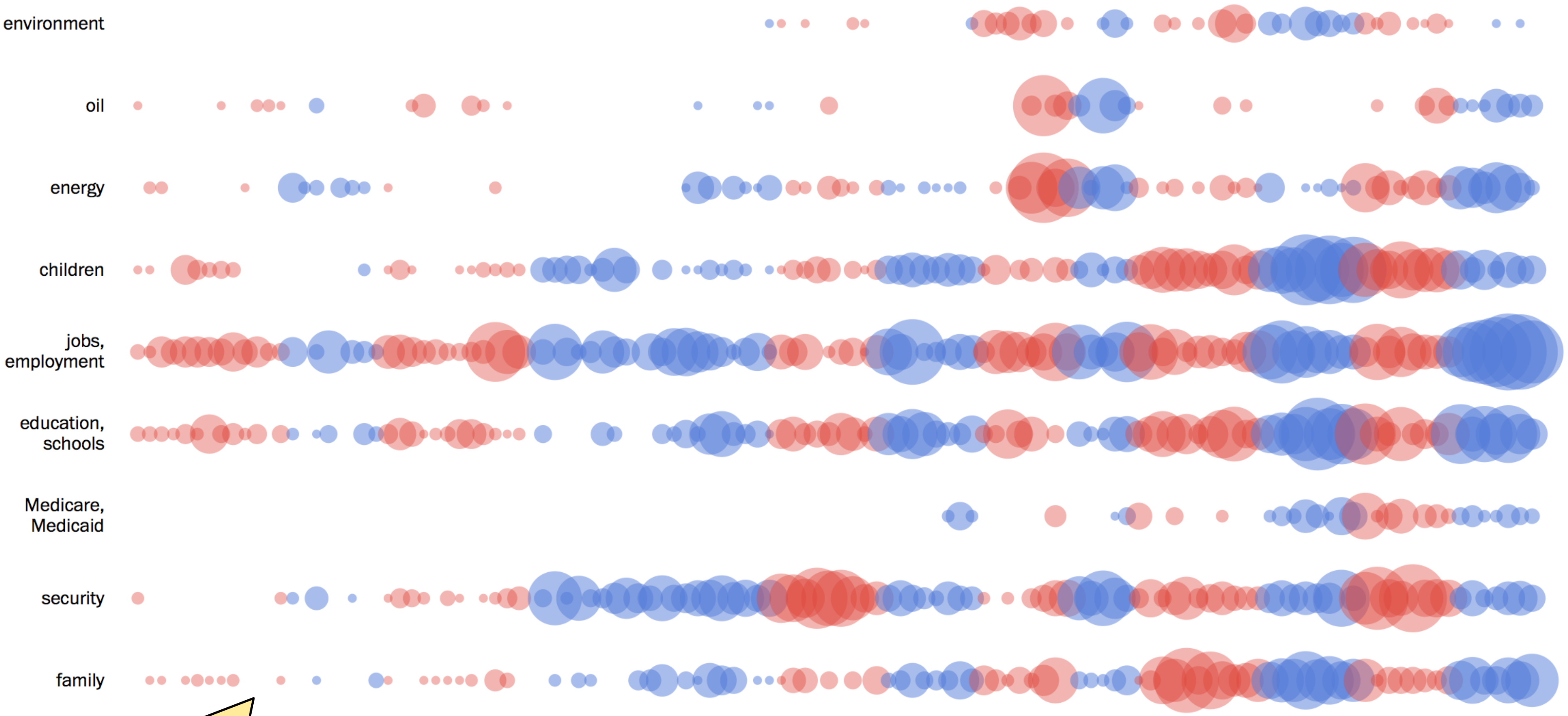
UNEMPLOYMENT Many Republican speakers brought up the still-high unemployment rate and the number of Americans who remain jobless, while Democrats largely avoided the topic.

Democrats mentioned Romney
92 times per 25,000 words

Republicans mentioned Romney
109 times per 25,000 words

Document Comparison: Washington Post

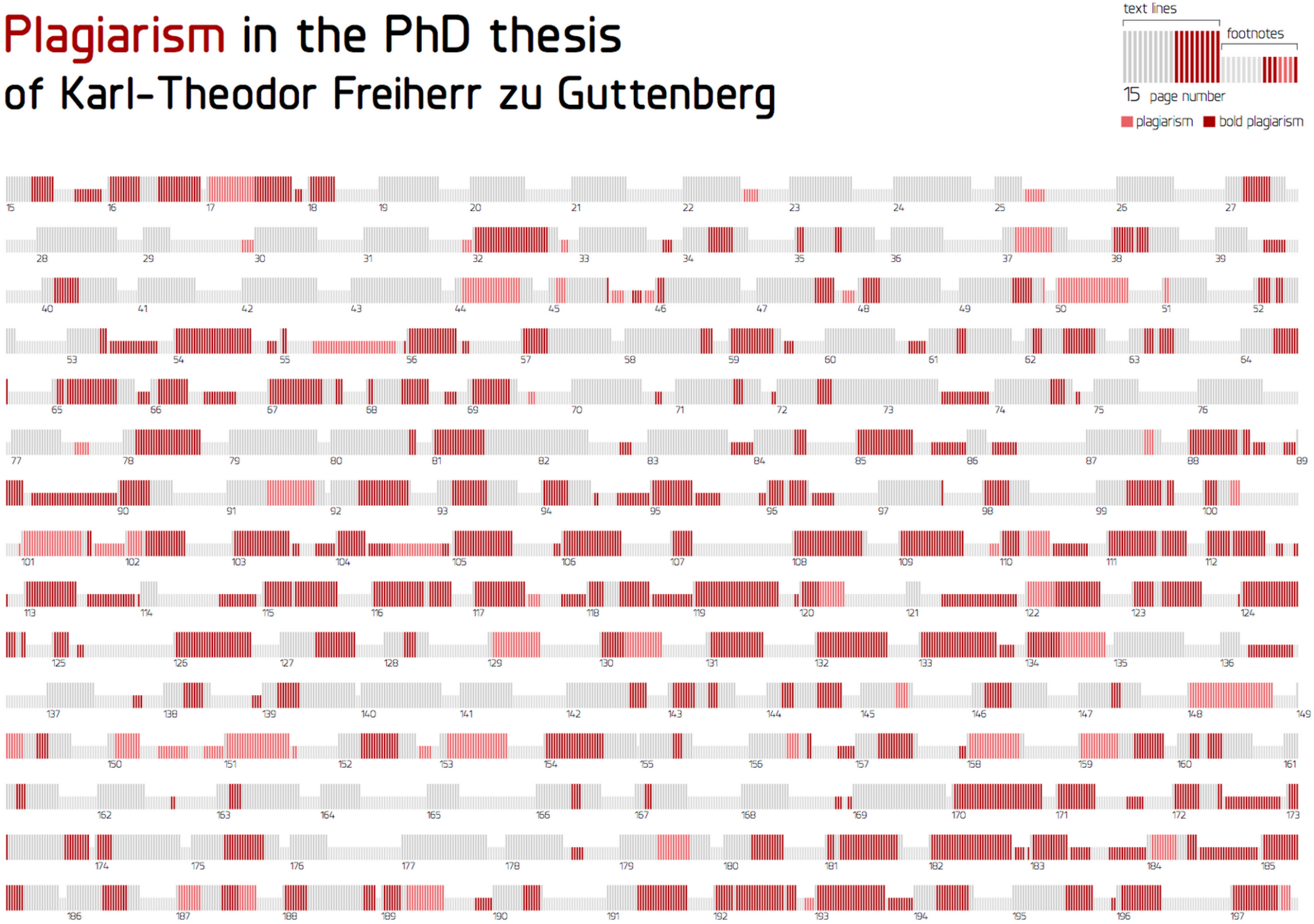
State of the Union Speeches, 1900-2016



What year? ☹

http://www.washingtonpost.com/graphics/politics/2015-sotu/language/

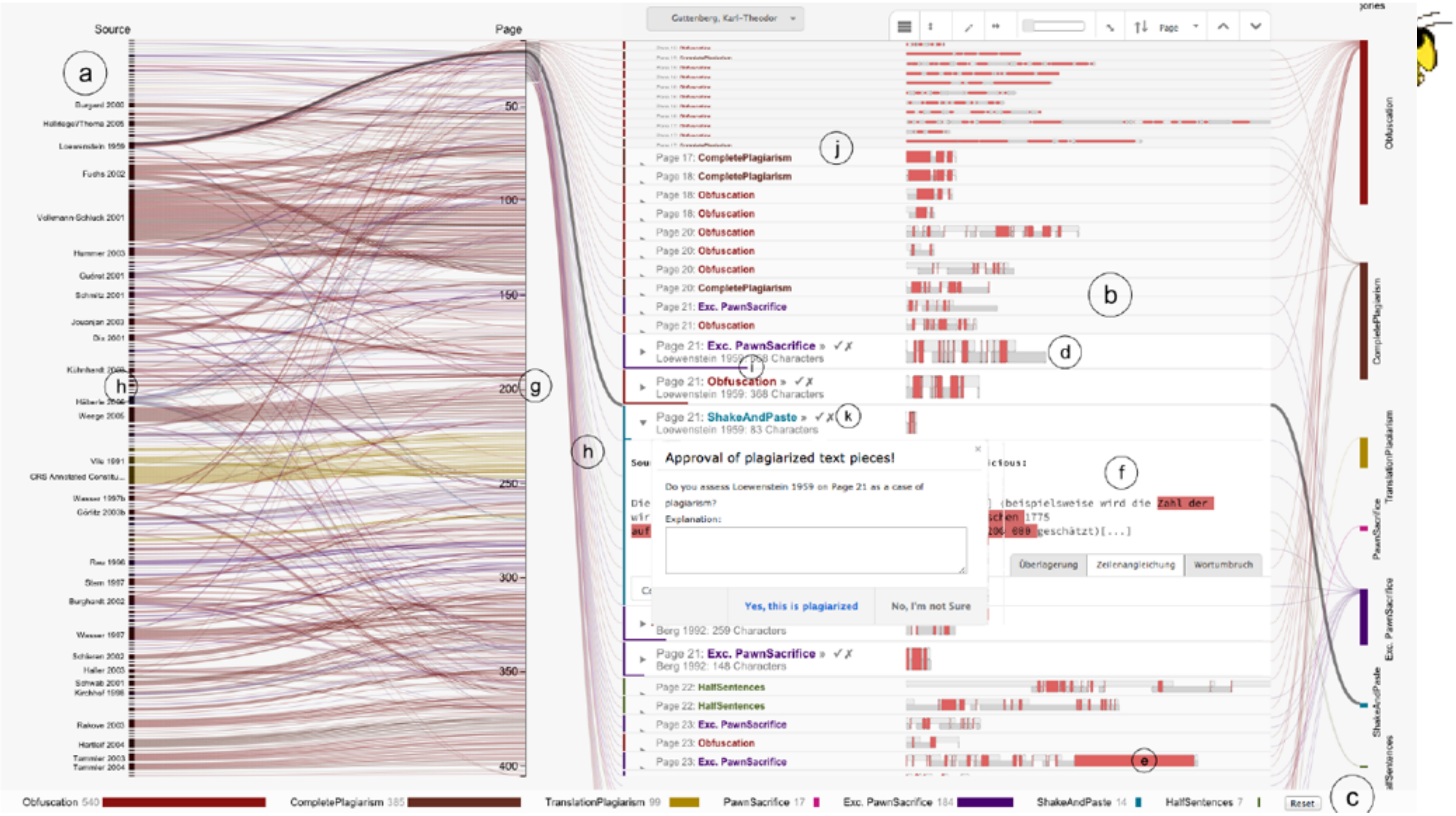
Plagiarism Example



Plagiarism in the PhD thesis
of Karl-Theodor Freiherr zu Guttenberg

Visualization of Alleged Plagiarism

Computer Graphics Forum
Volume 34, Issue 3, pages 61-70, 20 JUL 2015 DOI: 10.1111/cgf.12618
http://onlinelibrary.wiley.com/doi/10.1111/cgf.12618/full#cgf12618-fig-0001



Approval of plagiarized text pieces!

Do you assess Loewenstein 1959 on Page 21 as a case of plagiarism?

Explanation:

Yes, this is plagiarized

No, I'm not Sure

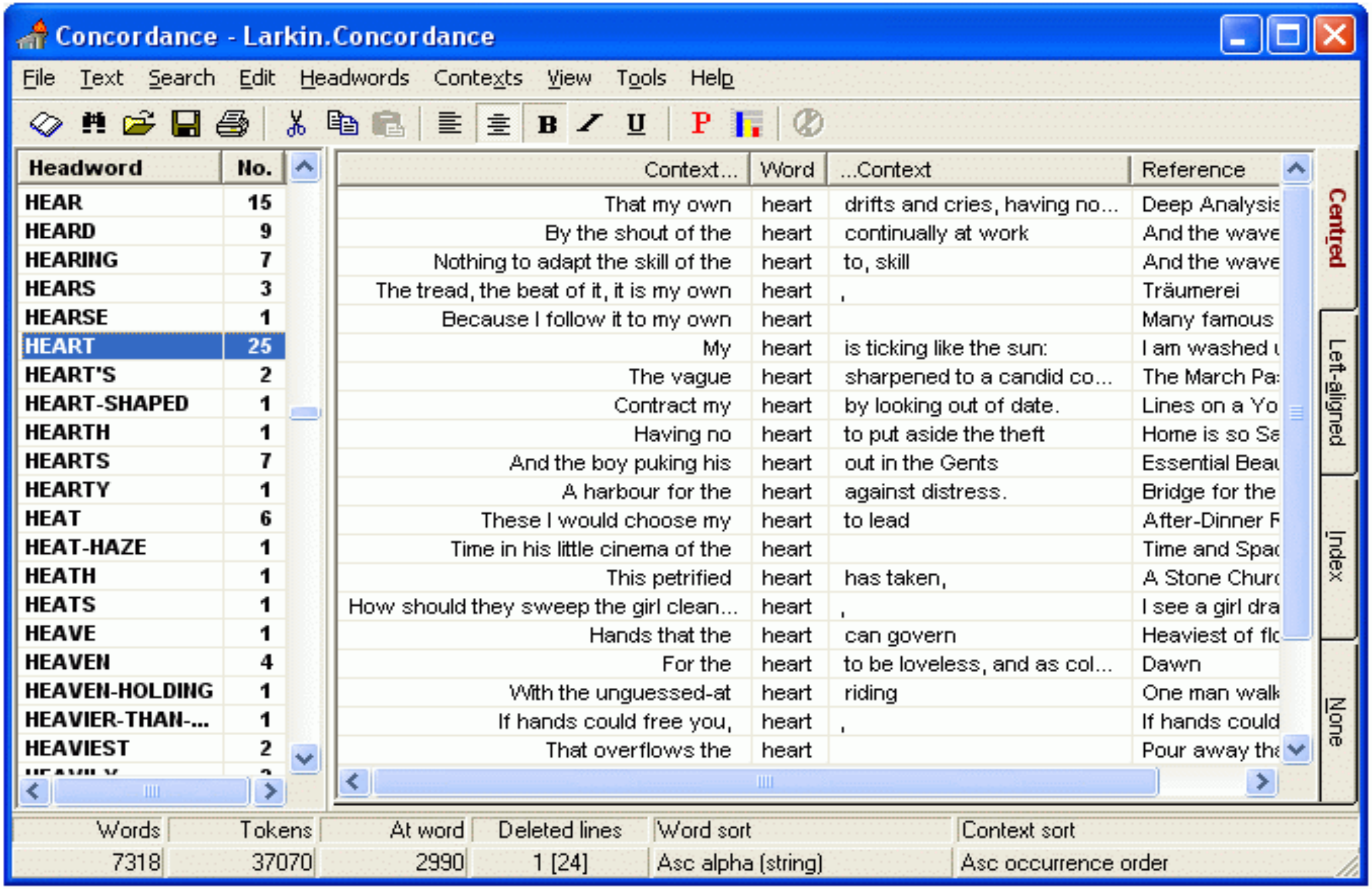
Die … wir … auf …] (beispielsweise wird die Zahl der … chen 1775 … 200 000 geschätzt)[...]

Types of plagiarism at bottom (c), a list of difflines (b) (glyph-based visualization of the finding spots (d)) in the center, overview on left (a). Copy-and-pasted passages marked in red (e). A finding spot can be opened for side-by-side comparison of suspicious and original text fragments. Overview reveals the distribution of finding spots across the document (g) and relationship to sources (h). Overview supports brushing and selection to define subset of finding spots displayed in diffline view.

Digging into Document Details

- A **concordance** is an alphabetical list of the principal words used in a book or body of work, with their immediate contexts (sometimes called KWIC or Key Word in Context)
- A **frequency list** is a sorted list of words together with their frequencies
- Word Trees
- Grammatical Structures

Concordance & Frequency List Together



Headword	No.
HEAR	15
HEARD	9
HEARING	7
HEARS	3
HEARSE	1
HEART	25
HEART'S	2
HEART-SHAPED	1
HEARTH	1
HEARTS	7
HEARTY	1
HEAT	6
HEAT-HAZE	1
HEATH	1
HEATS	1
HEAVE	1
HEAVEN	4
HEAVEN-HOLDING	1
HEAVIER-THAN-...	1
HEAVIEST	2

Context...	Word	...Context	Reference
That my own	heart	drifts and cries, having no...	Deep Analysis
By the shout of the	heart	continually at work	And the wave
Nothing to adapt the skill of the	heart	to, skill	And the wave
The tread, the beat of it, it is my own	heart	,	Träumerei
Because I follow it to my own	heart		Many famous
My	heart	is ticking like the sun:	I am washed u
The vague	heart	sharpened to a candid co...	The March Pa:
Contract my	heart	by looking out of date.	Lines on a Yo
Having no	heart	to put aside the theft	Home is so Sa
And the boy puking his	heart	out in the Gents	Essential Beau
A harbour for the	heart	against distress.	Bridge for the
These I would choose my	heart	to lead	After-Dinner R
Time in his little cinema of the	heart		Time and Spac
This petrified	heart	has taken,	A Stone Churc
How should they sweep the girl clean...	heart	,	I see a girl dra
Hands that the	heart	can govern	Heaviest of flo
For the	heart	to be loveless, and as col...	Dawn
With the unguessed-at	heart	riding	One man walk
If hands could free you,	heart	,	If hands could
That overflows the	heart		Pour away tha

Words	Tokens	At word	Deleted lines	Word sort	Context sort
7318	37070	2990	1 [24]	Asc alpha (string)	Asc occurrence order

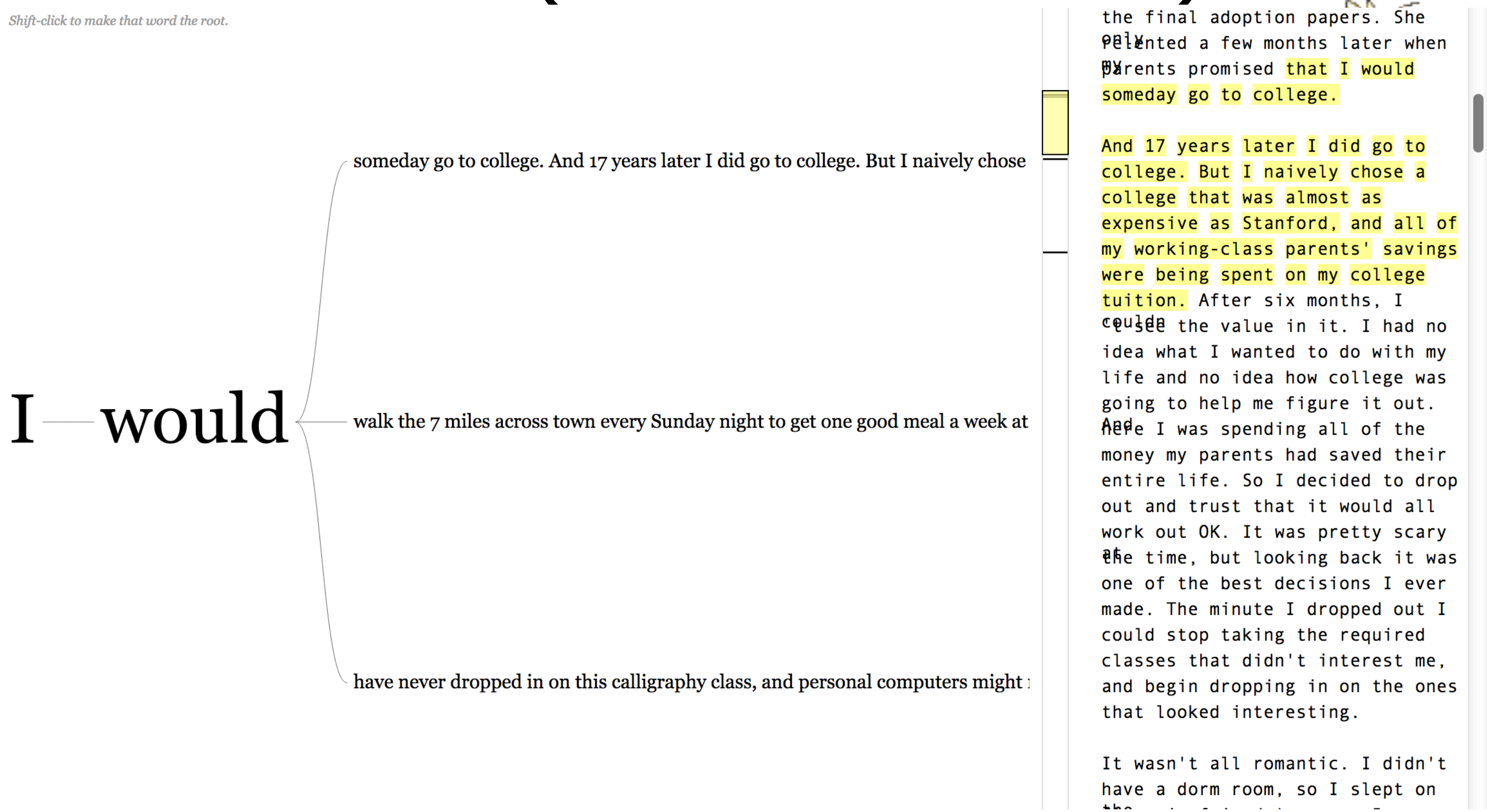
From www.concordancesoftware.co.uk

Concordance: Word Tree

- Shows context of a word or words
  - Follow word with all the phrases that follow it
- Font size shows frequency of appearance
- Continue branch until hitting unique phrase
- Clicking on phrase makes it the focus
- Ordered alphabetically, by frequency, or by first appearance

Wattenberg & Viégas
TVCG '08

Word Trees (Words in context)



Shift-click to make that word the root.

I would

someday go to college. And 17 years later I did go to college. But I naively chose

walk the 7 miles across town every Sunday night to get one good meal a week at

have never dropped in on this calligraphy class, and personal computers might

the final adoption papers. She only relented a few months later when my parents promised that I would someday go to college.

And 17 years later I did go to college. But I naively chose a college that was almost as expensive as Stanford, and all of my working-class parents' savings were being spent on my college tuition. After six months, I couldn't see the value in it. I had no idea what I wanted to do with my life and no idea how college was going to help me figure it out. And here I was spending all of the money my parents had saved their entire life. So I decided to drop out and trust that it would all work out OK. It was pretty scary at the time, but looking back it was one of the best decisions I ever made. The minute I dropped out I could stop taking the required classes that didn't interest me, and begin dropping in on the ones that looked interesting.

It wasn't all romantic. I didn't have a dorm room, so I slept on the

https://www.jasondavies.com/wordtree/?source=steve-jobs-commencement.txt&prefix=I&reverse=0

Multiple Words

- How about sequences or pairs of words?
- Are there good ways to present them?

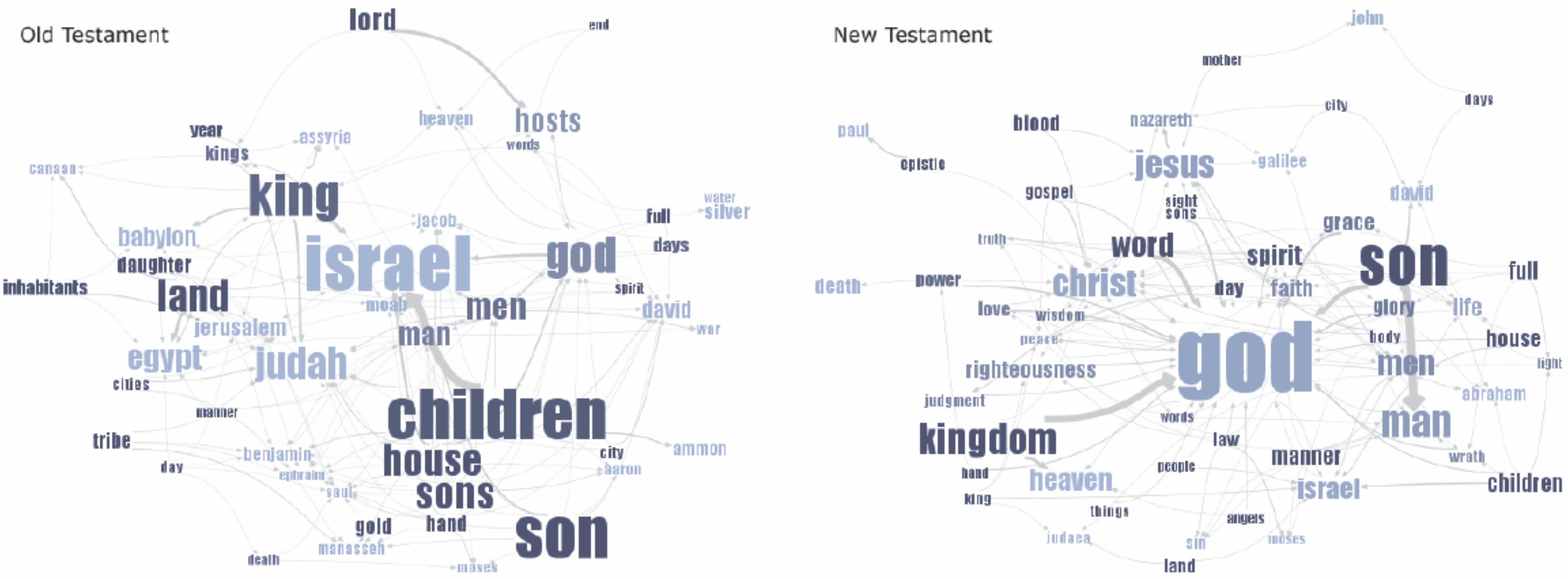
In Many Eyes now

Phrase Nets

- Examine unstructured text documents
- Presents pairs of terms from phrases such as
  - X and Y (as in "pride and prejudice")
  - X's Y (as in "Jim's trains")
  - X at Y (as in "Macy's at Lenox")
  - X (is|are|was|were) Y
- Uses special graph layout algorithm with compression and simplification

van Ham et al
TVCG '09

Phrase Net Examples


Old Testament

New Testament

Fig 4. Matching the same pattern on different texts. Here we used the pattern "X of Y" to compare the old and new testaments. Israel takes a central place in the Old Testament, while God acts as the main pattern receiver in the New Testament.

Phrase Net Examples

Requires "Entity Extraction" and Natural Language Processing

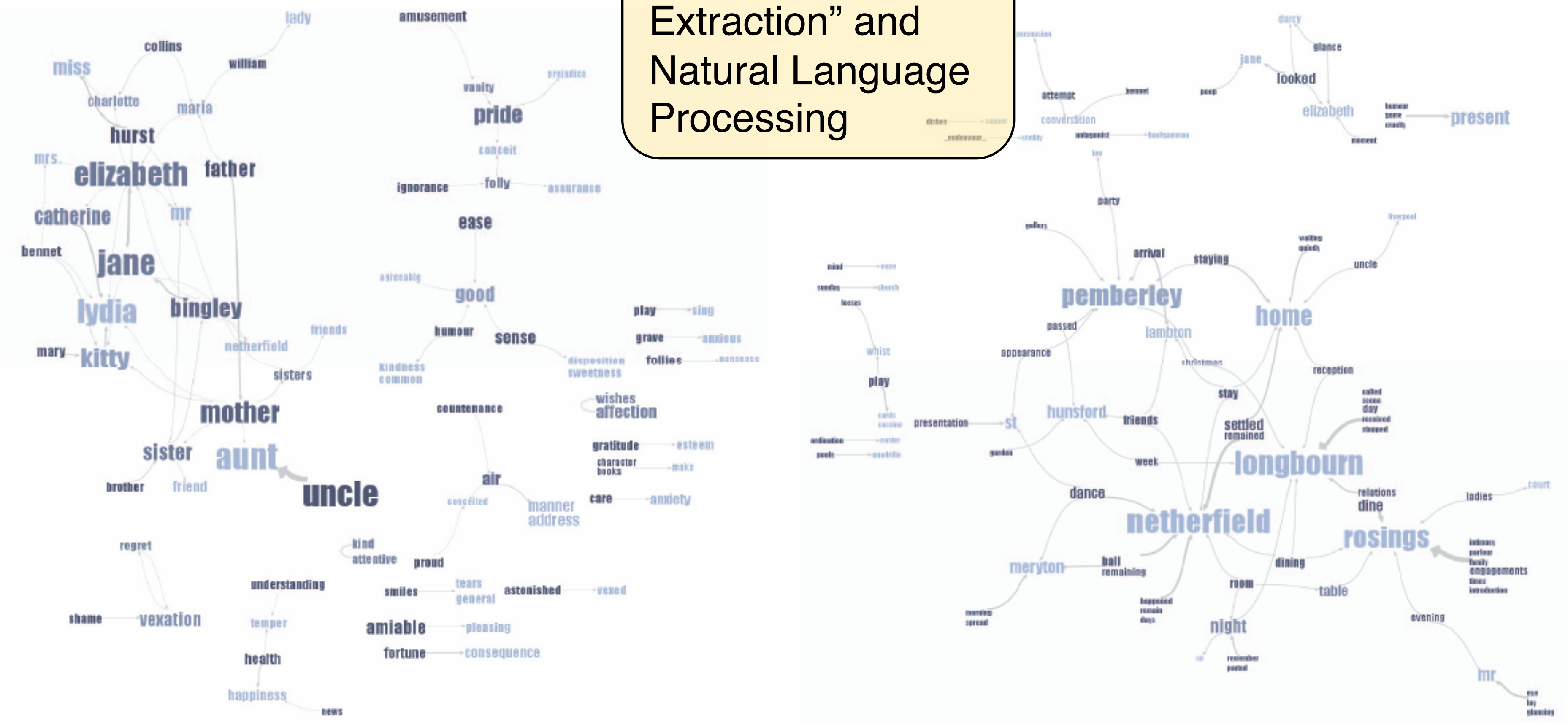


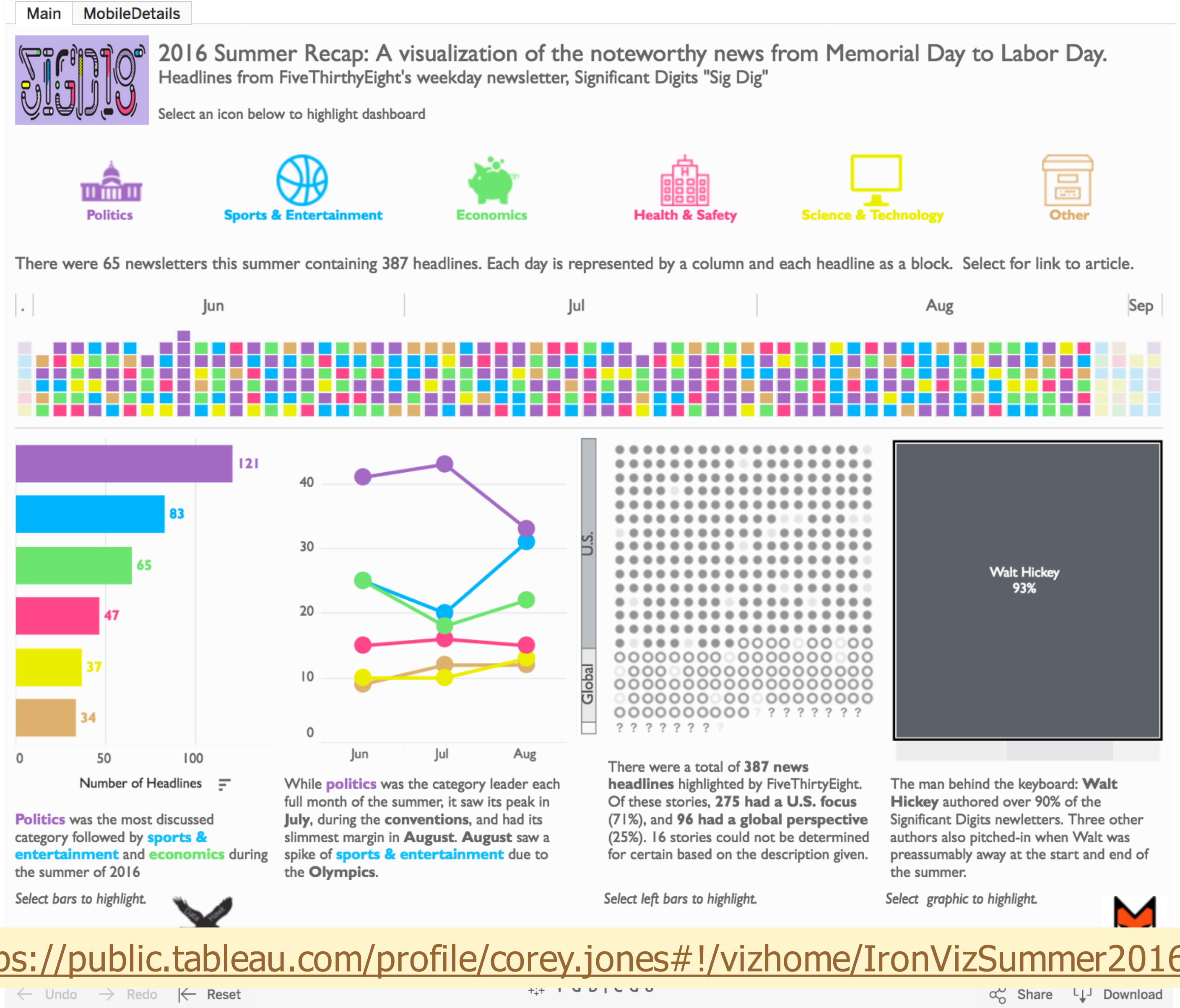
Fig 5. Matching different patterns on the same text. Here we analyzed Jane Austen's *Pride and Prejudice* with "X and Y" and "X at Y" respectively. The left image shows relationships between the main characters amongst others, while the right image shows relationships between locations.

Structured Document

- Lots of meta-data

- “Sig Dig”example
- FacetMaps
- PaperLens
- ResultMaps

News Story Trends



2016 Summer Recap: A visualization of the noteworthy news from Memorial Day to Labor Day.

Headlines from FiveThirthyEight's weekday newsletter, Significant Digits "Sig Dig"

Select an icon below to highlight dashboard

There were 65 newsletters this summer containing 387 headlines. Each day is represented by a column and each headline as a block. Select for link to article.

Politics was the most discussed category followed by sports & entertainment and economics during the summer of 2016

Select bars to highlight.

While politics was the category leader each full month of the summer, it saw its peak in July, during the conventions, and had its slimmest margin in August. August saw a spike of sports & entertainment due to the Olympics.

There were a total of 387 news headlines highlighted by FiveThirtyEight. Of these stories, 275 had a U.S. focus (71%), and 96 had a global perspective (25%). 16 stories could not be determined for certain based on the description given.

Select left bars to highlight.

The man behind the keyboard: Walt Hickey authored over 90% of the Significant Digits newletters. Three other authors also pitched-in when Walt was preassumably away at the start and end of the summer.

Select graphic to highlight.

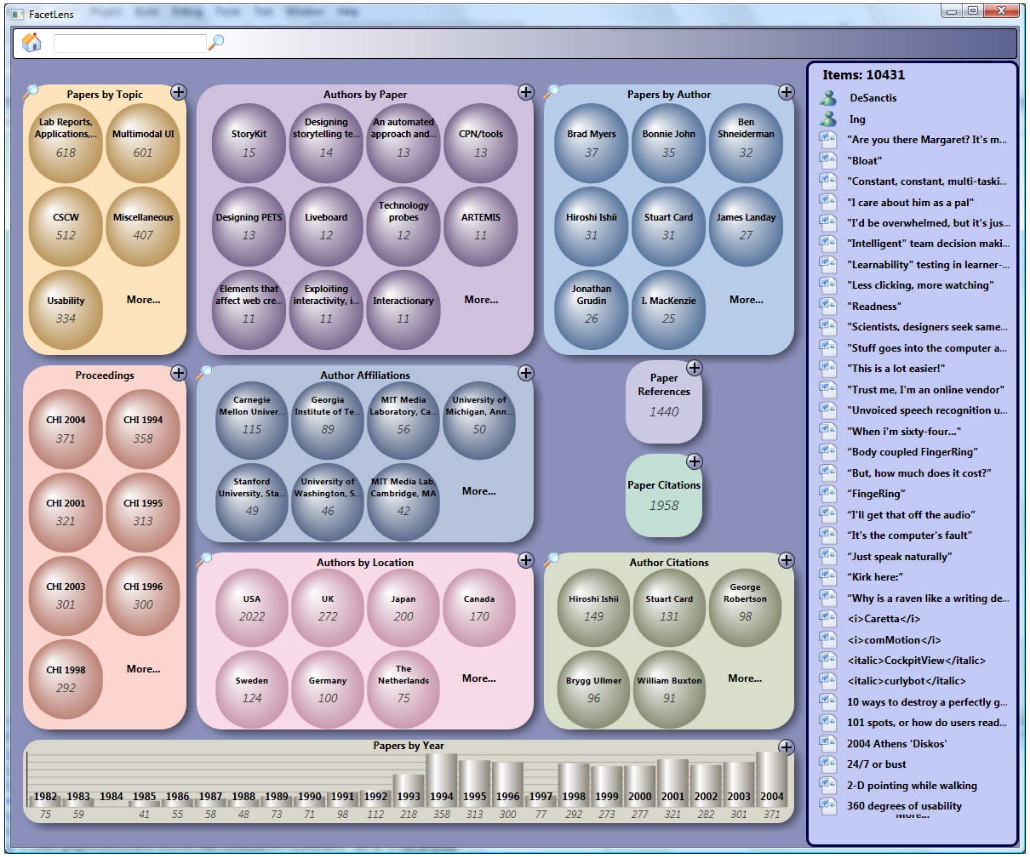
https://public.tableau.com/profile/corey.jones#!/vizhome/IronVizSummer2016/Main

Structured Info Spaces: FacetMaps

Facets: aka attributes

Displays attributes of a "case"

Drill down to one or a few cases (papers in this example)

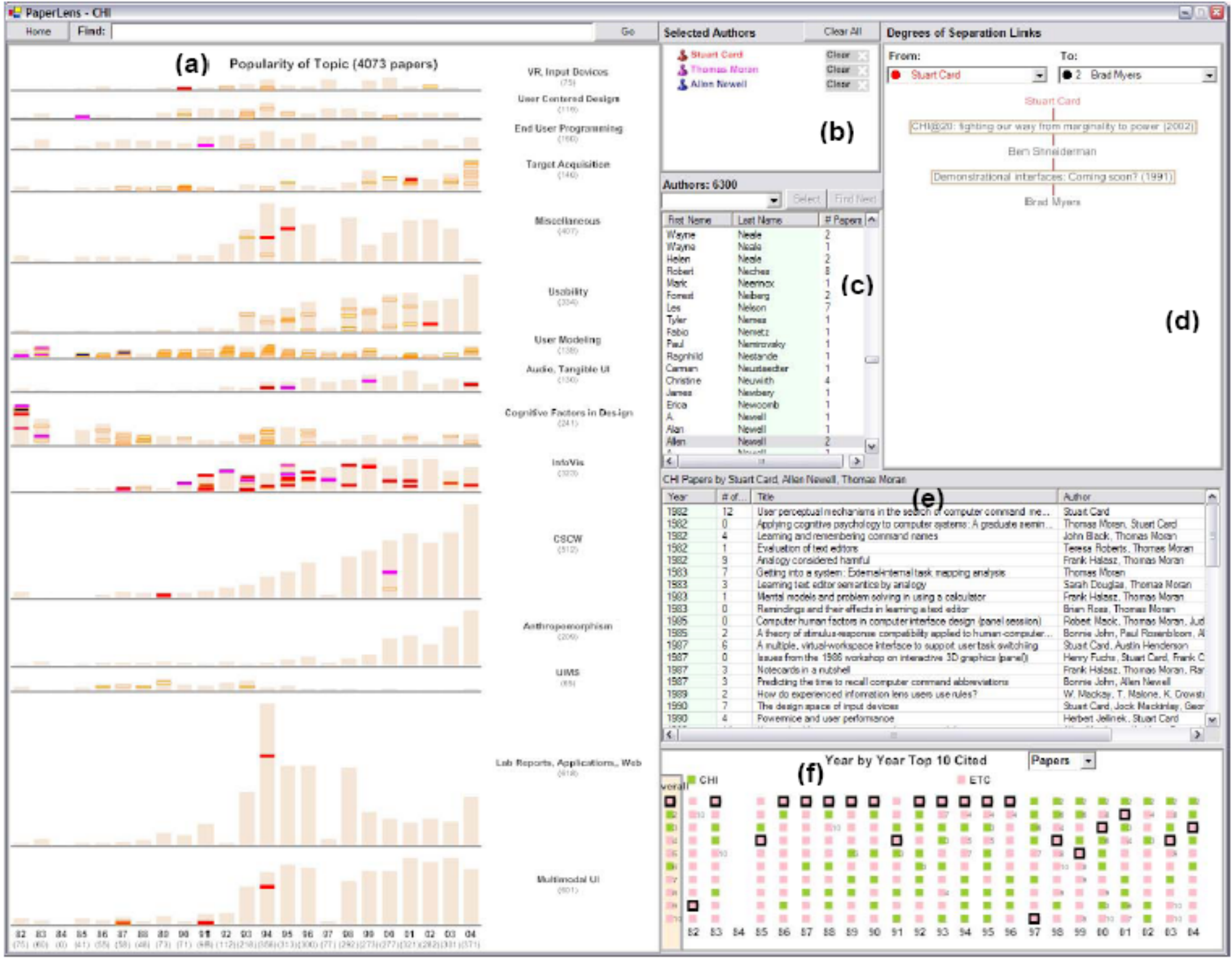


Cases (papers in this example)

Facets (attributes)

http://research.microsoft.com/en-us/um/redmond/groups/cue/facetlens/FacetLens-Video-CHI.wmv

Structured Info Spaces: PaperLens



First Name	Last Name	# Papers
Wayne	Neale	2
Wayne	Neale	1
Helen	Neale	2
Robert	Neches	8
Mark	Neerincx	1
Forrest	Neiberg	2
Les	Nelson	7
Tyler	Nemes	1
Fabio	Nemetz	1
Paul	Nemirovsky	1
Ragnhild	Nestande	1
Carman	Neustaedter	1
Christine	Neuwirth	4
James	Newbery	1
Erica	Newcomb	1
A.	Newell	1
Alan	Newell	1
Allen	Newell	2

CHI Papers by Stuart Card, Allen Newell, Thomas Moran

Year	# of...	Title	Author
1982	12	User perceptual mechanisms in the search of computer command me...	Stuart Card
1982	0	Applying cognitive psychology to computer systems: A graduate semin...	Thomas Moran, Stuart Card
1982	4	Learning and remembering command names	John Black, Thomas Moran
1982	1	Evaluation of text editors	Teresa Roberts, Thomas Moran
1982	9	Analogy considered harmful	Frank Halasz, Thomas Moran
1983	7	Getting into a system: External-internal task mapping analysis	Thomas Moran
1983	3	Learning text editor semantics by analogy	Sarah Douglas, Thomas Moran
1983	1	Mental models and problem solving in using a calculator	Frank Halasz, Thomas Moran
1983	0	Remindings and their effects in learning a text editor	Brian Ross, Thomas Moran
1985	0	Computer human factors in computer interface design (panel session)	Robert Mack, Thomas Moran, Jud
1985	2	A theory of stimulus-response compatibility applied to human-computer...	Bonnie John, Paul Rosenbloom, Al
1987	6	A multiple, virtual-workspace interface to support user task switching	Stuart Card, Austin Henderson
1987	0	Issues from the 1986 workshop on interactive 3D graphics (panel)	Henry Fuchs, Stuart Card, Frank C
1987	3	Notecards in a nutshell	Frank Halasz, Thomas Moran, Rar
1987	3	Predicting the time to recall computer command abbreviations	Bonnie John, Allen Newell
1989	2	How do experienced information lens users use rules?	W. Mackay, T. Malone, K. Crowst
1990	7	The design space of input devices	Stuart Card, Jock Mackinlay, Geor
1990	4	Powermice and user performance	Herbert Jellinek, Stuart Card

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

Video on next slide

PaperLens - Conclusion

- Shows all the data at once but not all the relationships
- Leverages tightly coupled views to show correlations
- Effective in answering questions regarding:
  - Patterns such as frequency of authors and papers cited
  - Themes are clearly visible
  - Trends such as number of papers published in a topic area over time
  - Correlations between authors, topics and citations
- User study
  - Average task performed in less than 20 seconds
  - Participants provided correct answers to tasks 97% of the time
- Scaling challenge
  - Fish-eye technique not as effective with larger number of papers
  - Over-lapping highlights
  - No-more than four authors can be color-coded in the overview

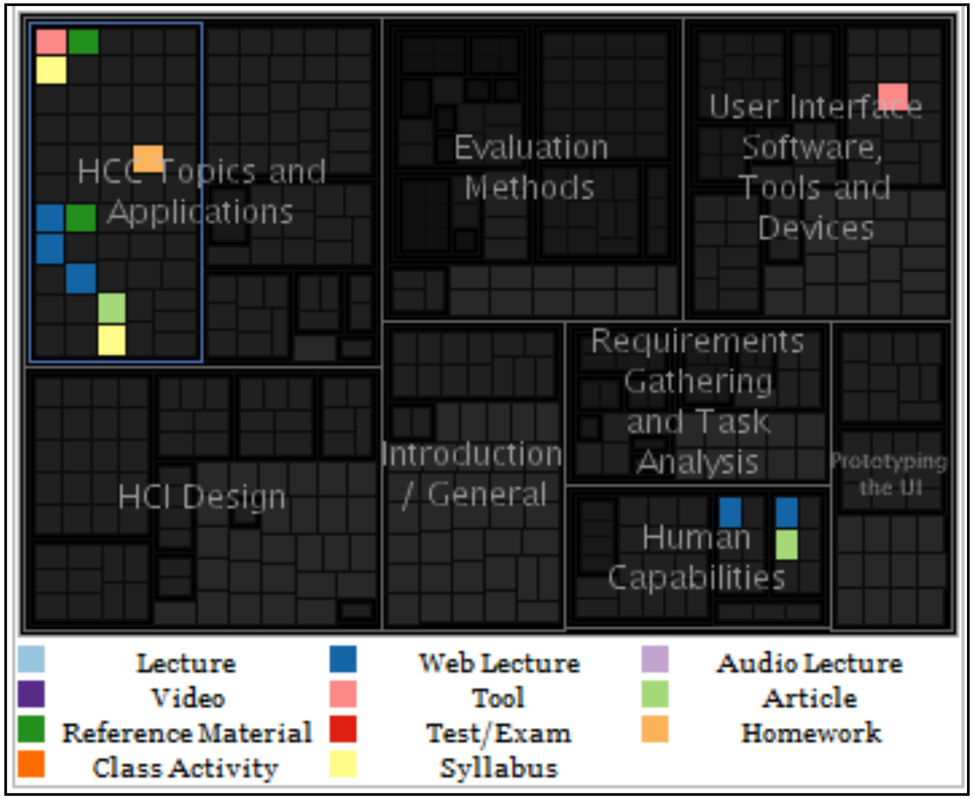
- Lee, Czerwinski, Robertson, Bederson, Understanding Research Trends in Conferences using PaperLens, Extended Abstracts of CHI 2005, pp. 1969-1972.

PaperLens Slides by Jennifer Stoll, GT

Structured Info Spaces: ResultMaps

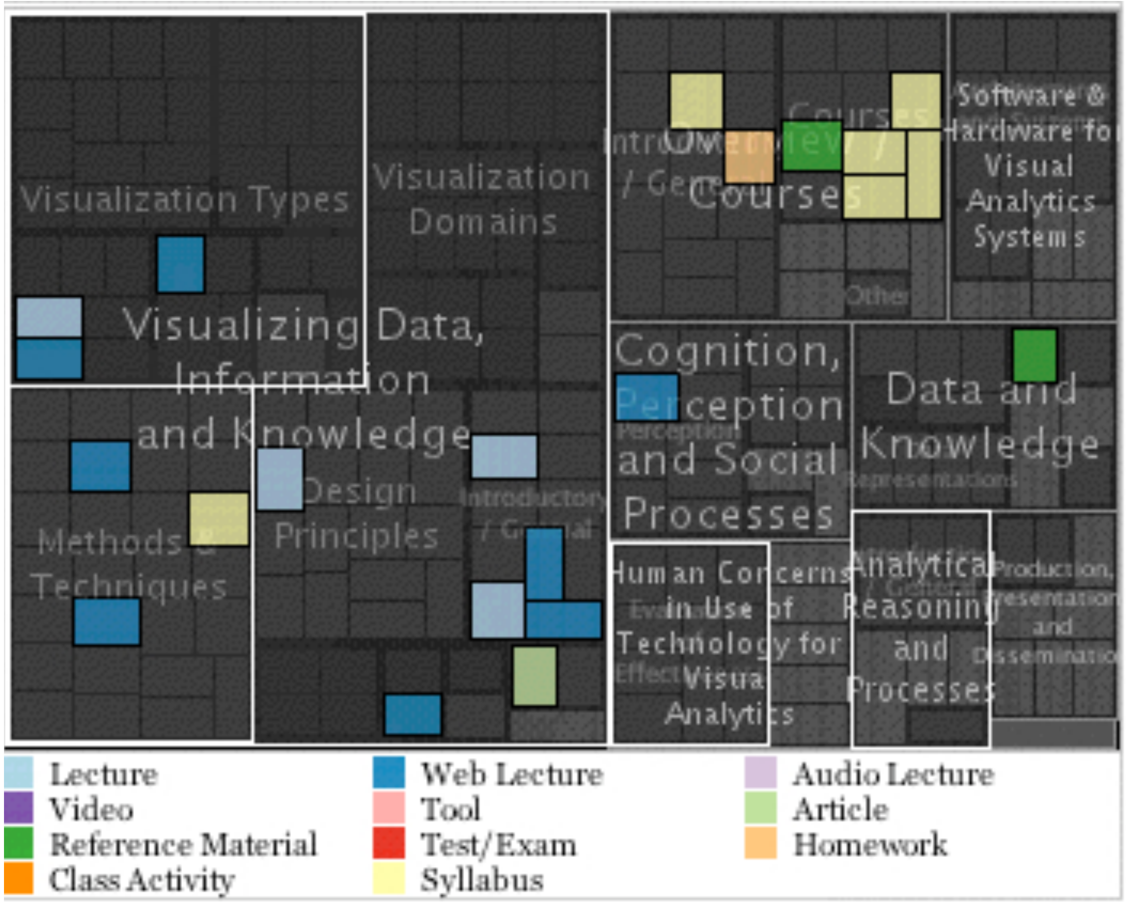
- Problem - understand a HIERARCHICAL information space and how retrieved results fit into that space
  - Contrast with PaperLens
- Solution - ResultMaps
  - Based on TreeMaps
  - Highlight documents retrived via text query
  - Linkage from highlight to document in retrieval list

- BTW – do a quick critique of this InfoVis



ResultMap

- Experimental evaluation
  - Compare with Google-style results list
  - Not much help ☹
  - Some evidence that they are subjectively preferred and help understand overall document collection structure


Visualization Types
Visualization Domains
Intro/Overview / General
Courses
Software & Hardware for Visual Analytics Systems
Other
Visualizing Data, Information and Knowledge
Design Principles
Introductory / General
Methods & Techniques
Cognition, Perception and Social Processes
Perception
Data and Knowledge Representations
Human Concerns in Use of Technology for Visual Analytics
Analytical Reasoning and Processes
Production, Presentation and Dissemination
Lecture
Video
Reference Material
Class Activity
Web Lecture
Tool
Test/Exam
Syllabus
Audio Lecture
Article
Homework

- (Is this InfoVis any better than previous?)

How to Think About all of this?

- Remember this outline?
- Macro-level – searching larger document collections
  - Unstructured – no meta-data
  - Structured – explicit meta-data
- Micro-level
  - Inter-document methods for smaller document collections

    How do retrieved documents relate to a query?

    How do retrieved documents relate to one another?
  - Intra-document methods

    Word usage, grammatical style, …

- With the caveat that some methods can be used in multiple ways

Text and Documents Takeaways

- It’s a hugh space – need to understand
  - From searching everything (WWW) to analyzing a single document
- Many opportunities for creativity
- What are user activities with Text and Documents? How can InfoVis support those activities?

Text and Documents Takeaways

- Which methods scale from one or a few documents to thousands of docs on up to the WWW? Why? Why not?
- How do we know which methods are good and which are not so good?
- Are there places where using InfoVis does not make sense? What are they?

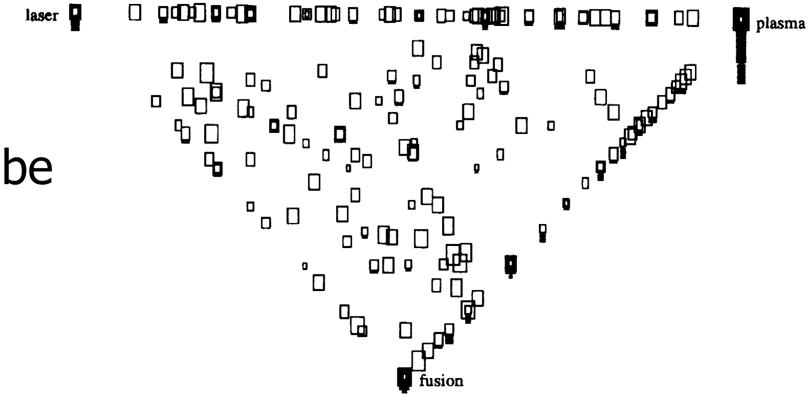
Can Combine, Mix and Match

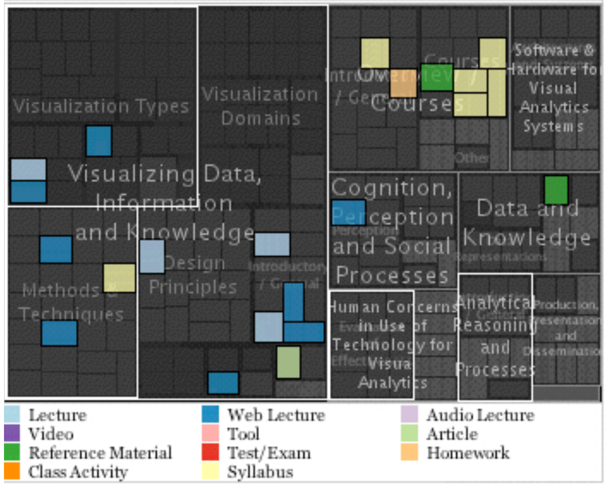
- From big picture overview of many docs to query-related views to detailed views of a few docs to within a single doc
- Add interactions to info presentations – the usual suspects
  - DoD
  - Dynamic queries/filters
  - Animation
  - Brush/Link

Compare and Contrast

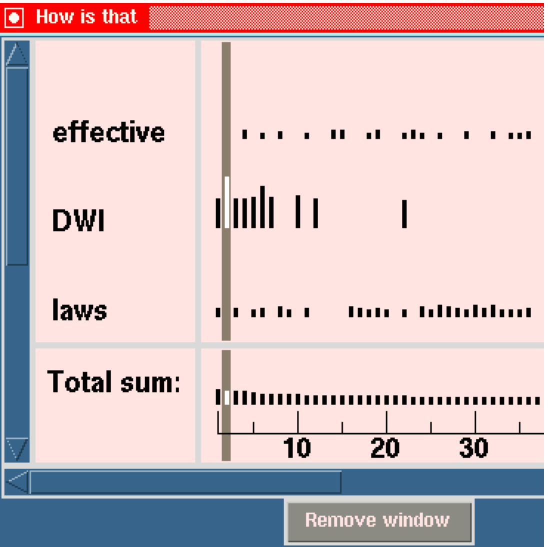
ResultMap

Vibe

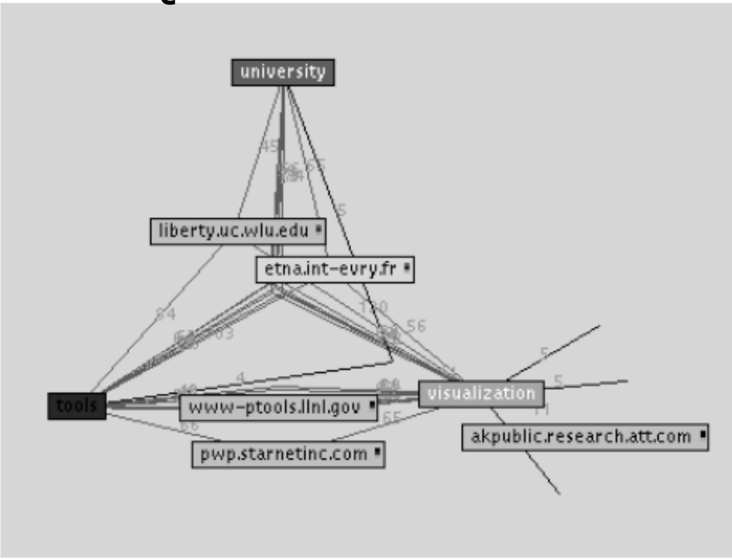



Visualization Types
Visualization Domains
Introduction / General
Courses
Software & Hardware for Visual Analytics Systems
Other
Visualizing Data, Information and Knowledge
Design Principles
Methods & Techniques
Cognition, Perception and Social Processes
Data and Knowledge Representations
Human Concerns in Use of Technology for Visual Analytics
Analytical Reasoning and Processes
Production, Presentation and Dissemination
Lecture
Video
Reference Material
Class Activity
Web Lecture
Tool
Test/Exam
Syllabus
Audio Lecture
Article
Homework

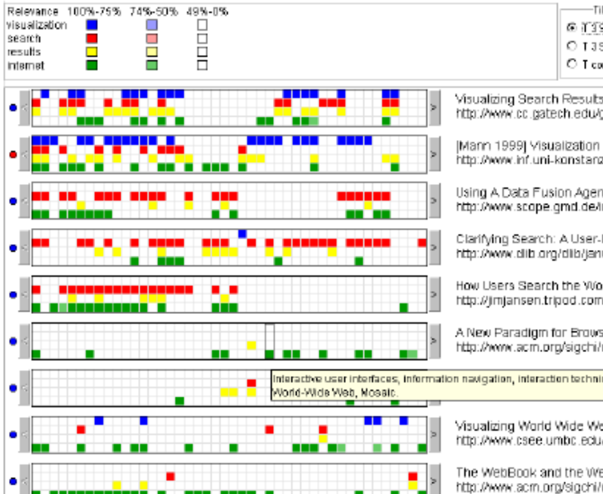
Verasamy



SQWID



TileBars


Relevance 100%-75% 74%-50% 49%-0%
visualization
search
results
internet
Visualizing Search Results
http://www.cc.gatech.edu/
[Mann 1999] Visualization
http://www.inf.uni-konstanz
Using A Data Fusion Agen
http://www.scope.gmd.de/
Clarifying Search: A User-
http://www.dlib.org/dlib/jan
How Users Search the Wo
http://jimjansen.tripod.com
A New Paradigm for Brows
http://www.acm.org/sigchi/
Interactive user interfaces, information navigation, interaction techni
World-Wide Web, Mosaic.
Visualizing World Wide We
http://www.csee.umbc.edu
The WebBook and the We
http://www.acm.org/sigchi/