Text and Document Visualization 1

CS 7450 - Information Visualization
October 31, 2016
John Stasko

Learning Objectives

- Explain key challenges in visualizing a large document or body of text
- Identify and explain different techniques for representing words and concepts in a document
  - Word cloud, Wordle, Parallel tag cloud, SeeSoft, WordTree, PhraseNet, SentenTree, TextArc
- Understand the positives and limitations of word clouds and Wordles
- Describe SeeSoft-style miniature visual representations
- Explain what word concordance is
- Describe how WordTree representation works
- Identify and explain the techniques:
  - Word cloud, Wordle, Parallel tag cloud, SeeSoft, WordTree, PhraseNet, SentenTree, TextArc

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
  - ...

Big Question

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

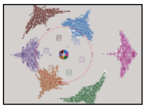
Challenge

- Text is nominal data
  - Does not seem to map to geometric/graphical presentation as easily as ordinal and quantitative data
- The “Raw data --> Data Table” mapping now becomes more important

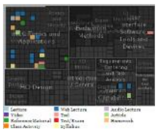
Related Topic - IR

- Information Retrieval
  - Active search process that brings back particular/specific items (will discuss that some today, but not always focus)
  - I think InfoVis and HCI can help some…
- 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

This Week's Agenda

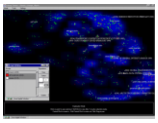


Visualization for IR
Helping search

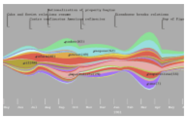


Visualizing text
Showing words, phrases, and sentences

→



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



Today

Next time

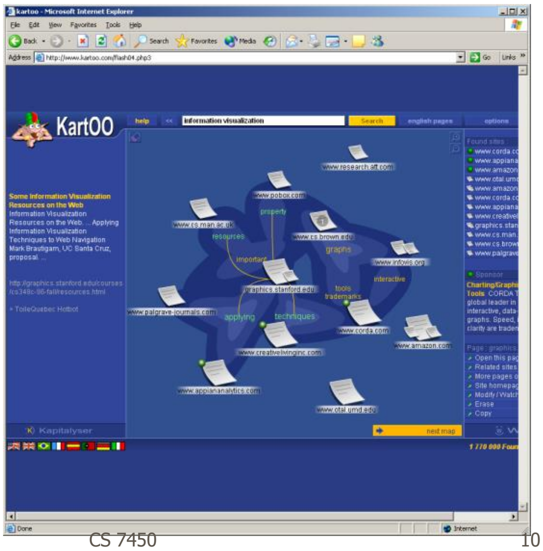
Information Retrieval

- Can InfoVis help IR?
- Assume there is some active search or query
  - Show results visually
  - Show how query terms relate to results
  - …

Generalize More

- How about the "holy grail" of a visual search engine?
  - Hot idea for a while
- My personal view: It's a mistake in the general case. Text is just better for this.

Search Visualization



Some Information Visualization Resources on the Web
Information Visualization Resources on the Web ... Applying Information Visualization Techniques to Web Navigation Mark Brautigam, UC Santa Cruz, proposal ...

http://graphics.stanford.edu/courses/cs348c-96-fall/resources.html

» ToileQuebec Hotbot

Charting/Graphi Tools CORDA T global leader in interactive, data- graphs. Speed, clarity are trader

http://www.kartoo.com

Defunct

RankSpiral

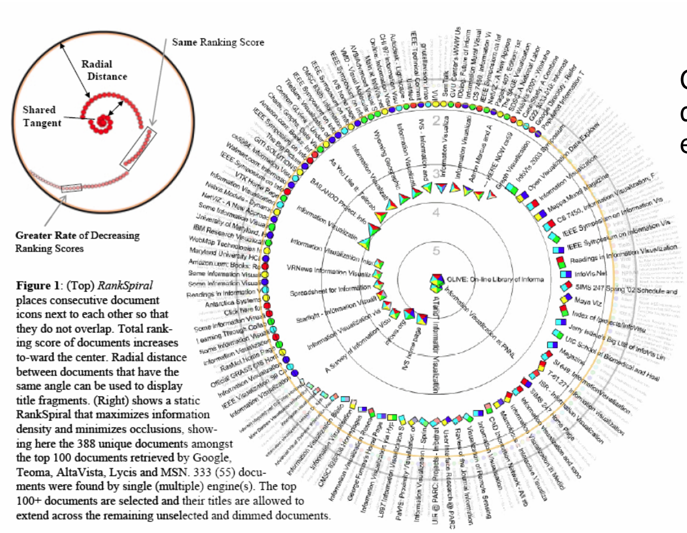
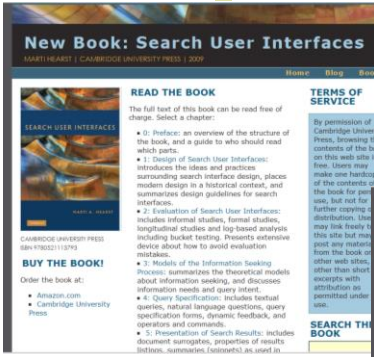


Figure 1: (Top) *RankSpiral* places consecutive document icons next to each other so that they do not overlap. Total ranking score of documents increases to-ward the center. Radial distance between documents that have the same angle can be used to display title fragments. (Right) shows a static RankSpiral that maximizes information density and minimizes occlusions, showing here the 388 unique documents amongst the top 100 documents retrieved by Google, Teoma, AltaVista, Lycis and MSN. 333 (55) documents were found by single (multiple) engine(s). The top 100+ documents are selected and their titles are allowed to extend across the remaining unselected and dimmed documents.

Color represents different search engines

Spoerri
InfoVis '04 poster

To Learn More



New Book: Search User Interfaces

MARTI HEARST | CAMBRIDGE UNIVERSITY PRESS | 2009

Home Blog Book

SEARCH USER INTERFACES

MARTI A. HEARST

CAMBRIDGE UNIVERSITY PRESS
ISBN 9780521113793

BUY THE BOOK!

Order the book at:

- Amazon.com
- Cambridge University Press

READ THE BOOK

The full text of this book can be read free of charge. Select a chapter:

- 0: Preface: an overview of the structure of the book, and a guide to who should read which parts.
- 1: Design of Search User Interfaces: introduces the ideas and practices surrounding search interface design, places modern design in a historical context, and summarizes design guidelines for search interfaces.
- 2: Evaluation of Search User Interfaces: includes informal studies, formal studies, longitudinal studies and log-based analysis including bucket testing. Presents extensive device about how to avoid evaluation mistakes.
- 3: Models of the Information Seeking Process: summarizes the theoretical models about information seeking, and discusses information needs and query intent.
- 4: Query Specification: includes textual queries, natural language questions, query specification forms, dynamic feedback, and operators and commands.
- 5: Presentation of Search Results: includes document surrogates, properties of results listings, summaries (snippets) as used in

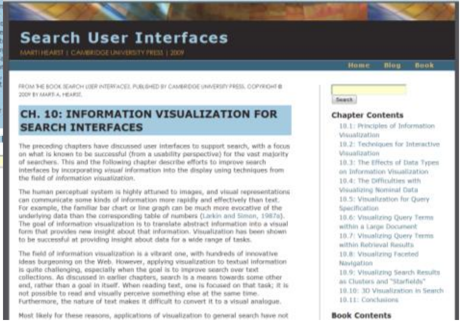
TERMS OF SERVICE

By permission of Cambridge University Press, browsing the contents of the book on this web site is free. Users may make one hardcopy of the contents of the book for per use, but not for further copying distribution. Use may link freely t this site but ma post any materia from the book o other web sites, other than short excerpts with attribution as permitted under use.

SEARCH THI BOOK

Marti Hearst's Book

Chapter 10



Search User Interfaces

MARTI HEARST | CAMBRIDGE UNIVERSITY PRESS | 2009

Home Blog Book

FROM THE BOOK SEARCH USER INTERFACES, PUBLISHED BY CAMBRIDGE UNIVERSITY PRESS. COPYRIGHT © 2009 BY MARTI A. HEARST.

CH. 10: INFORMATION VISUALIZATION FOR SEARCH INTERFACES

The preceding chapters have discussed user interfaces to support search, with a focus on what is known to be successful (from a usability perspective) for the vast majority of searchers. This and the following chapter describe efforts to improve search interfaces by incorporating *visual* information into the display using techniques from the field of *information visualization*.

The human perceptual system is highly attuned to images, and visual representations can communicate some kinds of information more rapidly and effectively than text. For example, the familiar bar chart or line graph can be much more evocative of the underlying data than the corresponding table of numbers (Larkin and Simon, 1987a). The goal of information visualization is to translate abstract information into a visual form that provides new insight about that information. Visualization has been shown to be successful at providing insight about data for a wide range of tasks.

The field of information visualization is a vibrant one, with hundreds of innovative ideas burgeoning on the Web. However, applying visualization to textual information is quite challenging, especially when the goal is to improve search over text collections. As discussed in earlier chapters, search is a means towards some other end, rather than a goal in itself. When reading text, one is focused on that task; it is not possible to read and visually perceive something else at the same time. Furthermore, the nature of text makes it difficult to convert it to a visual analogue.

Most likely for these reasons, applications of visualization to general search have not

Chapter Contents

- 10.1: Principles of Information Visualization
- 10.2: Techniques for Interactive Visualization
- 10.3: The Effects of Data Types on Information Visualization
- 10.4: The Difficulties with Visualizing Nominal Data
- 10.5: Visualization for Query Specification
- 10.6: Visualizing Query Terms within a Large Document
- 10.7: Visualizing Query Terms within Retrieval Results
- 10.8: Visualizing Faceted Navigation
- 10.9: Visualizing Search Results as Clusters and "Starfields"
- 10.10: 3D Visualization in Search
- 10.11: Conclusions

Book Contents

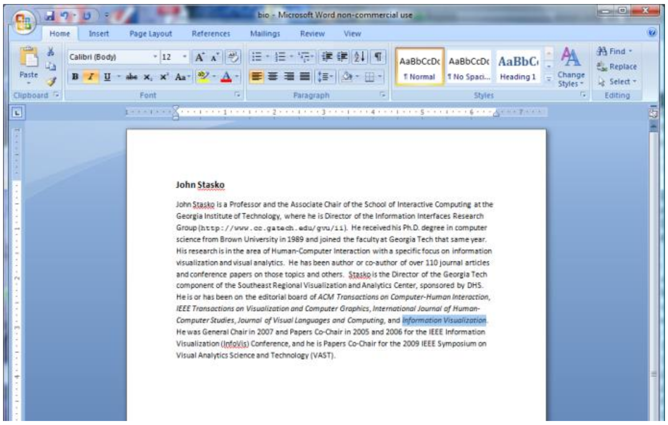
http://searchuserinterfaces.com/book/

Transition 1

- OK, let’s move up beyond just search/IR

- How do we represent the words, phrases, and sentences in a document or set of documents?
  - Main goal of *understanding* versus search

One Text Visualization



John Stasko

John Stasko is a Professor and the Associate Chair of the School of Interactive Computing at the Georgia Institute of Technology, where he is Director of the Information Interfaces Research Group (`http://www.cc.gatech.edu/gvu/ii`). He received his Ph.D. degree in computer science from Brown University in 1989 and joined the faculty at Georgia Tech that same year. His research is in the area of Human-Computer Interaction with a specific focus on information visualization and visual analytics. He has been author or co-author of over 110 journal articles and conference papers on those topics and others. Stasko is the Director of the Georgia Tech component of the Southeast Regional Visualization and Analytics Center, sponsored by DHS. He is or has been on the editorial board of *ACM Transactions on Computer-Human Interaction*, *IEEE Transactions on Visualization and Computer Graphics*, *International Journal of Human-Computer Studies*, *Journal of Visual Languages and Computing*, and *Information Visualization*. He was General Chair in 2007 and Papers Co-Chair in 2005 and 2006 for the IEEE Information Visualization (InfoVis) Conference, and he is Papers Co-Chair for the 2009 IEEE Symposium on Visual Analytics Science and Technology (VAST).

Uses:
Layout
Font
Style
Color
…

Design Challenge

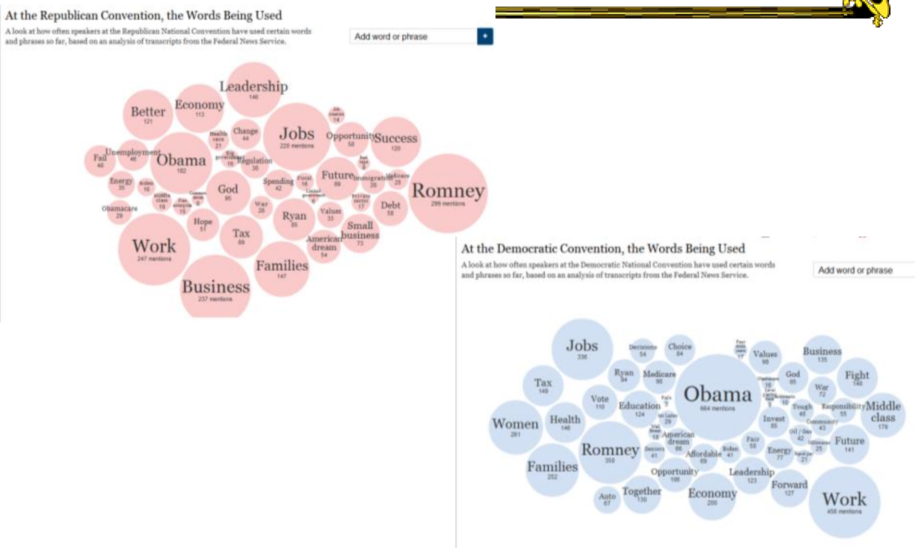
- How would you visualize one of the recent presidential debates?
- Ideas?

Tasks

- What kinds of questions or tasks would someone want to do with such a visualization?

http://www.nytimes.com/interactive/2012/08/28/us/politics/convention-word-counts.html

Word Counts



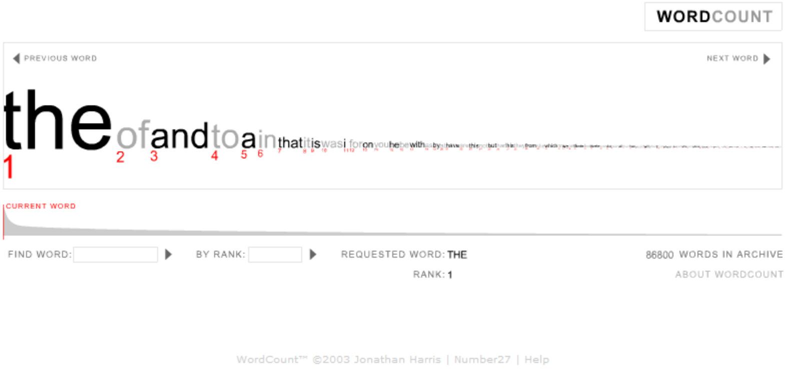
At the Republican Convention, the Words Being Used

A look at how often speakers at the Republican National Convention have used certain words and phrases so far, based on an analysis of transcripts from the Federal News Service.

At the Democratic Convention, the Words Being Used

A look at how often speakers at the Democratic National Convention have used certain words and phrases so far, based on an analysis of transcripts from the Federal News Service.

More Word Counting



WORDCOUNT
PREVIOUS WORD
NEXT WORD
the of and to a in that it is was i for on you he be with as by have ... this ... but ... from ... which
1 2 3 4 5 6 7 8 9 10 11 12
CURRENT WORD

FIND WORD: ▶ BY RANK: ▶ REQUESTED WORD: THE

RANK: 1

86800 WORDS IN ARCHIVE

ABOUT WORDCOUNT

WordCount™ ©2003 Jonathan Harris | Number27 | Help

http://www.wordcount.org

Tag/Word Clouds

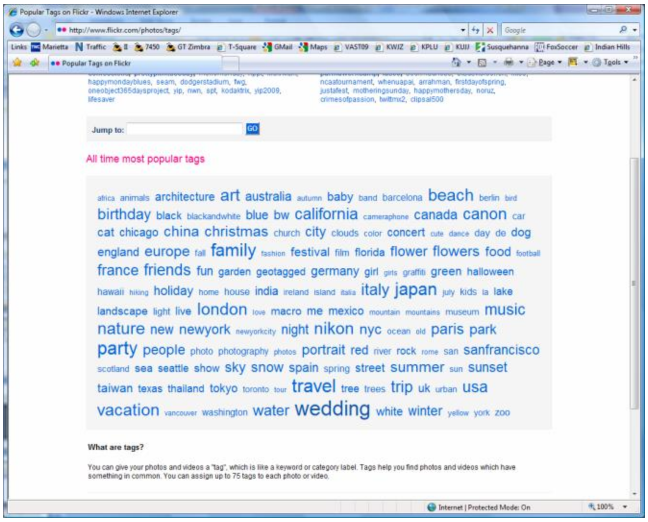
- Currently very "hot" in research community
- Have proven to be very popular on web
- Idea is to show word/concept importance through visual means
  - Tags: User-specified metadata (descriptors) about something
  - Sometimes generalized to just reflect word frequencies

History

- 90-year old Soviet Constructivism
- Milgram's '76 experiment to have people label landmarks in Paris
- Flanagan's '97 "Search referral Zeitgeist"
- Fortune's '01 Money Makes the World Go Round

Viégas & Wattenberg
interactions '08

Flickr Tag Cloud


Popular Tags on Flickr - Windows Internet Explorer
http://www.flickr.com/photos/tags/
Google
Links Marietta Traffic II 7450 GT Zimbra T-Square GMail Maps VAST09 KWJZ KPLU KUJI Susquehanna FoxSoccer Indian Hills
Popular Tags on Flickr
Page Tools

happymondayblues, seam, dodgerstadium, fwg, oneobject365daysproject, yip, nwn, spt, kodaktrix, yip2009, lifesaver

ncaatournament, whenuapai, arrahman, firstdayofspring, justafest, motheringsunday, happymothersday, noruz, crimesofpassion, twittmx2, clipsal500

Jump to: GO

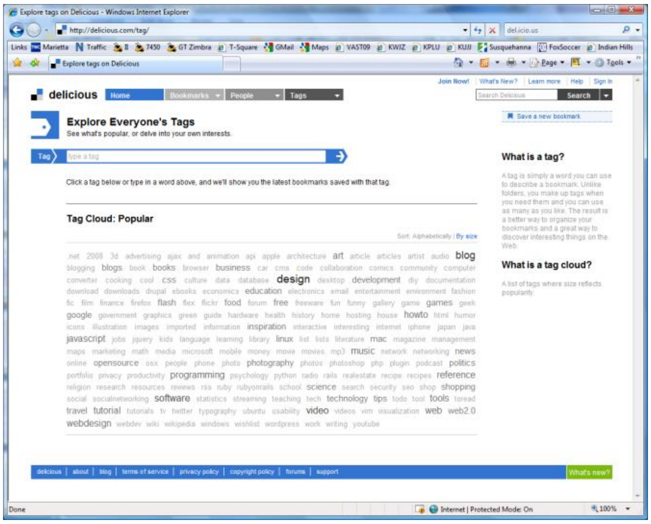
All time most popular tags

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

What are tags?

You can give your photos and videos a "tag", which is like a keyword or category label. Tags help you find photos and videos which have something in common. You can assign up to 75 tags to each photo or video.

delicious Tag Cloud


Explore tags on Delicious - Windows Internet Explorer
http://delicious.com/tag/
del.icio.us
Links Marietta Traffic II 7450 GT Zimbra T-Square GMail Maps VAST09 KWJZ KPLU KUJI Susquehanna FoxSoccer Indian Hills
Explore tags on Delicious
Page Tools
Join Now! What's New? Learn more Help Sign In
delicious
Home
Bookmarks
People
Tags
Search Delicious
Search
Save a new bookmark
Tag
type a tag

Explore Everyone's Tags

See what's popular, or delve into your own interests.

Click a tag below or type in a word above, and we'll show you the latest bookmarks saved with that tag.

Tag Cloud: Popular

Sort: Alphabetically | By size

.net 2008 3d advertising ajax and animation api apple architecture art article articles artist audio blog blogging blogs book books browser business car cms code collaboration comics community computer converter cooking cool css culture data database design desktop development diy documentation download downloads drupal ebooks economics education electronics email entertainment environment fashion fic film finance firefox flash flex flickr food forum free freeware fun funny gallery game games geek google government graphics green guide hardware health history home hosting house howto html humor icons illustration images imported information inspiration interactive interesting internet iphone japan java javascript jobs jquery kids language learning library linux list lists literature mac magazine management maps marketing math media microsoft mobile money movie movies mp3 music network networking news online opensource osx people phone photo photography photos photoshop php plugin podcast politics portfolio privacy productivity programming psychology python radio rails realestate recipe recipes reference religion research resources reviews rss ruby rubyonrails school science search security seo shop shopping social socialnetworking software statistics streaming teaching tech technology tips todo tool tools toread travel tutorial tutorials tv twitter typography ubuntu usability video videos vim visualization web web2.0 webdesign webdev wiki wikipedia windows wishlist wordpress work writing youtube

What is a tag?

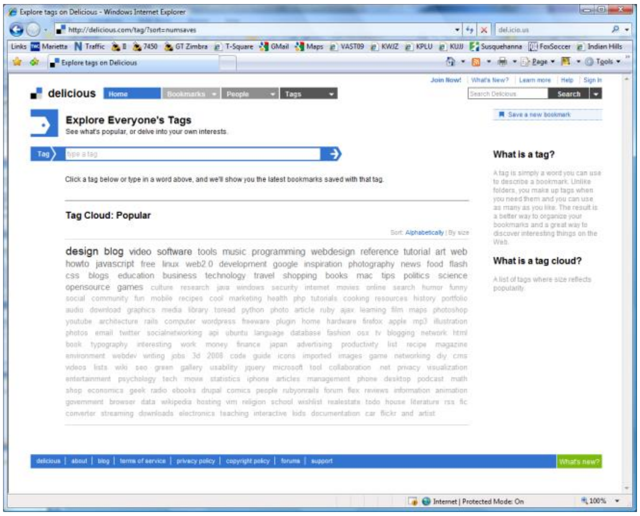
A tag is simply a word you can use to describe a bookmark. Unlike folders, you make up tags when you need them and you can use as many as you like. The result is a better way to organize your bookmarks and a great way to discover interesting things on the Web.

What is a tag cloud?

A list of tags where size reflects popularity.

delicious | about | blog | terms of service | privacy policy | copyright policy | forums | support
What's new?
Done
Internet | Protected Mode: On
100%

Alternate Order


Explore tags on Delicious - Windows Internet Explorer
http://delicious.com/tag/?sort=numsaves
del.icio.us
Links Marietta Traffic 7450 GT Zimbra T-Square GMail Maps VAST09 KWJZ KPLU KUJJ Susquehanna FoxSoccer Indian Hills
Explore tags on Delicious
Page Tools
Join Now! | What's New? | Learn more | Help | Sign In
delicious
Home
Bookmarks
People
Tags
Search Delicious
Search
Save a new bookmark
Tag
type a tag
delicious | about | blog | terms of service | privacy policy | copyright policy | forums | support
What's new?
Internet | Protected Mode: On
100%

Explore Everyone's Tags

See what's popular, or delve into your own interests.

Click a tag below or type in a word above, and we'll show you the latest bookmarks saved with that tag.

Tag Cloud: Popular

Sort: Alphabetically | By size

design blog video software tools music programming webdesign reference tutorial art web howto javascript free linux web2.0 development google inspiration photography news food flash css blogs education business technology travel shopping books mac tips politics science opensource games culture research java windows security internet movies online search humor funny social community fun mobile recipes cool marketing health php tutorials cooking resources history portfolio audio download graphics media library toread python photo article ruby ajax learning film maps photoshop youtube architecture rails computer wordpress freeware plugin home hardware firefox apple mp3 illustration photos email twitter socialnetworking api ubuntu language database fashion osx tv blogging network html book typography interesting work money finance japan advertising productivity list recipe magazine environment webdev writing jobs 3d 2008 code guide icons imported images game networking diy cms videos lists wiki seo green gallery usability jquery microsoft tool collaboration .net privacy visualization entertainment psychology tech movie statistics iphone articles management phone desktop podcast math shop economics geek radio ebooks drupal comics people rubyonrails forum flex reviews information animation government browser data wikipedia hosting vim religion school wishlist realestate todo house literature rss fic converter streaming downloads electronics teaching interactive kids documentation car flickr and artist

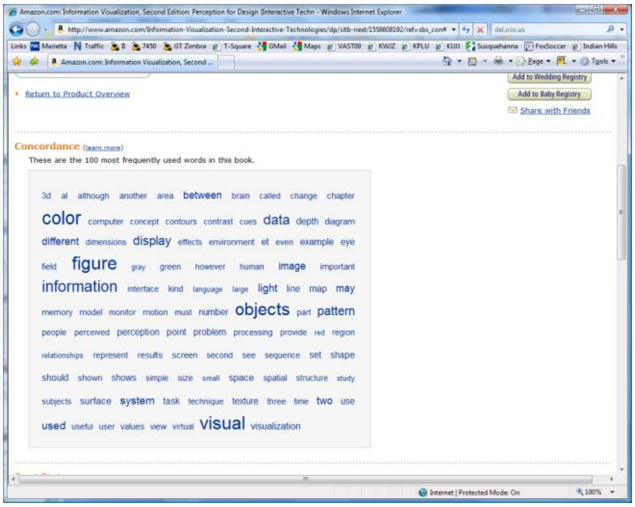
What is a tag?

A tag is simply a word you can use to describe a bookmark. Unlike folders, you make up tags when you need them and you can use as many as you like. The result is a better way to organize your bookmarks and a great way to discover interesting things on the Web.

What is a tag cloud?

A list of tags where size reflects popularity.

Amazon’s (old) Product Concordance


Amazon.com: Information Visualization, Second Edition: Perception for Design (Interactive Techn - Windows Internet Explorer
http://www.amazon.com/Information-Visualization-Second-Interactive-Technologies/dp/sitb-next/1558608192/ref=sbx_con#
del.icio.us
Links Marietta Traffic 7450 GT Zimbra T-Square GMail Maps VAST09 KWJZ KPLU KUJJ Susquehanna FoxSoccer Indian Hills
Amazon.com: Information Visualization, Second ...
Page Tools
Add to Wedding Registry
Add to Baby Registry
Share with Friends
Return to Product Overview
Internet | Protected Mode: On
100%

Concordance (learn more)

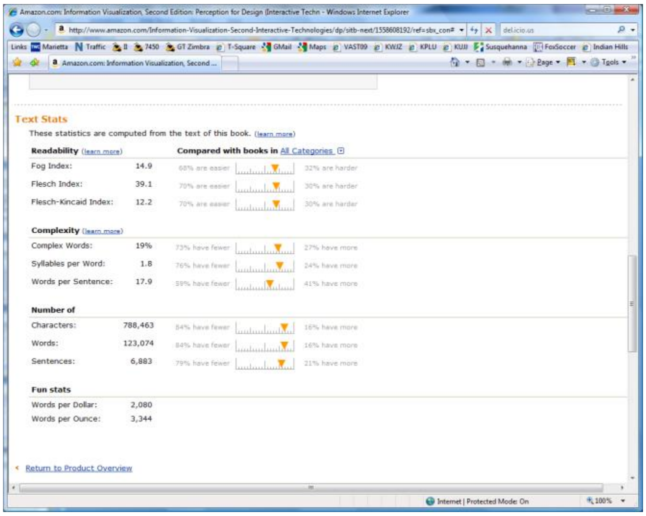
These are the 100 most frequently used words in this book.

3d al although another area between brain called change chapter color computer concept contours contrast cues data depth diagram different dimensions display effects environment et even example eye field figure gray green however human image important information interface kind language large light line map may memory model monitor motion must number objects part pattern people perceived perception point problem processing provide red region relationships represent results screen second see sequence set shape should shown shows simple size small space spatial structure study subjects surface system task technique texture three time two use used useful user values view virtual visual visualization

Maybe now a “word cloud”

More (old) Info

There are other types of info about a document on Amazon


Amazon.com: Information Visualization, Second Edition: Perception for Design (Interactive Techn - Windows Internet Explorer
http://www.amazon.com/Information-Visualization-Second-Interactive-Technologies/dp/sitb-next/1558608192/ref=sbx_con#
del.icio.us
Links Marietta Traffic 7450 GT Zimbra T-Square GMail Maps VAST09 KWJZ KPLU KUJI Susquehanna FoxSoccer Indian Hills
Amazon.com: Information Visualization, Second ...
Page Tools
Internet | Protected Mode: On
100%

Text Stats

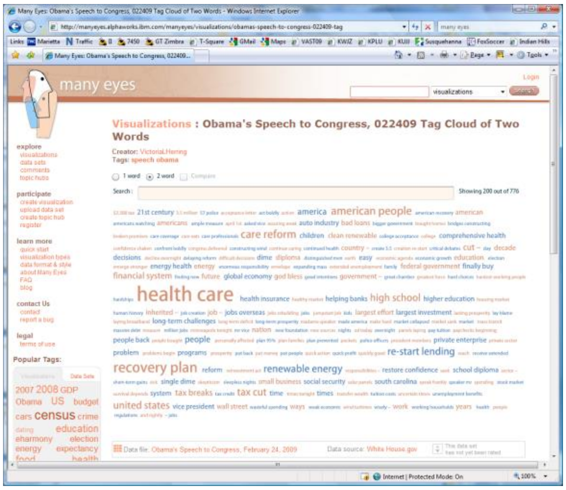
These statistics are computed from the text of this book. (learn more)

Readability (learn more)		Compared with books in All Categories	
Fog Index:	14.9	68% are easier	32% are harder
Flesch Index:	39.1	70% are easier	30% are harder
Flesch-Kincaid Index:	12.2	70% are easier	30% are harder
Complexity (learn more)			
Complex Words:	19%	73% have fewer	27% have more
Syllables per Word:	1.8	76% have fewer	24% have more
Words per Sentence:	17.9	59% have fewer	41% have more
Number of			
Characters:	788,463	84% have fewer	16% have more
Words:	123,074	84% have fewer	16% have more
Sentences:	6,883	79% have fewer	21% have more
Fun stats			
Words per Dollar:	2,080		
Words per Ounce:	3,344		

‹ Return to Product Overview

Many Eyes Tag Cloud

Here, pairs of words are shown



Visualizations : Obama's Speech to Congress, 022409 Tag Cloud of Two Words

Creator: VictoriaLHerring
Tags: speech obama

○ 1 word ● 2 word □ Compare

Search:

Showing 200 out of 776

Data file: Obama's Speech to Congress, February 24, 2009

Data source: White House.gov

This data set has not yet been rated

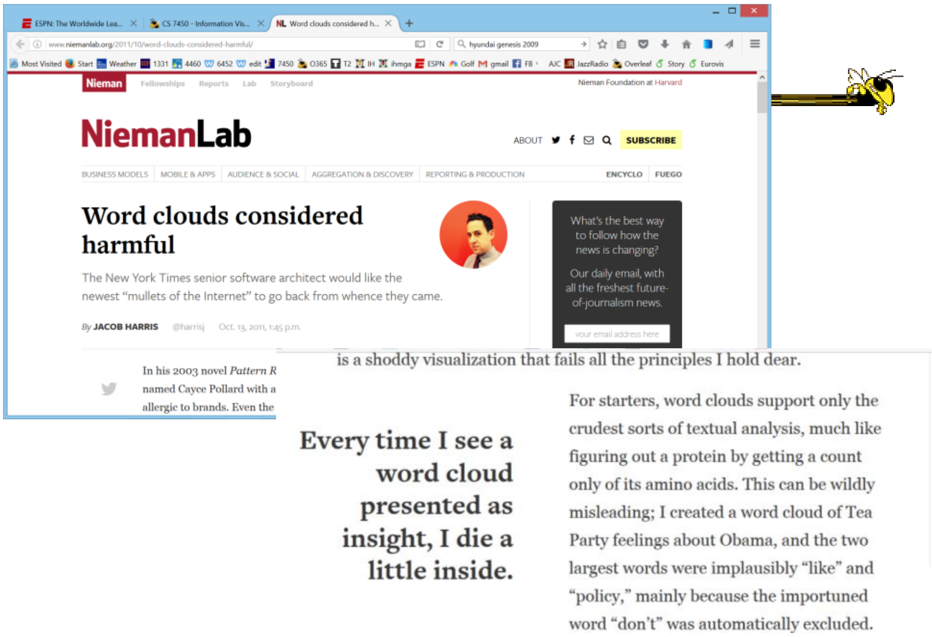
Problems

- 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

- Studies have even shown they underperform

Gruen et al
CHI '06

http://www.niemanlab.org/2011/10/word-clouds-considered-harmful/


ESPN: The Worldwide Lea...
CS 7450 - Information Vis...
NL Word clouds considered h...
www.niemanlab.org/2011/10/word-clouds-considered-harmful/
hyundai genesis 2009
Most Visited Start Weather 1331 4460 6452 edit 7450 O365 T2 IH ihmga ESPN Golf gmail FB AJC JazzRadio Overleaf Story Eurovis
Nieman Fellowships Reports Lab Storyboard
Nieman Foundation at Harvard
NiemanLab
ABOUT
SUBSCRIBE
BUSINESS MODELS MOBILE & APPS AUDIENCE & SOCIAL AGGREGATION & DISCOVERY REPORTING & PRODUCTION ENCYCLO FUEGO
your email address here

Word clouds considered harmful

The New York Times senior software architect would like the newest "mullets of the Internet" to go back from whence they came.

By JACOB HARRIS @harrisj Oct. 13, 2011, 1:45 p.m.

What's the best way to follow how the news is changing?

Our daily email, with all the freshest future-of-journalism news.

In his 2003 novel *Pattern R*
named Cayce Pollard with a
allergic to brands. Even the

is a shoddy visualization that fails all the principles I hold dear.

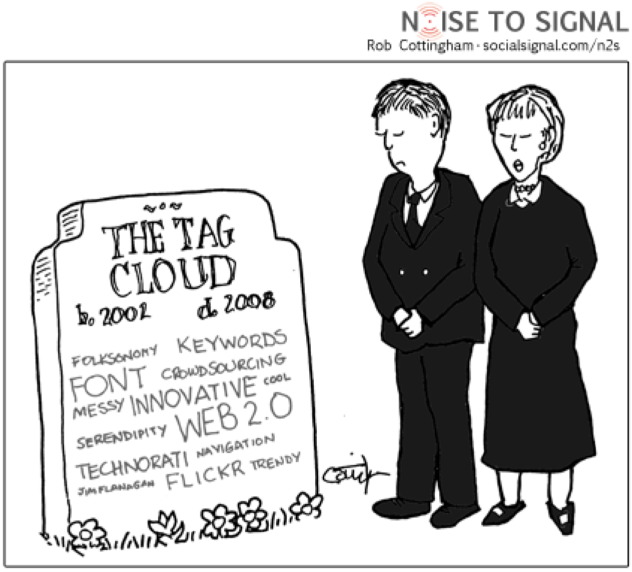
Every time I see a word cloud presented as insight, I die a little inside.

For starters, word clouds support only the crudest sorts of textual analysis, much like figuring out a protein by getting a count only of its amino acids. This can be wildly misleading; I created a word cloud of Tea Party feelings about Obama, and the two largest words were implausibly "like" and "policy," mainly because the importuned word "don't" was automatically excluded.

Why So 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
- Fun, not business-like

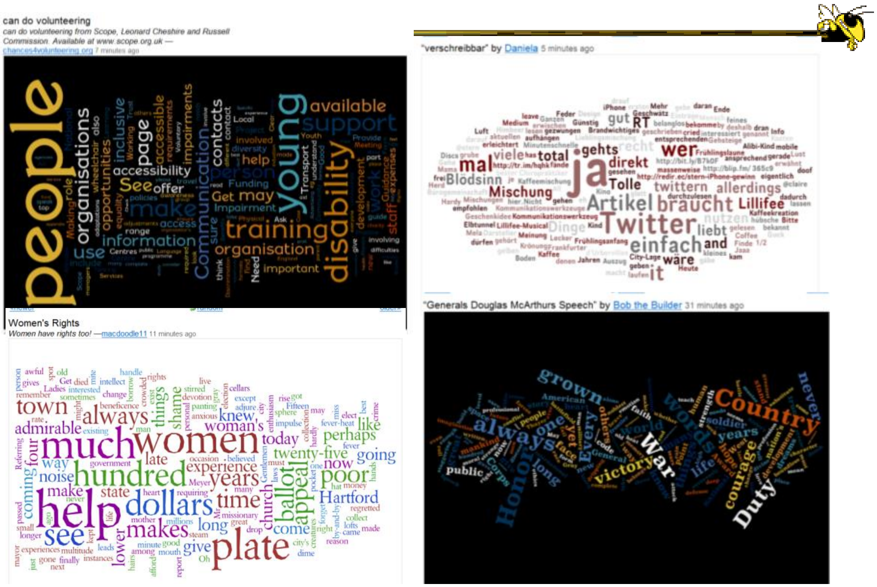
Hearst & Rosner
HICSS '08



http://www.socialsignal.com/system/files/images/2008-08-01-tagcloud.gif

Wordle

http://www.wordle.net



can do volunteering

can do volunteering from Scope, Leonard Cheshire and Russell Commission. Available at www.scope.org.uk — chances4volunteering.org 7 minutes ago

"verschreibbar" by Daniela 5 minutes ago

Women's Rights

*Women have rights too! —*macdoodle11 11 minutes ago

"Generals Douglas McArthurs Speech" by Bob the Builder 31 minutes ago

Wordle

- Tightly packed words, sometimes vertical or diagonal
- Word size is linearly correlated with frequency (typically square root in cloud)
- Multiple color palettes
- User gets some control

Viegas, Wattenberg, & Feinberg
TVCG (InfoVis) '09

Layout Algorithm

- Details not published
- Idea:

  -
    ```
    sort words by weight, decreasing order
    for each word w
      w.position := makeInitialPosition(w);
      while w intersects other words:
          updatePosition(w);
    ```
  - Init position randomly chosen according to distribution for target shape
  - Update position moves out radially

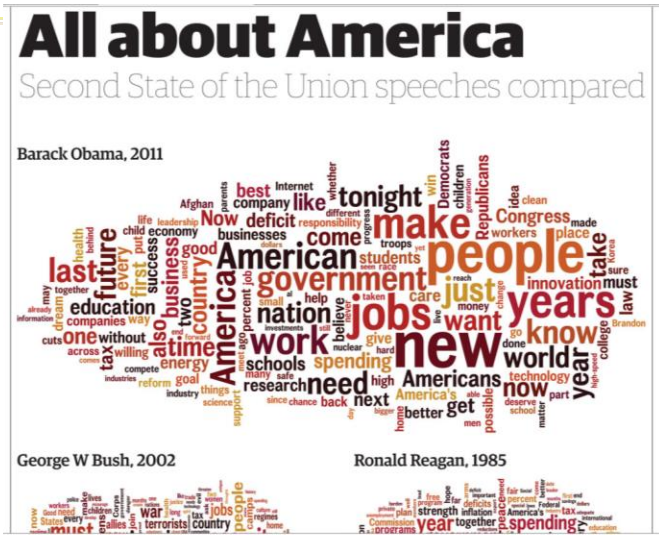
Fun Uses

- Political speeches
- Songs and poems
- Love letters (for “boyfriend points”)
- Wedding vows
- Course syllabi
- Teaching writing
- Gifts

2-day Survey in Jan. 09

- 2/3 respondents were women
- Interest came from design, visual appeal, beauty
- Why preferred over word clouds:
  - Emotional impact
  - Attention-keeping visuals
  - Organic, non-linear
- Fair percentage didn't know what size signified

SoTU Wordles



All about America

Second State of the Union speeches compared

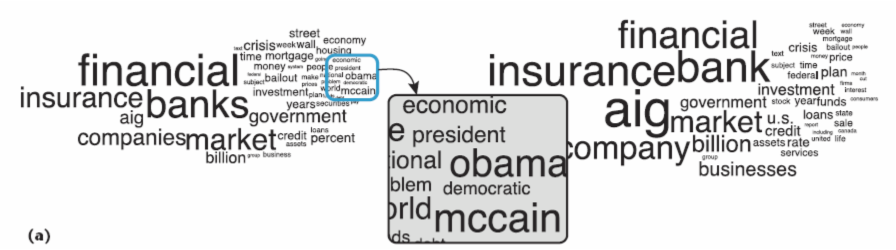
Barack Obama, 2011

George W Bush, 2002

Ronald Reagan, 1985

http://www.guardian.co.uk/news/datablog/2011/jan/25/state-of-the-union-text-obama#

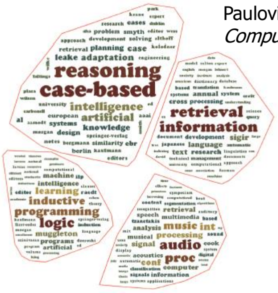
A Little More Order



Order the words more by frequency

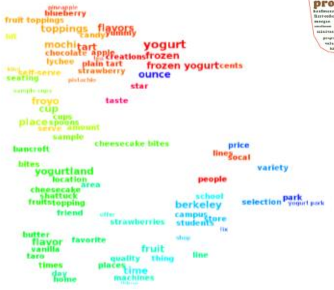
Cui et al
IEEE CG&A ‘10

Semantic/Context Word Clouds

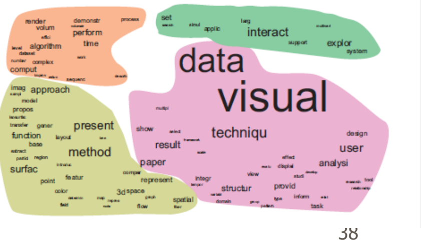


Paulovich et al
Computer Graphics Forum '12

Wang et al
Graphics Interface '14



Wu et al
Computer Graphics Forum '11



Wordle Characteristics

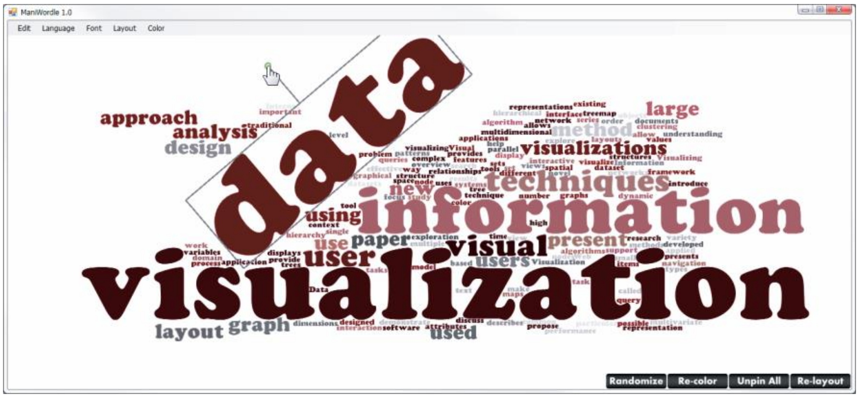
- Layout, words are automatic
- If you had some control, what would you like to change or alter?

Mani-Wordle

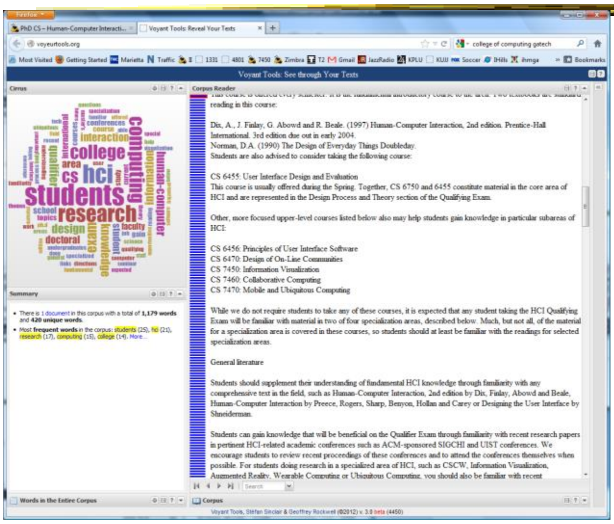
- Start with nice default algorithm
- Give user more control over design
  - Alter color (within a palette)
  - Pin words, redo the rest
  - Move and rotate words
  - Smooth animation and collision detection for tracking changes

Koh et al
TVCG (InfoVis) '10

Video



Text Analysis on Web


PhD CS – Human-Computer Interacti...
Voyant Tools: Reveal Your Texts
voyeurtools.org
college of computing gatech
Voyant Tools: See through Your Texts
Cirrus
Corpus Reader
Summary
Words in the Entire Corpus
Corpus
Search
students
research
hci
college
computing
human-computer
design
doctoral
knowledge

reading in this course:

Dix, A., J. Finlay, G. Abowd and R. Beale. (1997) Human-Computer Interaction, 2nd edition. Prentice-Hall International. 3rd edition due out in early 2004.
Norman, D.A. (1990) The Design of Everyday Things Doubleday.
Students are also advised to consider taking the following course:

CS 6455: User Interface Design and Evaluation
This course is usually offered during the Spring. Together, CS 6750 and 6455 constitute material in the core area of HCI and are represented in the Design Process and Theory section of the Qualifying Exam.

Other, more focused upper-level courses listed below also may help students gain knowledge in particular subareas of HCI:

CS 6456: Principles of User Interface Software
CS 6470: Design of On-Line Communities
CS 7450: Information Visualization
CS 7460: Collaborative Computing
CS 7470: Mobile and Ubiquitous Computing

While we do not require students to take any of these courses, it is expected that any student taking the HCI Qualifying Exam will be familiar with material in two of four specialization areas, described below. Much, but not all, of the material for a specialization area is covered in these courses, so students should at least be familiar with the readings for selected specialization areas.

General literature

Students should supplement their understanding of fundamental HCI knowledge through familiarity with any comprehensive text in the field, such as Human-Computer Interaction, 2nd edition by Dix, Finlay, Abowd and Beale, Human-Computer Interaction by Preece, Rogers, Sharp, Benyon, Hollan and Carey or Designing the User Interface by Shneiderman.

Students can gain knowledge that will be beneficial on the Qualifier Exam through familiarity with recent research papers in pertinent HCI-related academic conferences such as ACM-sponsored SIGCHI and UIST conferences. We encourage students to review recent proceedings of these conferences and to attend the conferences themselves when possible. For students doing research in a specialized area of HCI, such as CSCW, Information Visualization, Augmented Reality, Wearable Computing or Ubiquitous Computing, you should also be familiar with recent

- There is 1 document in this corpus with a total of **1,179 words** and **420 unique words**.
- Most **frequent words** in the corpus: students (25), hci (21), research (17), computing (15), college (14). More...

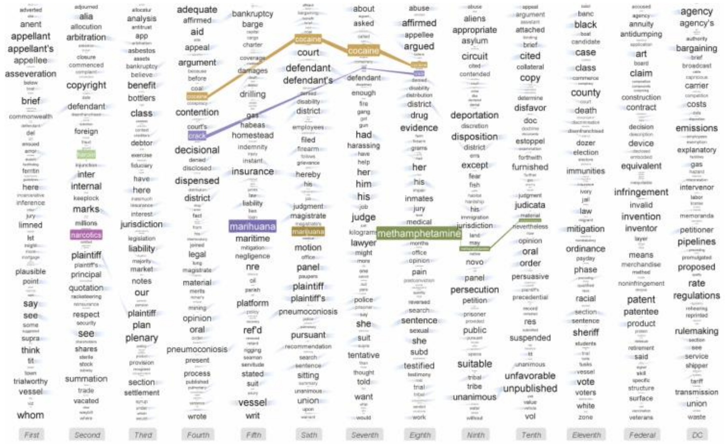
Voyant Tools, Stéfan Sinclair & Geoffrey Rockwell (©2012) v. 3.0 beta (4450)

http://voyant-tools.org/

Multiple Documents?

- How to show word frequencies across multiple related documents?

Parallel Tag Clouds



Video

Collins et al
VAST '09

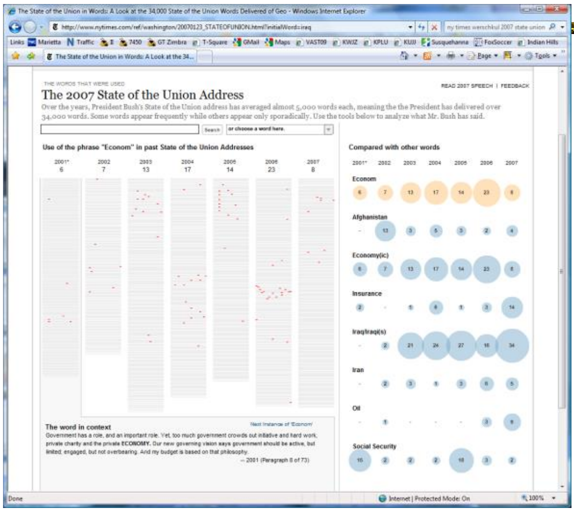
Different circuit courts

Analytic Support

- Note: Word Clouds and Wordles are really more overview-style visualizations
  - Don’t really support queries, searches, drill-down

- How might we also support queries and search?

Overview & Timeline



THE WORDS THAT WERE USED

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 tools below to analyze what Mr. Bush has said.

The word in context

Government has a role, and an important role. Yet, too much government crowds out initiative and hard work, private charity and the private ECONOMY. Our new governing vision says government should be active, but limited; engaged, but not overbearing. And my budget is based on that philosophy.

-- 2001 (Paragraph 8 of 73)

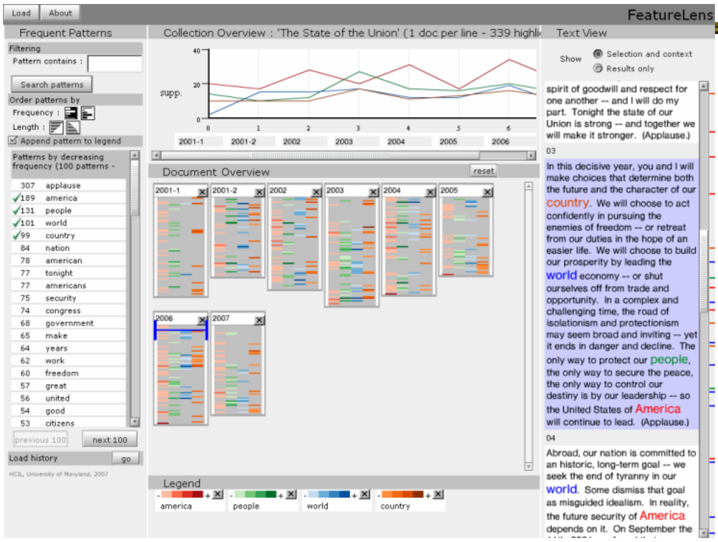
The State of the Union in Words: A Look at the 34,000 State of the Union Words Delivered of Geo - Windows Internet Explorer
http://www.nytimes.com/ref/washington/20070123_STATEOFUNION.html?initialWord=iraq
READ 2007 SPEECH | FEEDBACK
Search
or choose a word here.
Use of the phrase "Econom" in past State of the Union Addresses
2001* 6
2002 7
2003 13
2004 17
2005 14
2006 23
2007 8
Compared with other words
Econom
Afghanistan
Economy(ic)
Insurance
Iraq/Iraqi(s)
Iran
Oil
Social Security
Next Instance of 'Econom'
Done
Internet | Protected Mode: On
100%

State of the Union Addresses

http://www.nytimes.com/ref/washington/20070123_STATEOFUNION.html?initialWord=iraq

FeatureLens

Video



Patterns by decreasing frequency (100 patterns -

	307	applause
✓	189	america
✓	131	people
✓	101	world
✓	99	country
	84	nation
	78	american
	77	tonight
	77	americans
	75	security
	74	congress
	68	government
	65	make
	64	years
	62	work
	60	freedom
	57	great
	56	united
	54	good
	53	citizens

spirit of goodwill and respect for one another -- and I will do my part. Tonight the state of our Union is strong -- and together we will make it stronger. (Applause.)

03

In this decisive year, you and I will make choices that determine both the future and the character of our country. We will choose to act confidently in pursuing the enemies of freedom -- or retreat from our duties in the hope of an easier life. We will choose to build our prosperity by leading the world economy -- or shut ourselves off from trade and opportunity. In a complex and challenging time, the road of isolationism and protectionism may seem broad and inviting -- yet it ends in danger and decline. The only way to protect our people, the only way to secure the peace, the only way to control our destiny is by our leadership -- so the United States of America will continue to lead. (Applause.)

04

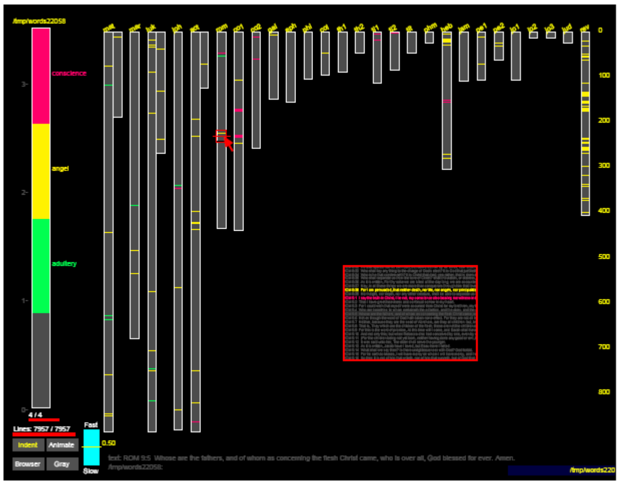
Abroad, our nation is committed to an historic, long-term goal -- we seek the end of tyranny in our world. Some dismiss that goal as misguided idealism. In reality, the future security of America depends on it. On September the

Show patterns of words or n-grams

Don et al
CIKM '07

http://www.cs.umd.edu/hcil/textvis/featurelens/

SeeSoft Display



text: ROM 9:5 Whose are the fathers, and of whom as concerning the flesh Christ came, who is over all, God blessed for ever. Amen.
/tmp/words22058:

Like taping text to the wall and walking far away

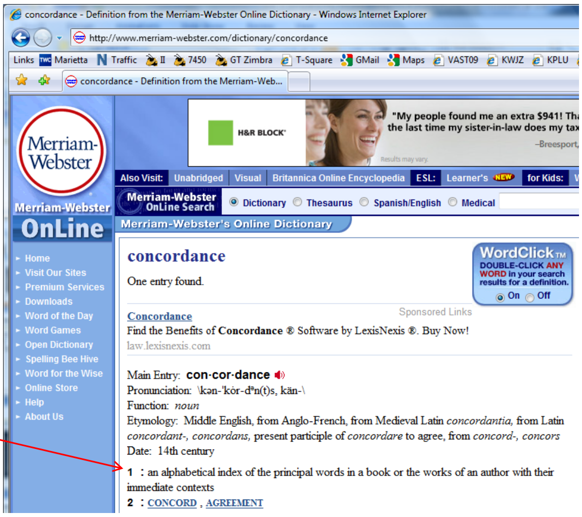
New Testament

Eick
Journal Comput. & Graph. Stats '94

Beyond Individual Words

- The previous techniques focus largely on words
  - Especially word clouds & wordles
- Can we show combinations of words, ie, actual phrases and sentences, in order to provide more context?

Concordance


concordance - Definition from the Merriam-Webster Online Dictionary - Windows Internet Explorer
http://www.merriam-webster.com/dictionary/concordance
Links Marietta Traffic II 7450 GT Zimbra T-Square GMail Maps VAST09 KWJZ KPLU
concordance - Definition from the Merriam-Web...
Merriam-Webster
H&R BLOCK
"My people found me an extra $941! Tha
the last time my sister-in-law does my tax
–Breesport,
Results may vary.
Also Visit: Unabridged Visual Britannica Online Encyclopedia ESL: Learner's NEW for Kids:
Merriam-Webster OnLine Search
● Dictionary ○ Thesaurus ○ Spanish/English ○ Medical
Merriam-Webster's Online Dictionary
Merriam-Webster OnLine
Home
Visit Our Sites
Premium Services
Downloads
Word of the Day
Word Games
Open Dictionary
Spelling Bee Hive
Word for the Wise
Online Store
Help
About Us
WordClick TM
DOUBLE-CLICK ANY WORD in your search results for a definition.
● On ○ Off
Sponsored Links

concordance

One entry found.

Concordance

Find the Benefits of **Concordance** ® Software by LexisNexis ®. Buy Now!

law.lexisnexis.com

Main Entry: **con·cor·dance**

Pronunciation: \kən-ˈkȯr-dᵊn(t)s, kän-\

Function: *noun*

Etymology: Middle English, from Anglo-French, from Medieval Latin *concordantia,* from Latin *concordant-, concordans,* present participle of *concordare* to agree, from *concord-, concors*

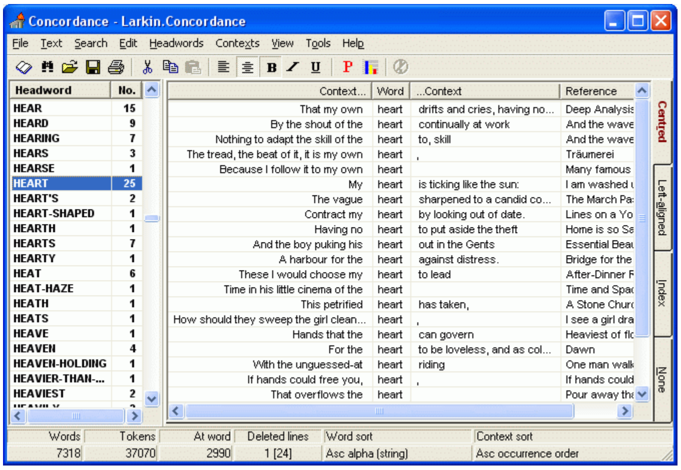
Date: 14th century

1 : an alphabetical index of the principal words in a book or the works of an author with their immediate contexts

2 : CONCORD , AGREEMENT

Definition

Concordance in Text


Concordance - Larkin.Concordance
File Text Search Edit Headwords Contexts View Tools Help
Centred
Left-aligned
Index
None

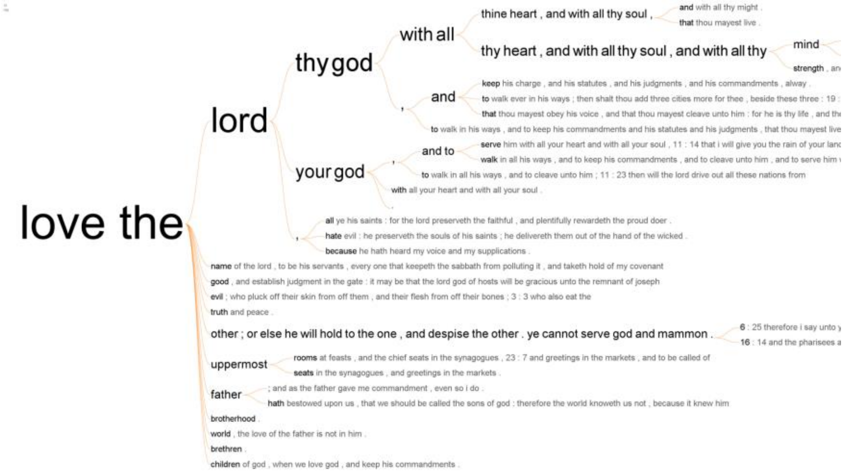
Headword	No.
HEAR	15
HEARD	9
HEARING	7
HEARS	3
HEARSE	1
HEART	25
HEART'S	2
HEART-SHAPED	1
HEARTH	1
HEARTS	7
HEARTY	1
HEAT	6
HEAT-HAZE	1
HEATH	1
HEATS	1
HEAVE	1
HEAVEN	4
HEAVEN-HOLDING	1
HEAVIER-THAN-...	1
HEAVIEST	2

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 the

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

http://www.concordancesoftware.co.uk

Word Tree


love the
lord
thy god
with all
thine heart , and with all thy soul ,
and with all thy might .
that thou mayest live .
thy heart , and with all thy soul , and with all thy
mind
strength , an
,
and
keep his charge , and his statutes , and his judgments , and his commandments , alway .
to walk ever in his ways ; then shalt thou add three cities more for thee , beside these three : 19 :
that thou mayest obey his voice , and that thou mayest cleave unto him : for he is thy life , and th
to walk in his ways , and to keep his commandments and his statutes and his judgments , that thou mayest live
your god
,
and to
serve him with all your heart and with all your soul , 11 : 14 that i will give you the rain of your land
walk in all his ways , and to keep his commandments , and to cleave unto him , and to serve him
to walk in all his ways , and to cleave unto him ; 11 : 23 then will the lord drive out all these nations from
with all your heart and with all your soul .
,
all ye his saints : for the lord preserveth the faithful , and plentifully rewardeth the proud doer .
hate evil : he preserveth the souls of his saints ; he delivereth them out of the hand of the wicked .
because he hath heard my voice and my supplications .
name of the lord , to be his servants , every one that keepeth the sabbath from polluting it , and taketh hold of my covenant
good , and establish judgment in the gate : it may be that the lord god of hosts will be gracious unto the remnant of joseph
evil ; who pluck off their skin from off them , and their flesh from off their bones ; 3 : 3 who also eat the
truth and peace .
other ; or else he will hold to the one , and despise the other . ye cannot serve god and mammon .
6 : 25 therefore i say unto y
16 : 14 and the pharisees a
uppermost
rooms at feasts , and the chief seats in the synagogues , 23 : 7 and greetings in the markets , and to be called of
seats in the synagogues , and greetings in the markets .
father
; and as the father gave me commandment , even so i do .
hath bestowed upon us , that we should be called the sons of god : therefore the world knoweth us not , because it knew him
brotherhood .
world , the love of the father is not in him .
brethren .
children of god , when we love god , and keep his commandments .

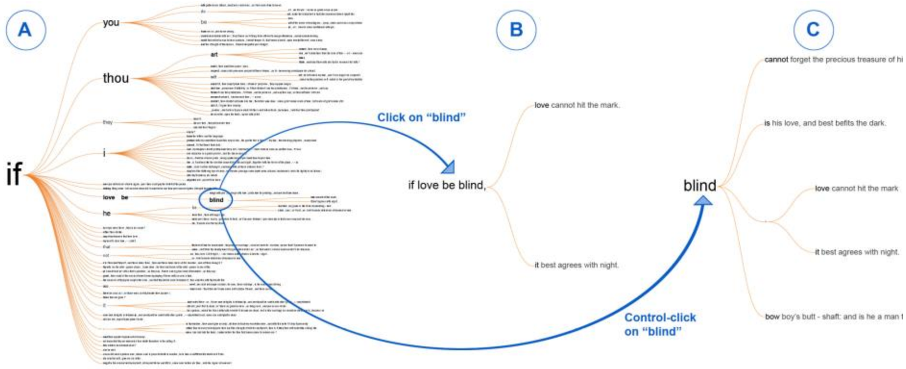
From King James Bible

Word Tree

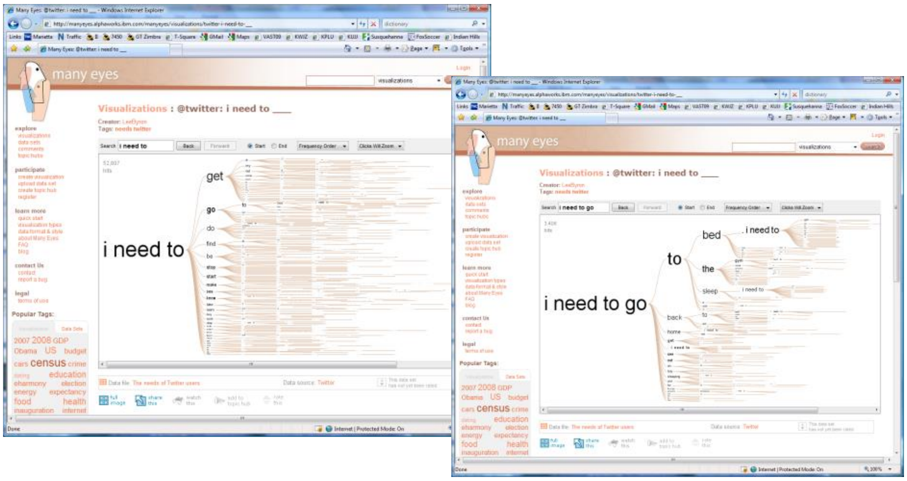
- Shows context of a word or words
  - Follow word with all the phrases that follow it
- Font size shows frequency of appearance
- Continue branch until hitting unique phrase
- Clicking on phrase makes it the focus
- Ordered alphabetically, by frequency, or by first appearance

Wattenberg & Viégas
TVCG (InfoVis) '08

Interaction


A
you
thou
if
love be
blind
B
Click on "blind"
love cannot hit the mark.
if love be blind,
it best agrees with night.
C
blind
cannot forget the precious treasure of hi
is his love, and best befits the dark.
love cannot hit the mark
it best agrees with night.
bow boy's butt - shaft: and is he a man t
Control-click on "blind"

Many Eyes' WordTree



Was added to Many Eyes

Phrase Nets

- Examine unstructured [illegible] documents
- Presents pairs of terms [illegible] 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

van Ham et al
TVCG (InfoVis) '09

Examples

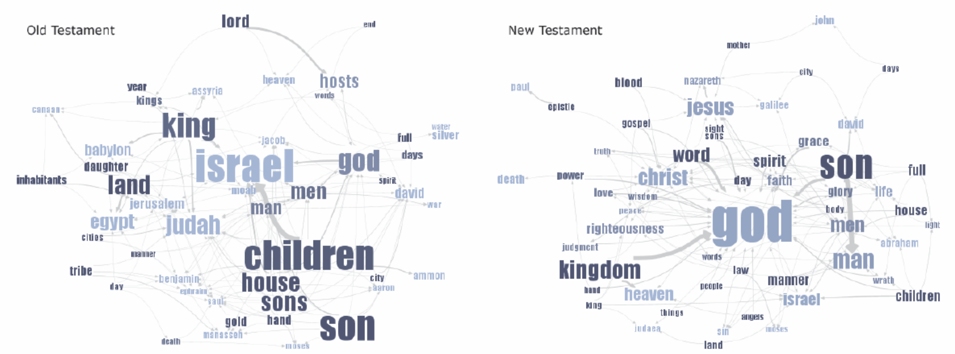


Fig 4. Matching the same pattern on different texts. Here we used the pattern “X of Y” to compare the old and new testaments. Israel takes a central place in the Old Testament, while God acts as the main pattern receiver in the New Testament.

Examples

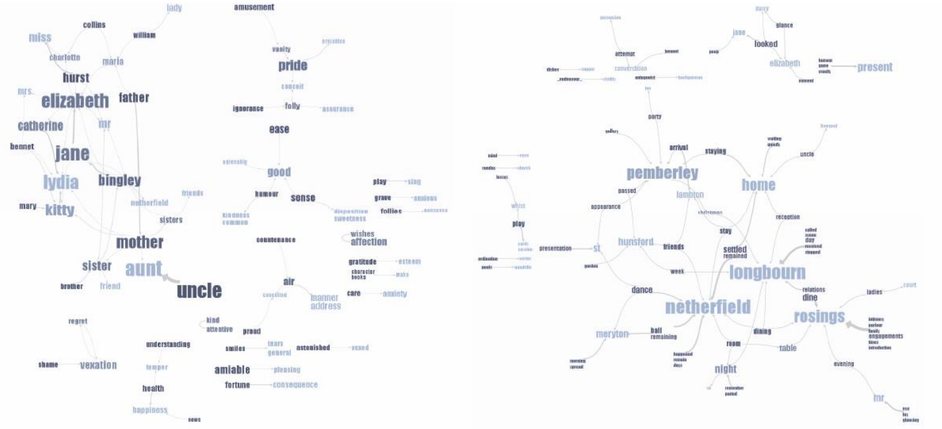
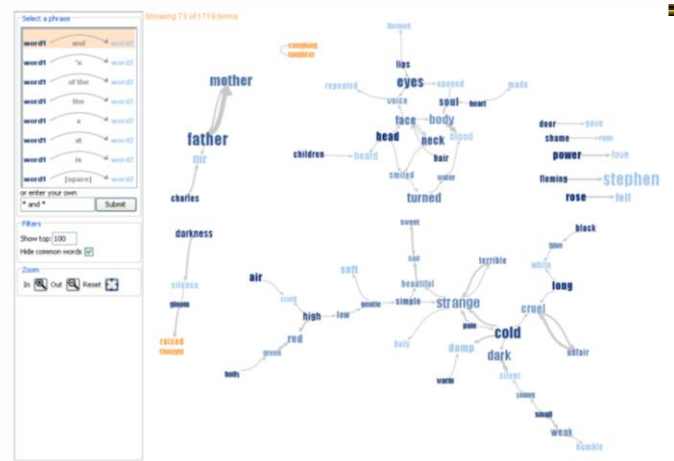


Fig 5. Matching different patterns on the same text. Here we analyzed Jane Austen’s *Pride and Prejudice* with “X and Y” and “X at Y” respectively. The left image shows relationships between the main characters amongst others, while the right image shows relationships between locations.

User Interface



or enter your own: * and *

Show top: 100

Hide common words ☑

Fig 3. The Phrase Net user interface applied to James Joyce's Portrait of the Artist as a Young Man. The user can select a predefined pattern from the list of patterns on the left or define a custom pattern in the box below. This list of patterns simultaneously serves as a legend, a list of presets and an interactive training mechanism for regular expressions. Here the user has selected "...X and Y...", revealing two main clusters, one almost exclusively consisting of adjectives, the other of verbs and nouns. The highlighted clusters of terms have been aggregated by our edge compression algorithm.

Words and Context

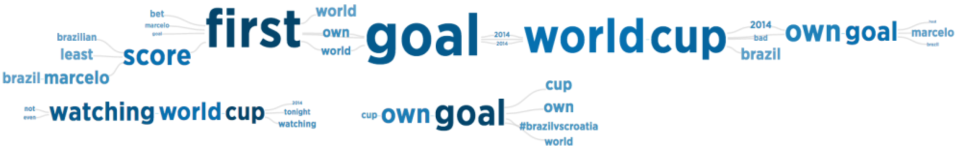
- Can we show most frequent words like a word cloud but also provide context?
  - Should each word appear one time?
  - But then how to show context?
  - If appears multiple times, how to make that work?

SentenTree

- Elements of word clouds and word trees
  - Highlight keywords using size
  - Show sentence fragments
  - Provide a summary of the dataset
  - Enable drill-down into details

Hu, Wongsuphasawat, and Stasko
TVCG '17 (InfoVis '16)

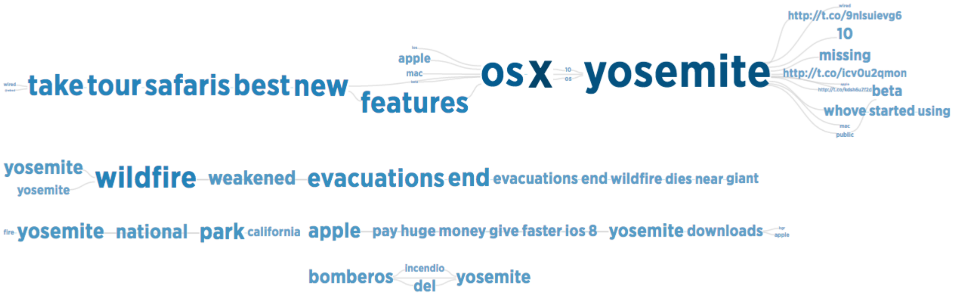
Example



Summary of 189,450 tweets (108,702 unique) posted in a 15 minute time window around the first goal of the opening game of the 2014 Soccer World Cup

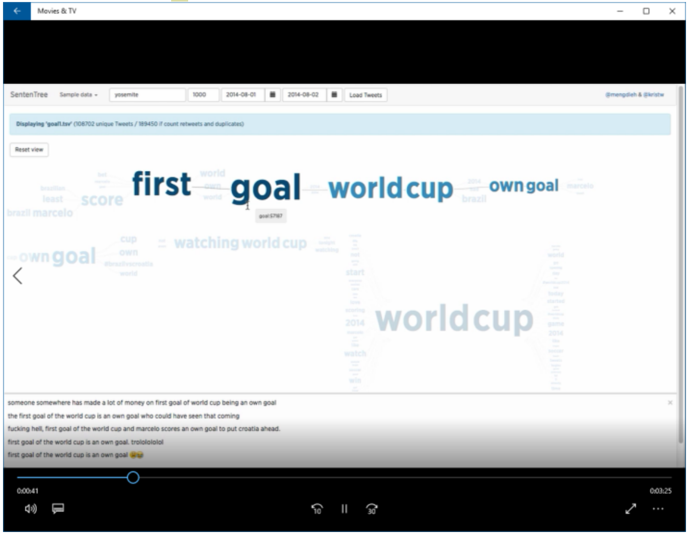
(Interaction is key & not shown here)

Example



Tweets mentioning word "Yosemite" from Aug 1, 2014

Video



Displaying 'goal1.tsv' (108702 unique Tweets / 189450 if count retweets and duplicates)

someone somewhere has made a lot of money on first goal of world cup being an own goal

the first goal of the world cup is an own goal who could have seen that coming

fucking hell, first goal of the world cup and marcelo scores an own goal to put croatia ahead.

first goal of the world cup is an own goal. trololololol

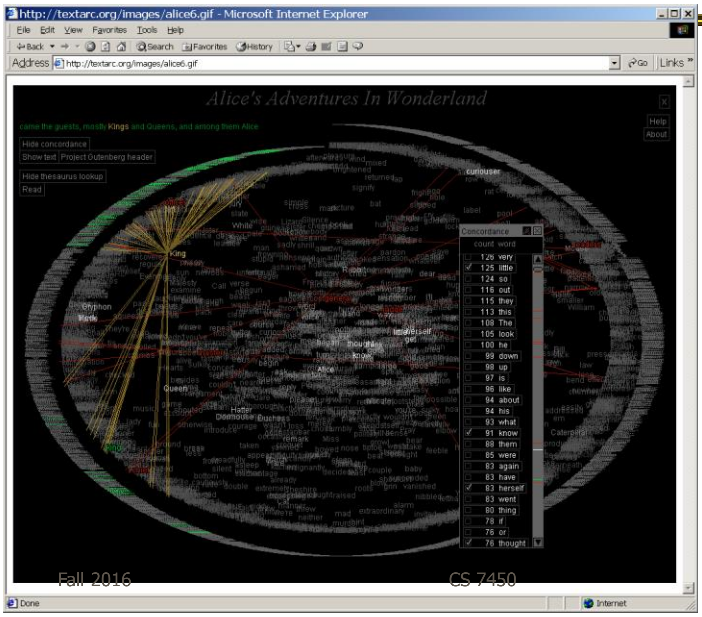
first goal of the world cup is an own goal 😂😂

Another Challenge

- Visualize an entire book
- What does that mean?
  - Word appearances
  - Sentences
  - …

TextArc

http://textarc.org


http://textarc.org/images/alice6.gif - Microsoft Internet Explorer

Address http://textarc.org/images/alice6.gif

Alice's Adventures In Wonderland

came the guests, mostly Kings and Queens, and among them Alice

Concordance

	count	word
□	126	very
✓	125	little
□	124	so
□	116	out
□	115	they
□	113	this
□	108	The
□	105	look
□	100	he
□	99	down
□	98	up
□	97	is
□	96	like
□	94	about
□	94	his
□	93	what
✓	91	know
□	88	them
□	85	were
□	83	again
□	83	have
✓	83	herself
□	83	went
□	80	thing
□	78	if
□	76	or
✓	76	thought

Sentences laid out in order of appearance

Words near to where they appear

Significant interaction

Brad Paley

Next Time

- More about collections of documents and showing other characteristics of documents
  - Analysis metrics
  - Entities
  - Concepts & themes

Learning Objectives

- Explain key challenges in visualizing a large document or body of text
- Identify and explain different techniques for representing words and concepts in a document
  - Word cloud, Wordle, Parallel tag cloud, SeeSoft, WordTree, PhraseNet, SentenTree, TextArc
- Understand the positives and limitations of word clouds and Wordles
- Describe SeeSoft-style miniature visual representations
- Explain what word concordance is
- Describe how WordTree representation works
- Identify and explain the techniques:
  - Word cloud, Wordle, Parallel tag cloud, SeeSoft, WordTree, PhraseNet, SentenTree, TextArc

Reading

- Viegas & Wattenberg '08

Project

- Meetings with TAs
  - Once every two weeks
  - Will be logged
  - Conveys an impression

Upcoming

- Text and Documents 2
- Hierarchical (Tree) Data 1

References

- Marti Hearst's i247 slides
- All referred to papers