Topic Notes

Text and Document Visualization 1

CS 7450 - Information Visualization
March 15, 2011
John Stasko

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?

Tasks/Goals

- What kinds of analysis questions might a person ask about text & documents?

Example Tasks & Goals

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

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

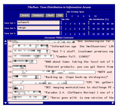
Related Topic - Sensemaking

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

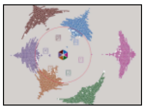
Challenge

- 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

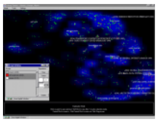
This Week's Agenda



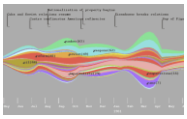
Visualization for IR
Helping search



Visualizing text
Showing words, phrases, and sentences



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



Information Retrieval

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

Improving Text Searches

- What's wrong with the common search?
- Visualizing the results of search operations is another big area in text infovis

What Hearst Thinks is Wrong

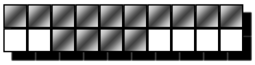
- Query responses do not include include:
  - How strong the match is
  - How frequent each term is
  - How each term is distributed in the document
  - Overlap between terms
  - Length of document
- Document ranking is opaque
- Inability to compare between results
- Input limits term relationships

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

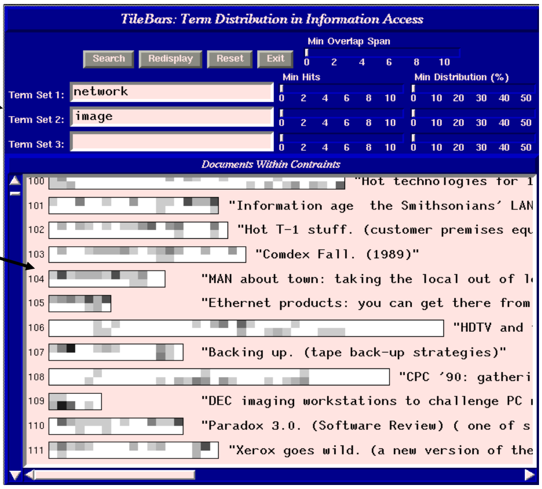


- Graphical representation of term distribution and overlap
- Simultaneously indicate:
  - Relative document length
  - Frequency of term sets in document
  - Distribution of term sets with respect to the document and each other

Interface

Search terms

Presentation



TileBars: Term Distribution in Information Access

Term Set 1: network

Term Set 2: image

Term Set 3:

Documents Within Contraints

100 "Hot technologies for I

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

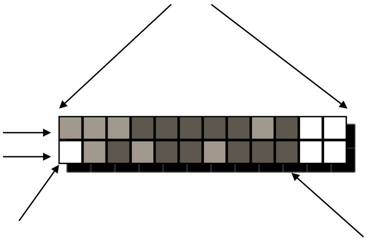
109 "DEC imaging workstations to challenge PC

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

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

Technique

Video



Relative length of document

Two search terms

Blocks indicate "chunks" of text, such as paragraphs

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

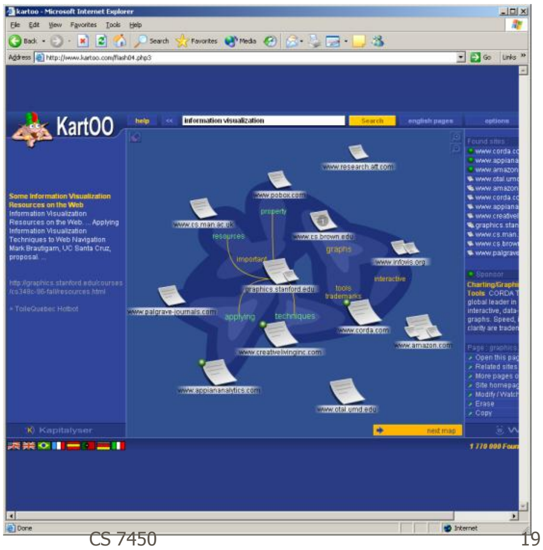
Issues

- Horizontal alignment doesn’t match mental model
- May not be the best solution for web searches
  - Non-linear material
  - Images? Java apps?
- Anything else?

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 traden

`http://www.kartoo.com`

Defunct

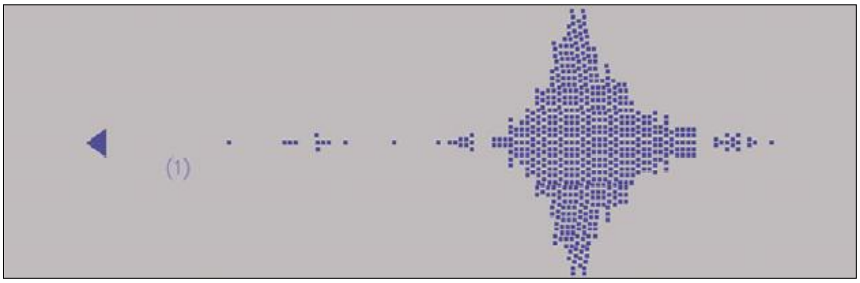
Sparkler

- Abstract result documents more
- Show “distance” from query in order to give user better feel for quality of match(es)
- Also shows documents in responses to multiple queries

Havre et al
InfoVis ‘01

Visualizing One Query

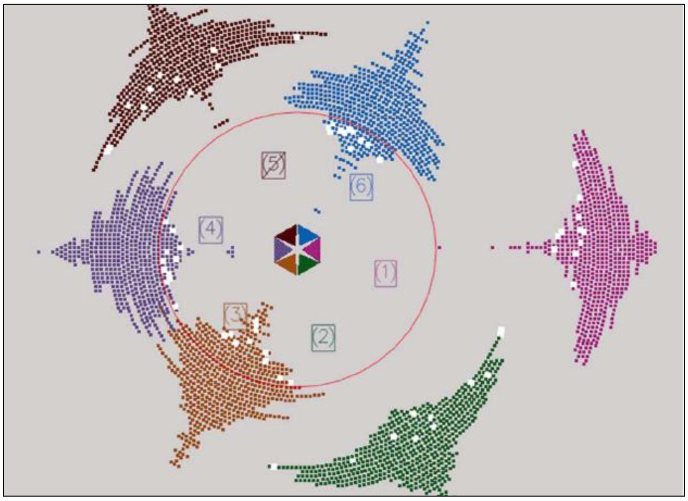
- Triangle – query
- Square – document
- Distance between query and documents represents their relevance



Visualizing Multiple Queries

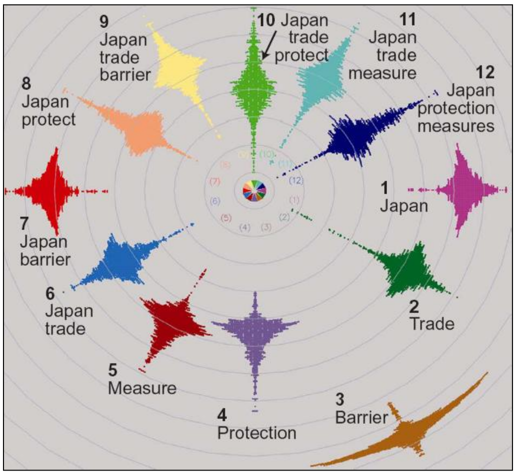
Six queries here

Bullseye allows viewer to select quality results

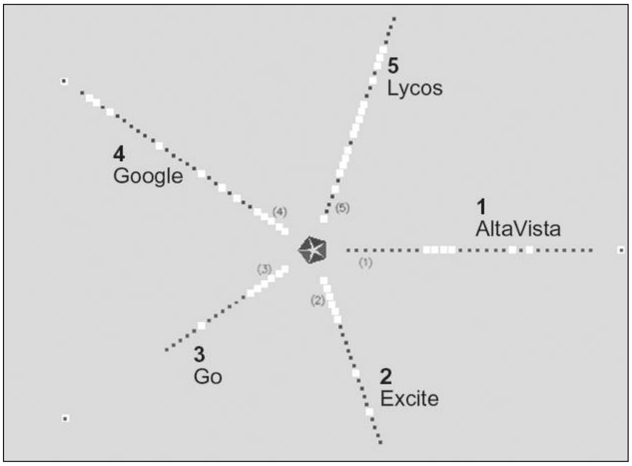


Test Example

- Text Retrieval Conference (TREC-3) test document collection
- AP news stories from June 24–30, 1990
- TREC topic: Japan Protectionist Measures
- Sparkler found 16 of 17 relevant documents



Another Idea



Use it to compare search results from different search engines

RankSpiral

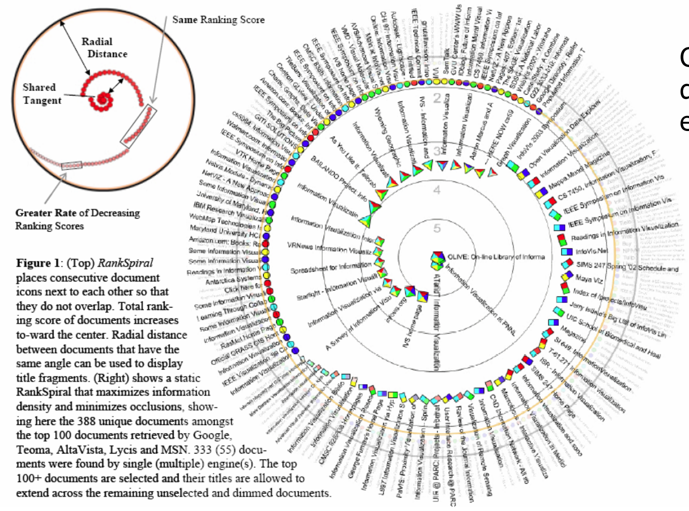
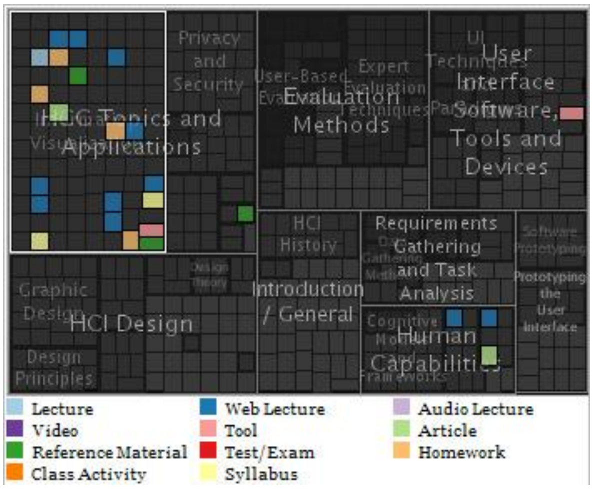

Same Ranking Score
Radial Distance
Shared Tangent
Greater Rate of Decreasing Ranking Scores
2
3
4
5
OLIVE: On-line Library of Informa

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

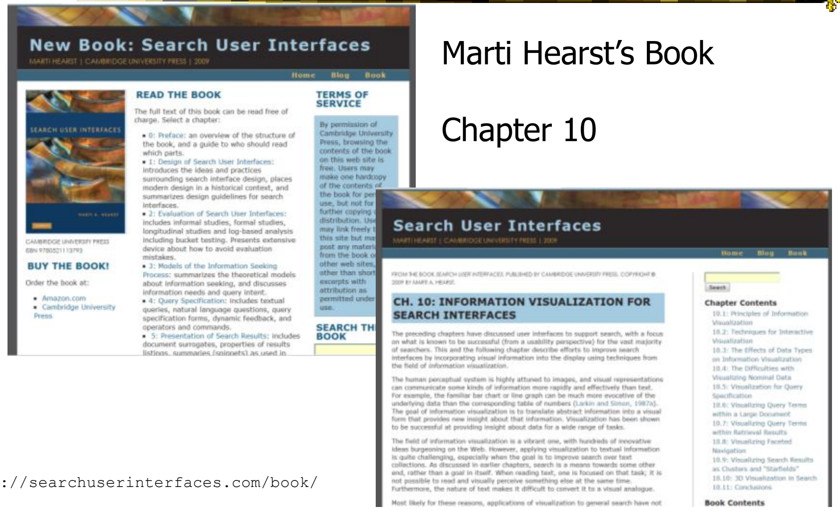
ResultMaps



Treemap-style vis for showing query results in a digital library

Clarkson, Desai & Foley
TVCG (InfoVis) '09

To Learn More



New Book: Search User Interfaces

MARTI HEARST | CAMBRIDGE UNIVERSITY PRESS | 2009

SEARCH USER INTERFACES

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 materi from the book o other web sites, other than short excerpts with attribution as permitted under use.

SEARCH TH BOOK

Search User Interfaces

MARTI HEARST | CAMBRIDGE UNIVERSITY PRESS | 2009

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

Marti Hearst's Book

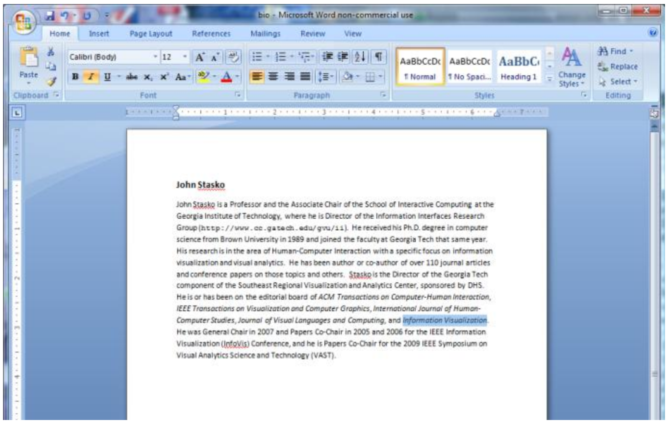
Chapter 10

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
…

Tag/Word Clouds

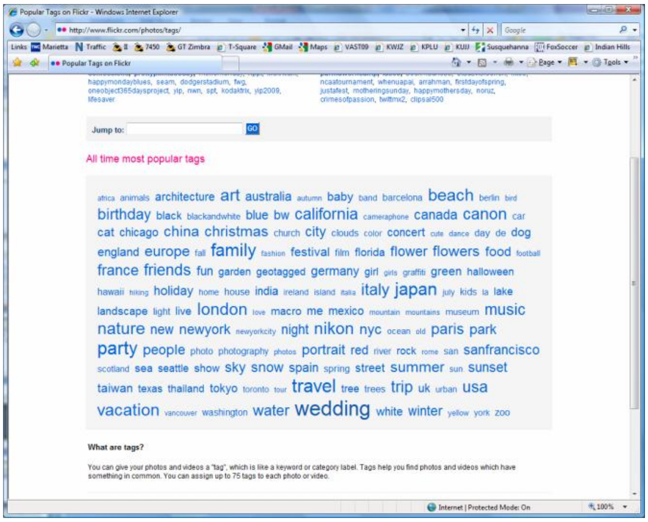
- Currently very “hot” in research community
- Have proven to be very popular on web
- Idea is to show word/concept importance through visual means
  - Tags: User-specified metadata (descriptors) about something
  - Sometimes generalized to just reflect word frequencies

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



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

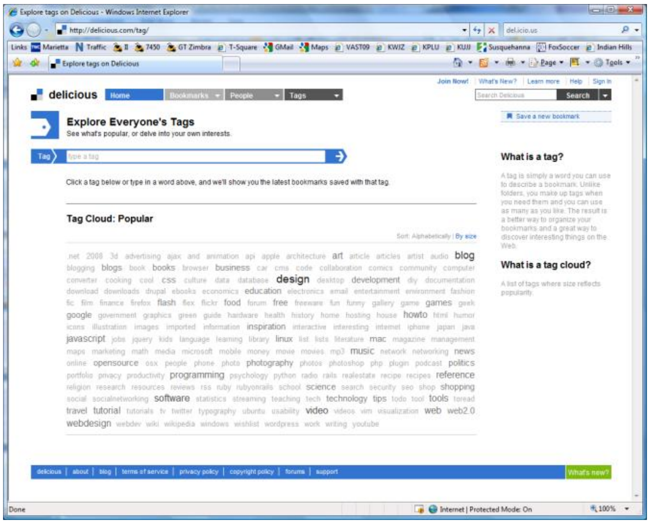
All time most popular tags

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
Done
Internet | Protected Mode: On
100%

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

Tag type a tag

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?

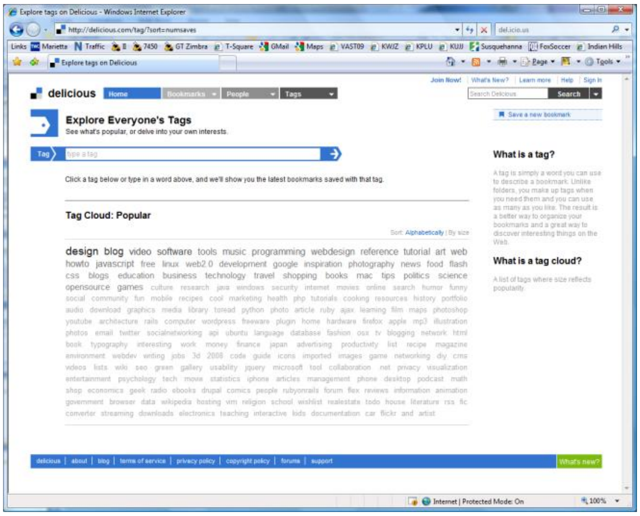
A tag is simply a word you can use to describe a bookmark. Unlike folders, you make up tags when you need them and you can use as many as you like. The result is a better way to organize your bookmarks and a great way to discover interesting things on the Web.

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?

Alternate Order


Explore tags on Delicious - Windows Internet Explorer
http://delicious.com/tag/?sort=numsaves
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
Internet | Protected Mode: On
100%

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

Tag type a tag

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.

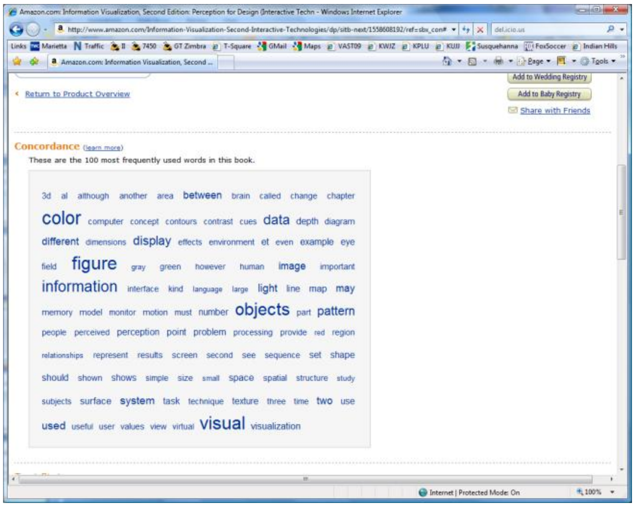
What is a tag cloud?

A list of tags where size reflects popularity.

delicious | about | blog | terms of service | privacy policy | copyright policy | forums | support What's new?

Amazon’s Product Concordance

Maybe now a “word cloud”


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 II 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
Internet | Protected Mode: On
100%

‹ Return to Product Overview

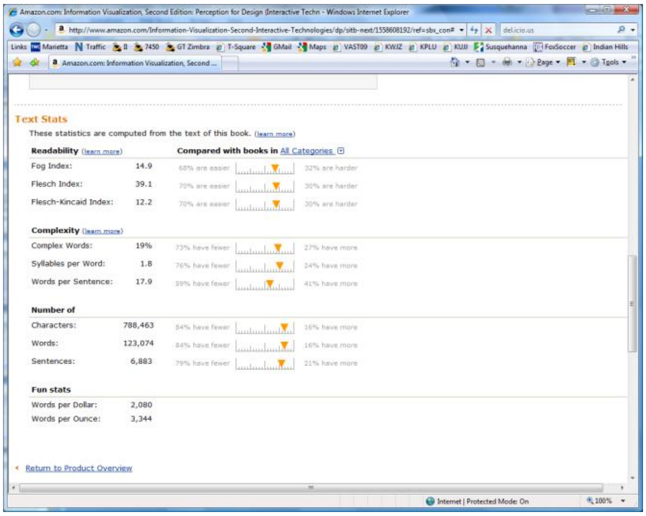
Concordance (learn more)

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

Sidenote

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 II 7450 GT Zimbra T-Square GMail Maps VAST09 KWJZ KPLU KUJJ Susquehanna FoxSoccer Indian Hills
Amazon.com: Information Visualization, Second ...
Page Tools
Internet | Protected Mode: On
100%

Text Stats

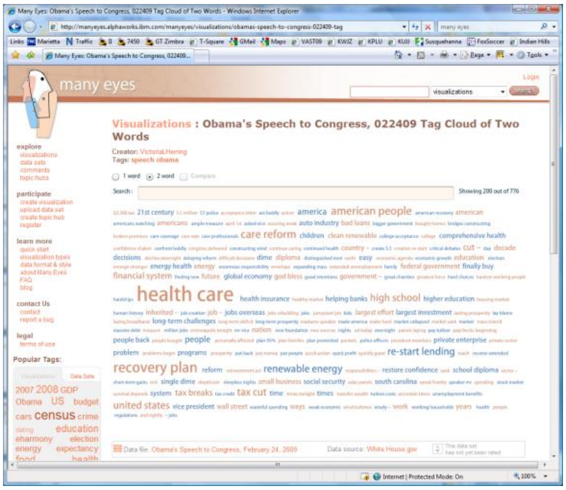
These statistics are computed from the text of this book. (learn more)

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



Many Eyes: Obama's Speech to Congress, 022409 Tag Cloud of Two Words - Windows Internet Explorer

http://manyeyes.alphaworks.ibm.com/manyeyes/visualizations/obamas-speech-to-congress-022409-tag

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

many eyes
Login
visualizations
Search
explore
visualizations
data sets
comments
topic hubs
participate
create visualization
upload data set
create topic hub
register
learn more
quick start
visualization types
data format & style
about Many Eyes
FAQ
blog
contact Us
contact
report a bug
legal
terms of use
Popular Tags:
Data Sets
2007 2008 GDP
Obama US budget
cars census crime
dating education
eharmony election
energy expectancy
food health
21st century america american people american
americans auto industry bad loans
care reform children clean renewable comprehensive health
country cut decade
decisions dime diploma easy education
energy health energy federal government finally buy
financial system future global economy god bless government
health care health insurance helping banks high school higher education
inherited job jobs overseas largest effort largest investment
long-term challenges
nation
people back people private enterprise
problem programs re-start lending
recovery plan reform renewable energy restore confidence school diploma
single dime small business social security south carolina
system tax breaks tax cut time times
united states vice president wall street ways work years
Internet | Protected Mode: On
100%

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

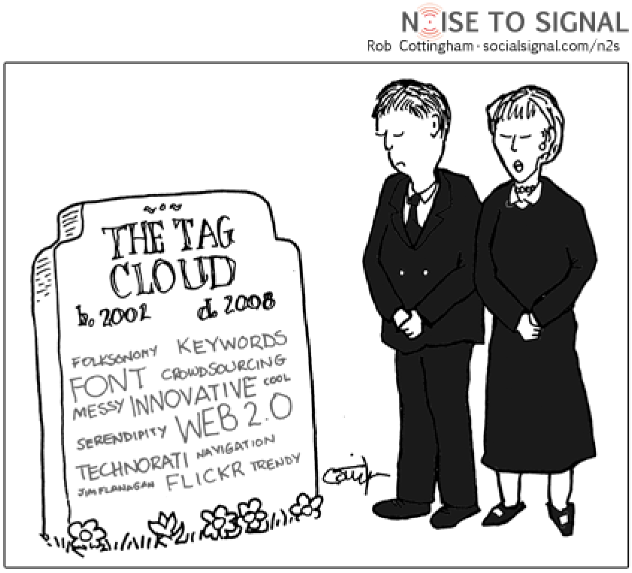
- Studies have even shown they underperform

Gruen et al
CHI '06

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

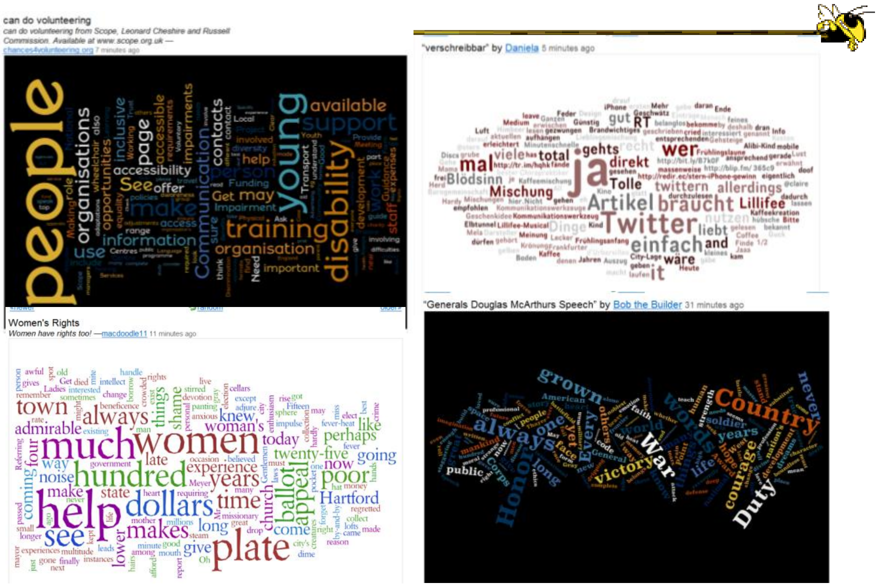
Hearst & Rosner
HICSS '08



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

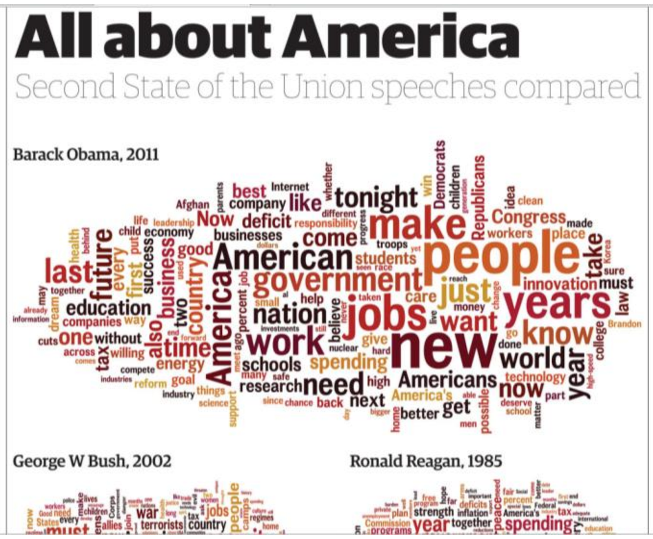
Fun Uses

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

2-day Survey in Jan. 09

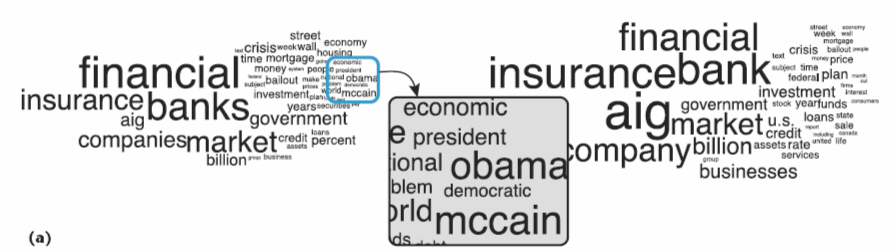
- 2/3 respondents were women
- Interest came from design, visual appeal, beauty
- Why preferred over word clouds:
  - Emotional impact
  - Attention-keeping visuals
  - Organic, non-linear
- Fair percentage didn’t know what size signified

SoTU Wordles



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

Wordle Characteristics

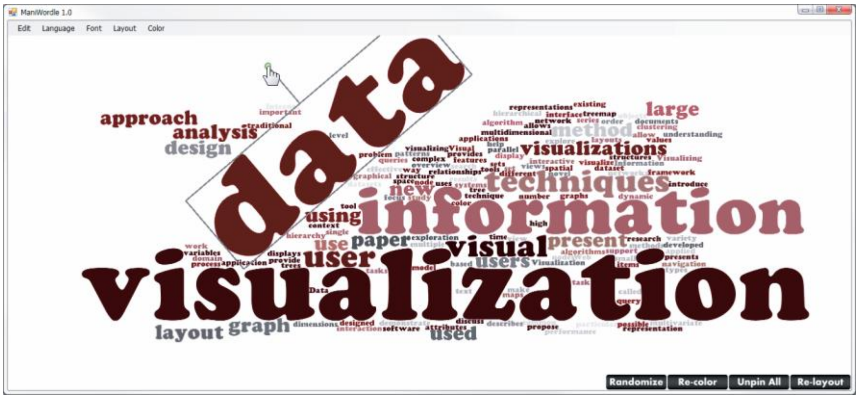
- Layout, words are automatic
- If you had some control, what would you like to change or alter?

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

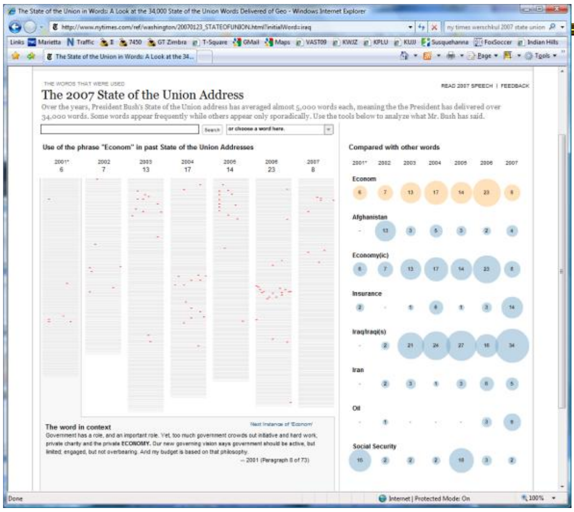
Video



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 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
ny times werschkul 2007 state union
Links Marietta Traffic 8 7450 GT Zimbra T-Square GMail Maps VAST09 KWJZ KPLU KUJJ Susquehanna FoxSoccer Indian Hills
The State of the Union in Words: A Look at the 34...
Page Tools
THE WORDS THAT WERE USED
READ 2007 SPEECH | FEEDBACK
Search
or choose a word here.
Next Instance of 'Econom'
Done
Internet | Protected Mode: On
100%

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.

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

2001*	2002	2003	2004	2005	2006	2007
6	7	13	17	14	23	8

Compared with other words

	2001*	2002	2003	2004	2005	2006	2007
Econom	6	7	13	17	14	23	8
Afghanistan	-	13	3	5	3	2	4
Economy(ic)	6	7	13	17	14	23	8
Insurance	2	-	1	4	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	16	2	2	2	18	3	2

The word in context

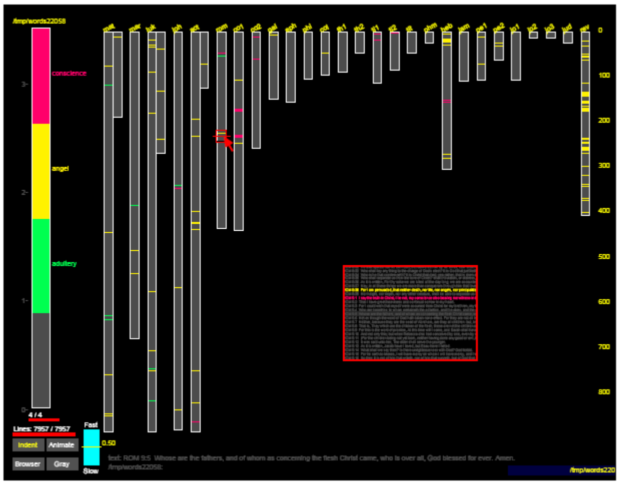
Government has a role, and an important role. Yet, too much government crowds out initiative and hard work, private charity and the private **ECONOMY.** Our new governing vision says government should be active, but limited; engaged, but not overbearing. And my budget is based on that philosophy.

— 2001 (Paragraph 8 of 73)

State of the Union Addresses

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

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

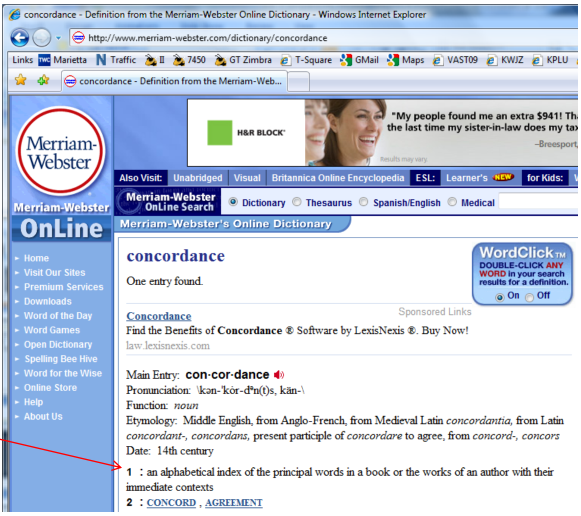
Eick
Journal Comput. & Graph. Stats ‘94

Beyond Individual Words

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

Concordance

Definition



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
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
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
WordClick™ DOUBLE-CLICK ANY WORD in your search results for a definition. ● On ○ Off

H&R BLOCK "My people found me an extra $941! Th the last time my sister-in-law does my tax –Breesport, Results may vary.

concordance

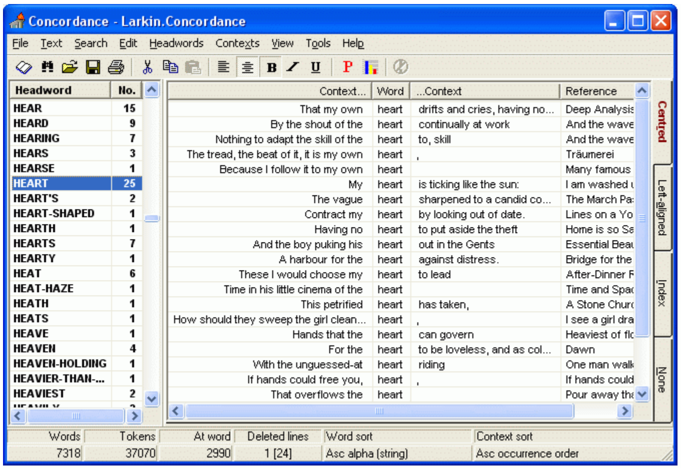
One entry found.

Sponsored Links
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

Concordance in Text



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

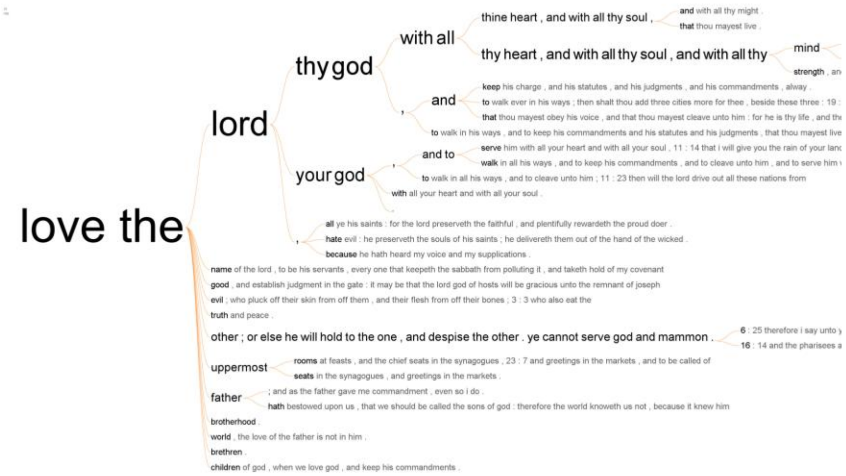
Headword	No.
HEAR	15
HEARD	9
HEARING	7
HEARS	3
HEARSE	1
HEART	25
HEART'S	2
HEART-SHAPED	1
HEARTH	1
HEARTS	7
HEARTY	1
HEAT	6
HEAT-HAZE	1
HEATH	1
HEATS	1
HEAVE	1
HEAVEN	4
HEAVEN-HOLDING	1
HEAVIER-THAN-...	1
HEAVIEST	2

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
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 Bea
A harbour for the	heart	against distress.	Bridge for the
These I would choose my	heart	to lead	After-Dinner F
Time in his little cinema of the	heart		Time and Spa
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 flc
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 th

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 lanc
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 .

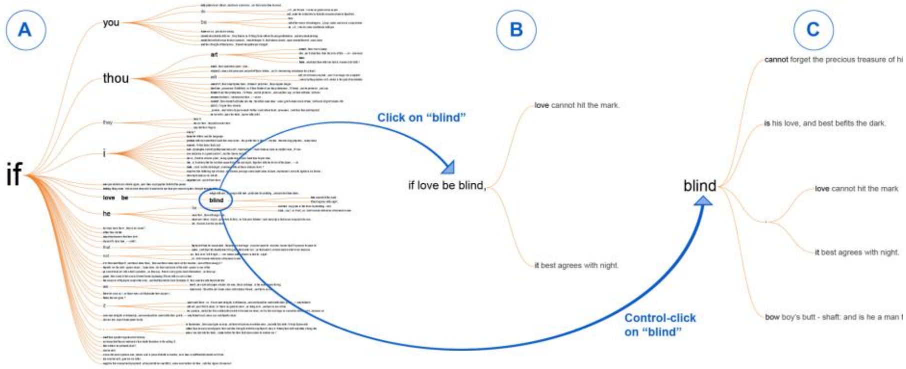
From King James Bible

Word Tree

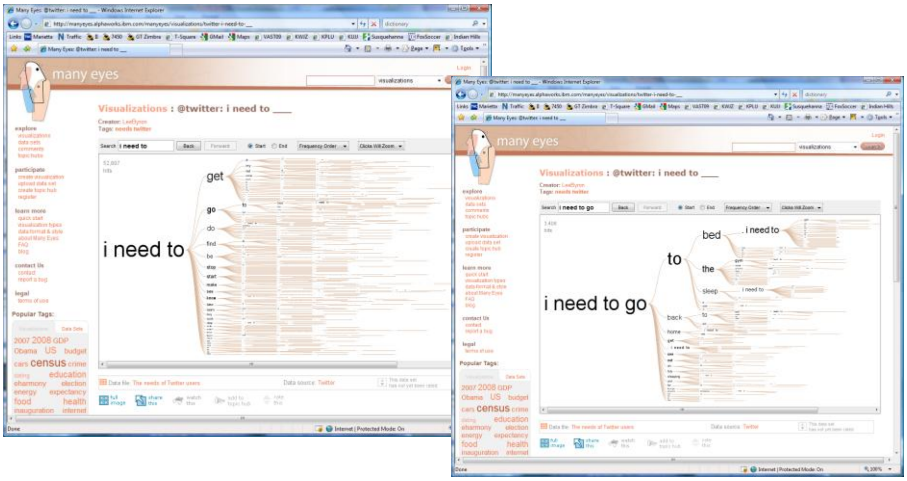
- Shows context of a word or words
  - Follow word with all the phrases that follow it
- Font size shows frequency of appearance
- Continue branch until hitting unique phrase
- Clicking on phrase makes it the focus
- Ordered alphabetically, by frequency, or by first appearance

Wattenberg & Viégas
TVCG (InfoVis) '08

Interaction


A
B
C
if
you
thou
i
love be
he
blind
Click on “blind”
if love be blind,
love cannot hit the mark.
it best agrees with night.
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


many eyes
Visualizations : @twitter: i need to ___
i need to
get
go
do
find
many eyes
Visualizations : @twitter: i need to ___
i need to go
to
bed
the
sleep
back
home

In Many Eyes now

Phrase Nets

- Examine unstructured text documents
- Presents pairs of terms from phrases such as
  - X and Y
  - X’s Y
  - X at Y
  - X (is|are|was|were) Y
- Uses special graph layout algorithm with compression and simplification

van Ham et al
TVCG (InfoVis) ‘09

Examples

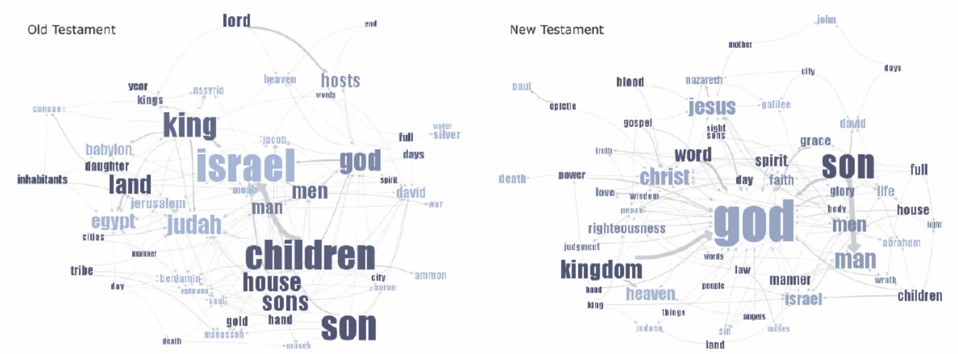


Fig 4. Matching the same pattern on different texts. Here we used the pattern “X of Y” to compare the old and new testaments. Israel takes a central place in the Old Testament, while God acts as the main pattern receiver in the New Testament.

Examples

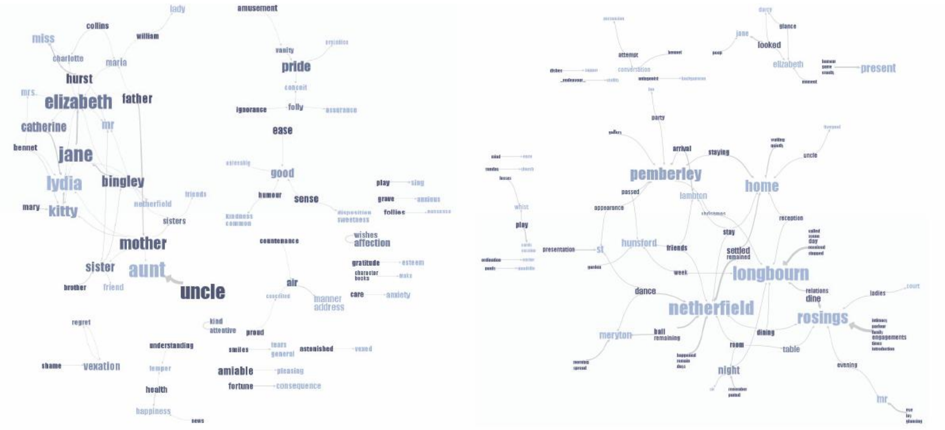
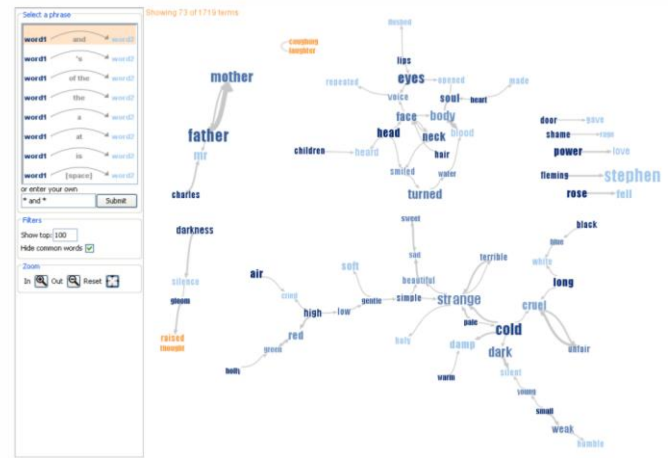


Fig 5. Matching different patterns on the same text. Here we analyzed Jane Austen's *Pride and Prejudice* with “X and Y” and “X at Y” respectively. The left image shows relationships between the main characters amongst others, while the right image shows relationships between locations.

User Interface



Select a phrase

word1 and word2
word1 's word2
word1 of the word2
word1 the word2
word1 a word2
word1 at word2
word1 is word2
word1 [space] word2

or enter your own
* and *
Submit

Filters
Show top: 100
Hide common words ☑

Zoom
In Out Reset

Showing 73 of 1719 terms

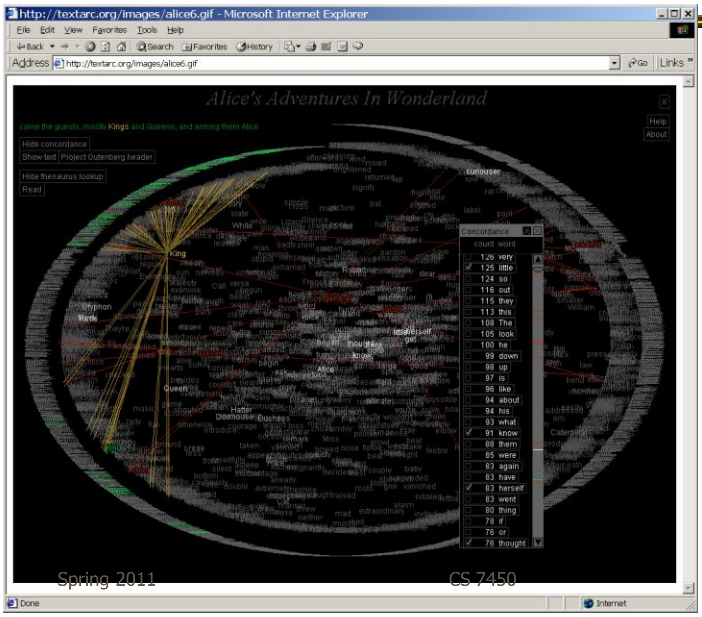
Fig 3. The Phrase Net user interface applied to James Joyce's Portrait of the Artist as a Young Man. The user can select a predefined pattern from the list of patterns on the left or define a custom pattern in the box below. This list of patterns simultaneously serves as a legend, a list of presets and 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.

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
File Edit View Favorites Tools Help
Back Search Favorites History
Address http://textarc.org/images/alice6.gif
Go Links
Alice's Adventures In Wonderland
came the guests, mostly Kings and Queens, and among them Alice
Hide concordance
Show text
Project Gutenberg header
Hide thesaurus lookup
Read
Help
About
King
Queen
Alice
curiouser
Concordance
Done
Internet

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

Upcoming

- Text and Documents 2
  - Reading

    Keim & Oelke '07

- Spring Break
- Visual Analytics 1
  - Reading

    Keim et al '08

References

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