Information Visualization:
Text Visualization

Dr. Christopher Collins

Acknowledgement: Parts of this lecture are based on material prepared by Tamara Munzner.

TEXT VISUALIZATION

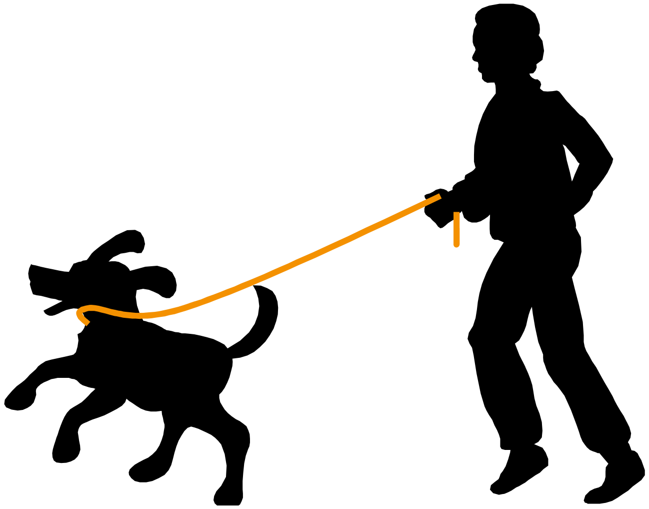
The dog.

The excited dog .

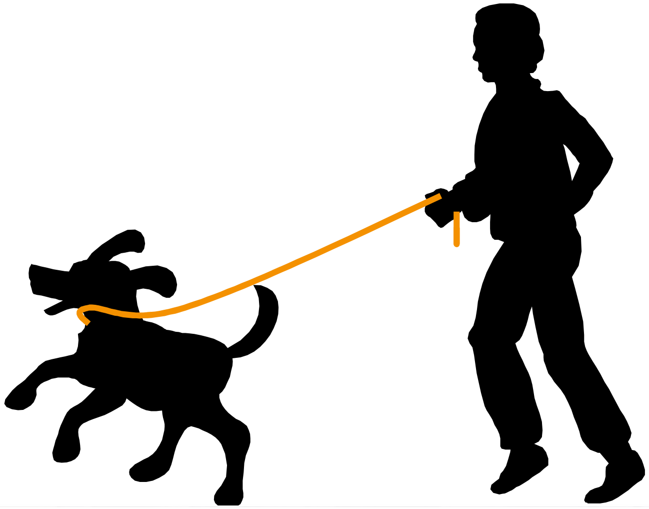
The man.

The man walks.

The man walks the excited dog.



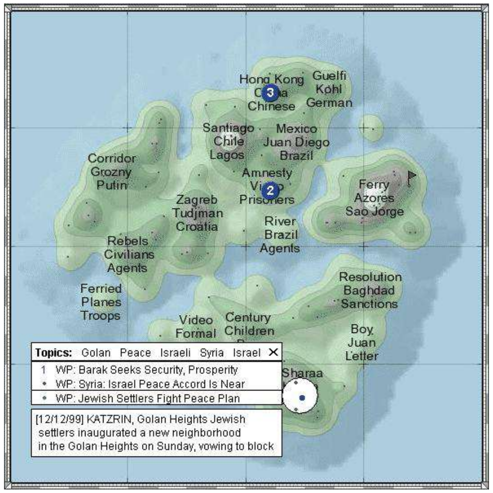
As the man walks the excited dog, he daydreams of the coming spring, and is filled with dread, as he is every year when the days drag on longer, the happy sun grinning sardonically at him as he enters his windowless workplace prison for the most hectic and stressful time of year.



Example based on lecture notes of Marti Hearst, 2006

Why Visualize Language?

- To assist information retrieval
- To enable linguistic analysis
- To augment analytics on mixed data

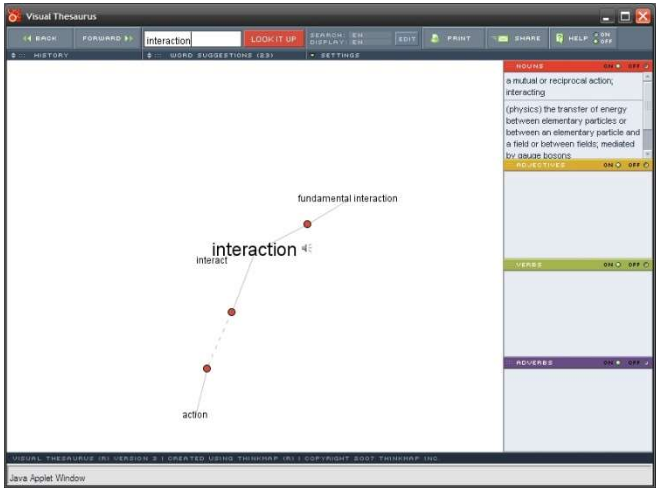


Topics: Golan Peace Israeli Syria Israel ✕

1 WP: Barak Seeks Security, Prosperity
• WP: Syria: Israel Peace Accord Is Near
• WP: Jewish Settlers Fight Peace Plan

[12/12/99] KATZRIN, Golan Heights Jewish settlers inaugurated a new neighborhood in the Golan Heights on Sunday, vowing to block

Themescape

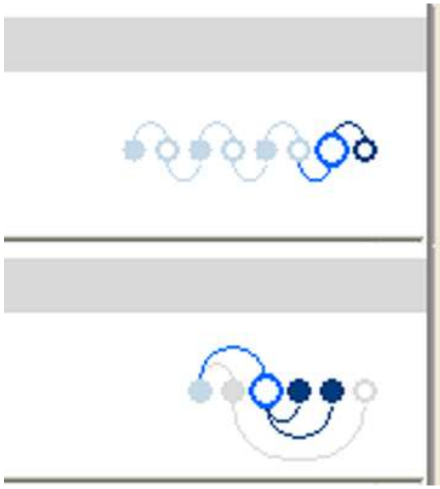


a mutual or reciprocal action; interacting

(physics) the transfer of energy between elementary particles or between an elementary particle and a field or between fields; mediated by gauge bosons

VISUAL THESAURUS (R) VERSION 3 | CREATED USING THINKMAP (R) | COPYRIGHT 2007 THINKMAP INC

Visual Thesaurus



Thread Arcs

Visualizing Language is Difficult

- Many of the common challenges still exist
- Can you name some?

Visualizing Language is Difficult

- Many of the common challenges still exist:
  - Screen real estate / occlusion
  - Choosing appropriate visual variable mappings
  - Colour and perception issues
  - Maintaining “graphical integrity”
  - Interaction and usability

- Specific challenges for language?

Difficult Data

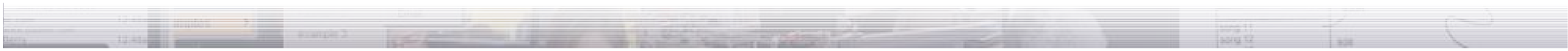
- Too much data – what to use?
  - Millions of blog posts,
  - Hundreds of thousands of news stories,
  - 183 billion emails,
  - **... per day**
- Data is noisy:
  - Newswire stories are syndicated (but differ slightly)
  - 70-72% of email is spam
  - Text contains section headings, figure captions, and direct quotes

Once you have the data...

- Most meaning comes from our minds and common understanding.

- “How much is that doggy in the window?”
  - how much: social system of barter and trade (not the size of the dog)
  - “doggy” implies childlike, plaintive, probably cannot do the purchasing on their own
  - “in the window” implies behind a store window, not really inside a window, requires notion of window shopping

(Hearst, 2006)



Language is Ambiguous

- Words and phrases can have many meanings, determined by context and world knowledge.
- Interesting language is often figurative:
  - “Tables encourage casual interaction.”

  vs

  - “I encouraged her to take a day off.”

Language is Ambiguous

- I saw Pathfinder on Mars with a telescope.
- Pathfinder photographed Mars.
- The Pathfinder photograph mars our perception of a lifeless planet.
- The Pathfinder photograph from Ford has arrived.
- The Pathfinder forded the river without marring its paint job.

(Hearst, 2006)

Data Processing Decisions

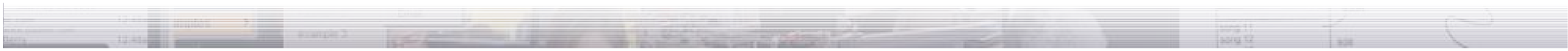
- Many levels of data processing can take place:
  - Word counting
  - Stemming
  - Parsing
  - Summarization
  - Sentiment analysis
  - Topic modelling
  - Word-Sense disambiguation
- Each step of extra processing introduces uncertainty

Visual Considerations

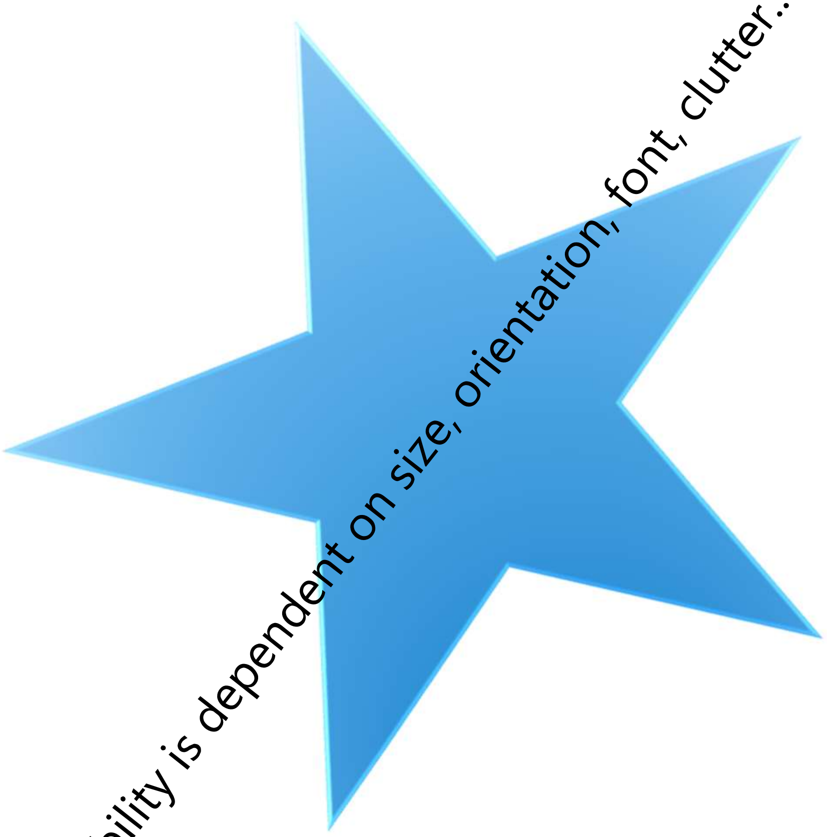
Supporters of Martin, who has been jailed without trial for more than two years, are calling on Prime Minister Stephen Harper to ask Mexican president Felipe Calderon to release Martin text is not preattentive under a section of the Mexican constitution that allows the government to expel undesirables from the country. Martin's supporters believe she has no chance of a fair trial in Mexico. Neither does Waage.

Visual Considerations

Supporters of Martin, who has been jailed without trial for more than two years, are calling on Prime Minister Stephen Harper to ask Mexican president Felipe Calderon to release Martin text is not preattentive under a section of the Mexican constitution that allows the government to expel undesirables from the country. Martin's supporters believe she has no chance of a fair trial in Mexico. Neither does Waage.



Visual Considerations



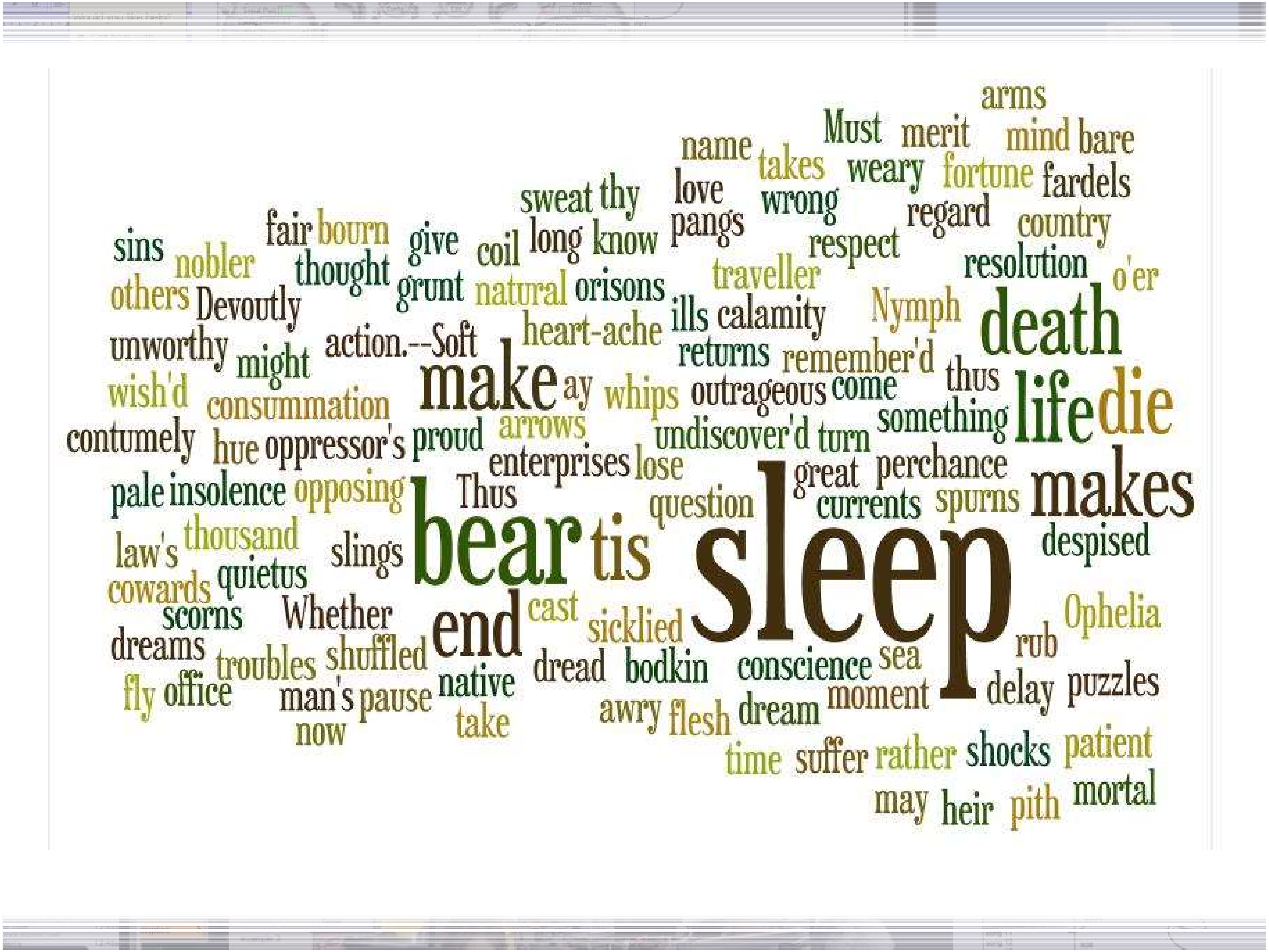
- Text readability is dependent on size, orientation, font, clutter…

Visual Considerations

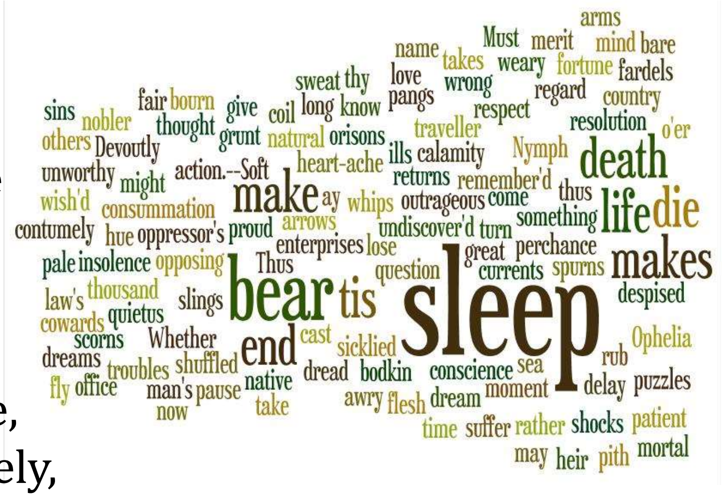
- Text readability is dependent on size, orientation, font, clutter...
- More likely to need large amounts of text in language visualization

Visualizing language is also easy!

- SO much data available for analysis
- (Mostly) readily computer readable
- Simple techniques can give instant summaries

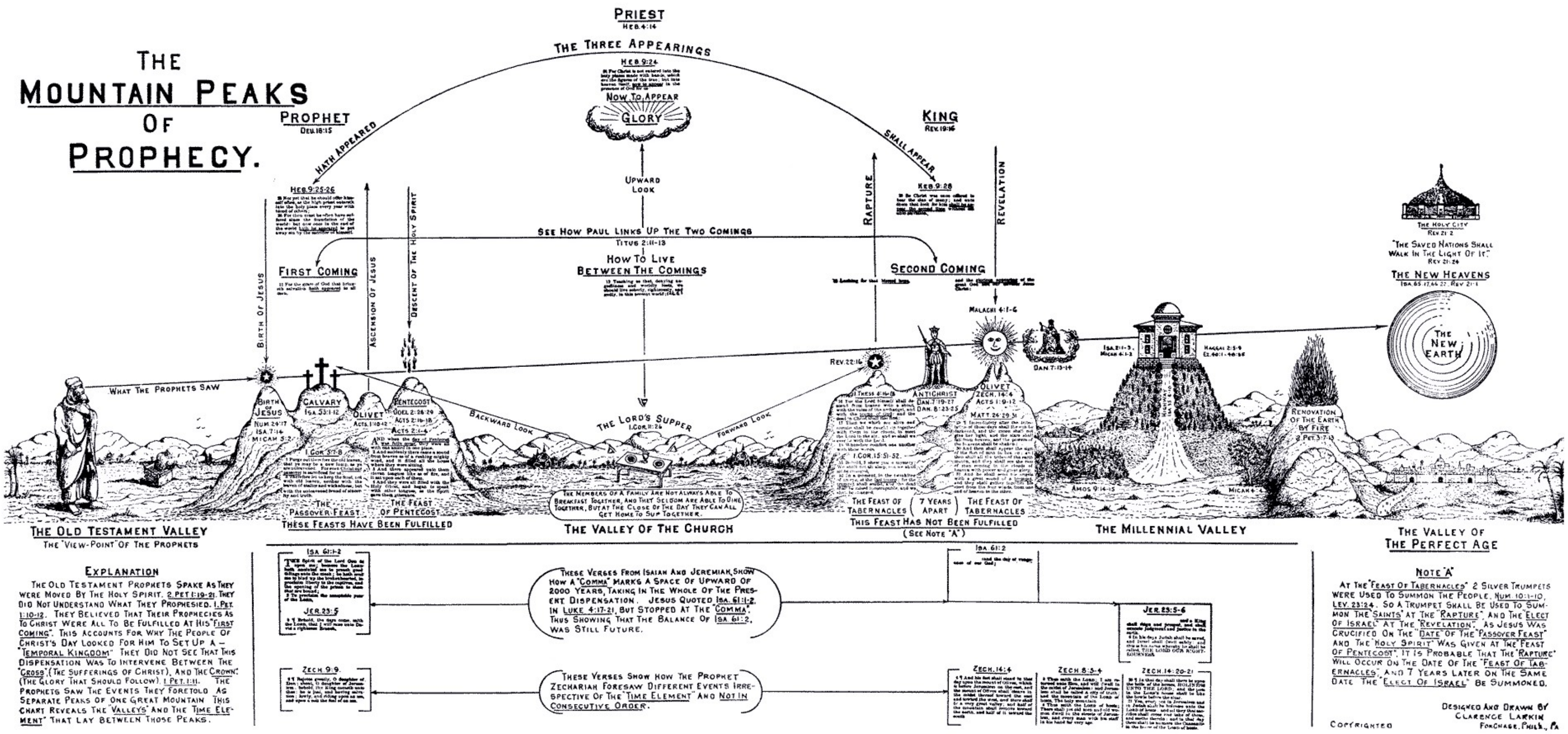

arms
Must
merit
mind
bare
name
takes
weary
fortune
fardels
love
wrong
regard
sweat
thy
pangs
country
fair
bourn
give
coil
long
know
respect
sins
nobler
thought
traveller
resolution
o'er
others
Devoutly
grunt
natural
orisons
ills
calamity
Nymph
death
unworthy
might
action.--Soft
heart-ache
returns
remember'd
thus
make
ay
whips
outrageous
come
life
die
wish'd
consummation
something
contumely
hue
oppressor's
proud
arrows
undiscover'd
turn
enterprises
lose
great
perchance
makes
pale
insolence
opposing
Thus
question
currents
spurns
law's
thousand
slings
bear
tis
sleep
despised
cowards
quietus
scorns
Whether
end
cast
sicklied
Ophelia
dreams
troubles
shuffled
dread
bodkin
conscience
sea
rub
fly
office
man's
pause
native
awry
flesh
dream
moment
delay
puzzles
now
take
time
suffer
rather
shocks
patient
may
heir
pith
mortal

To be, or not to be: that is the question:
Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune,
Or to take arms against a sea of troubles,
And by opposing end them? To die: to sleep;
No more; and by a sleep to say we end
The heart-ache and the thousand natural shocks
That flesh is heir to, 'tis a consummation
Devoutly to be wish'd. To die, to sleep;
To sleep: perchance to dream: ay, there's the rub;
For in that sleep of death what dreams may come
When we have shuffled off this mortal coil,
Must give us pause: there's the respect
That makes calamity of so long life;
For who would bear the whips and scorns of time,
The oppressor's wrong, the proud man's contumely,
The pangs of despised love, the law's delay,
The insolence of office and the spurns
That patient merit of the unworthy takes,
When he himself might his quietus make
With a bare bodkin? who would fardels bear,
…



Text Visualization

BACKGROUND



The Members Of A Family Are Not Always Able To Breakfast Together, And They Seldom Are Able To Dine Together, But At The Close Of The Day They Can All Get Home To Sup Together.

These Verses From Isaiah And Jeremiah Show How A "Comma" Marks A Space Of Upward Of 2000 Years, Taking In The Whole Of The Present Dispensation. Jesus Quoted Isa. 61:1-2, In Luke 4:17-21, But Stopped At The "Comma", Thus Showing That The Balance Of Isa. 61:2, Was Still Future.

These Verses Show How The Prophet Zechariah Foresaw Different Events Irrespective Of The "Time Element" And Not In Consecutive Order.

Explanation

The Old Testament Prophets Spake As They Were Moved By The Holy Spirit. 2 Pet 1:19-21. They Did Not Understand What They Prophesied. 1. Pet. 1:10-12. They Believed That Their Prophecies As To Christ Were All To Be Fulfilled At His "First Coming". This Accounts For Why The People Of Christ's Day Looked For Him To Set Up A – "Temporal Kingdom" They Did Not See That This Dispensation Was To Intervene Between The "Cross" (The Sufferings Of Christ), And The Crown" (The Glory That Should Follow). 1. Pet. 1:11. The Prophets Saw The Events They Foretold As Separate Peaks Of One Great Mountain This Chart Reveals The "Valleys" And The "Time Element" That Lay Between Those Peaks.

Note "A"

At The "Feast Of Tabernacles" 2 Silver Trumpets Were Used To Summon The People. Num. 10:1-10, Lev. 23:24. So A Trumpet Shall Be Used To Summon The "Saints" At The "Rapture", And The "Elect Of Israel" At The "Revelation". As Jesus Was Crucified On The "Date" Of The "Passover Feast" And The "Holy Spirit" Was Given At The "Feast Of Pentecost", It Is Probable That The "Rapture" Will Occur On The Date Of The "Feast Of Tabernacles", And 7 Years Later On The Same Date The "Elect Of Israel" Be Summoned.

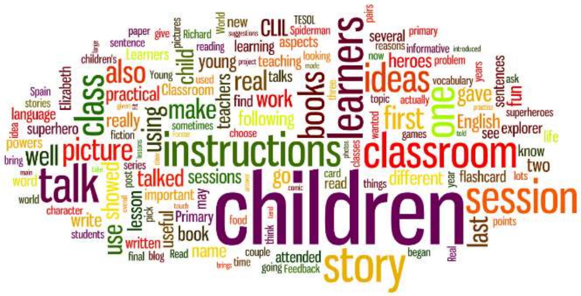
Designed And Drawn By
Clarence Larkin
Fox Chase, Phila., Pa

Copyrighted

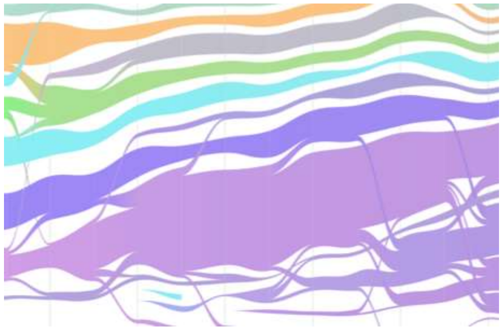
Mountain Peaks of Prophecy (Larkin, 1918)

Visual Text Analytics

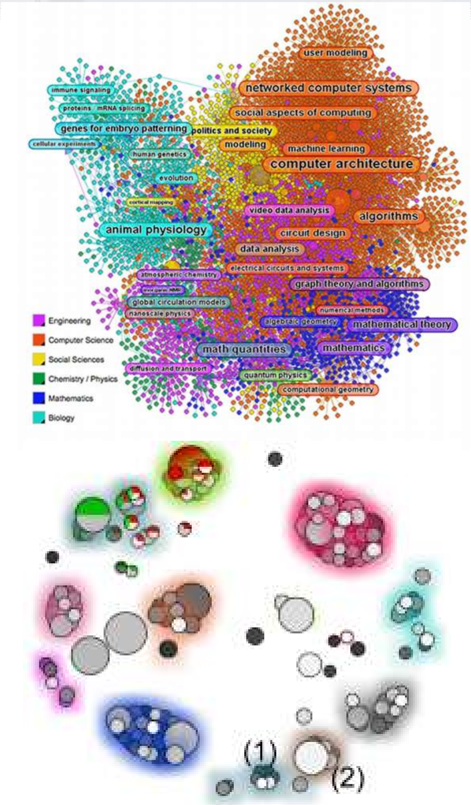
- Visual techniques for words, documents, sets of documents to support rapid summarization, trend analysis, exploration, search, comparative analysis, …
- Application areas include market analysis, legal studies, e-discovery, readability, literary studies, personal reflection, information retrieval and exploration, intelligence analysis

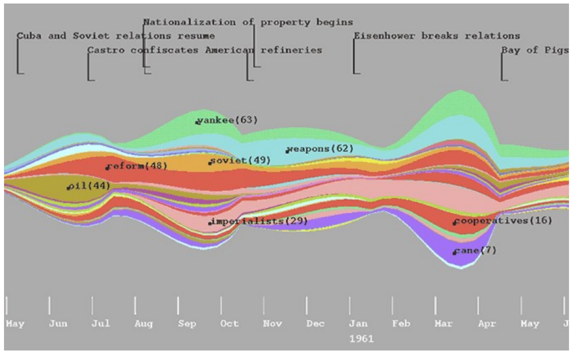


Word Clouds

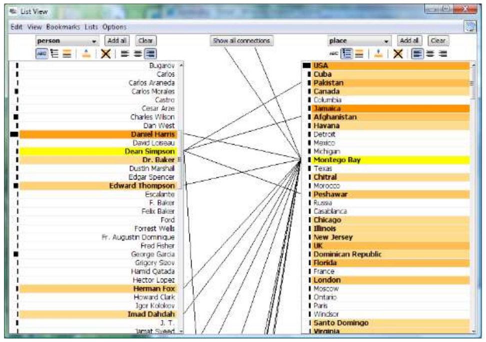


TextFlow



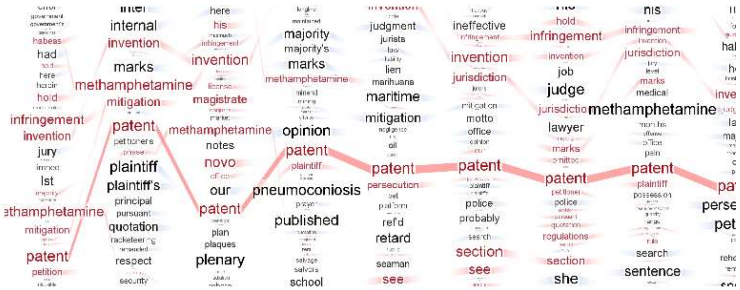


Theme River



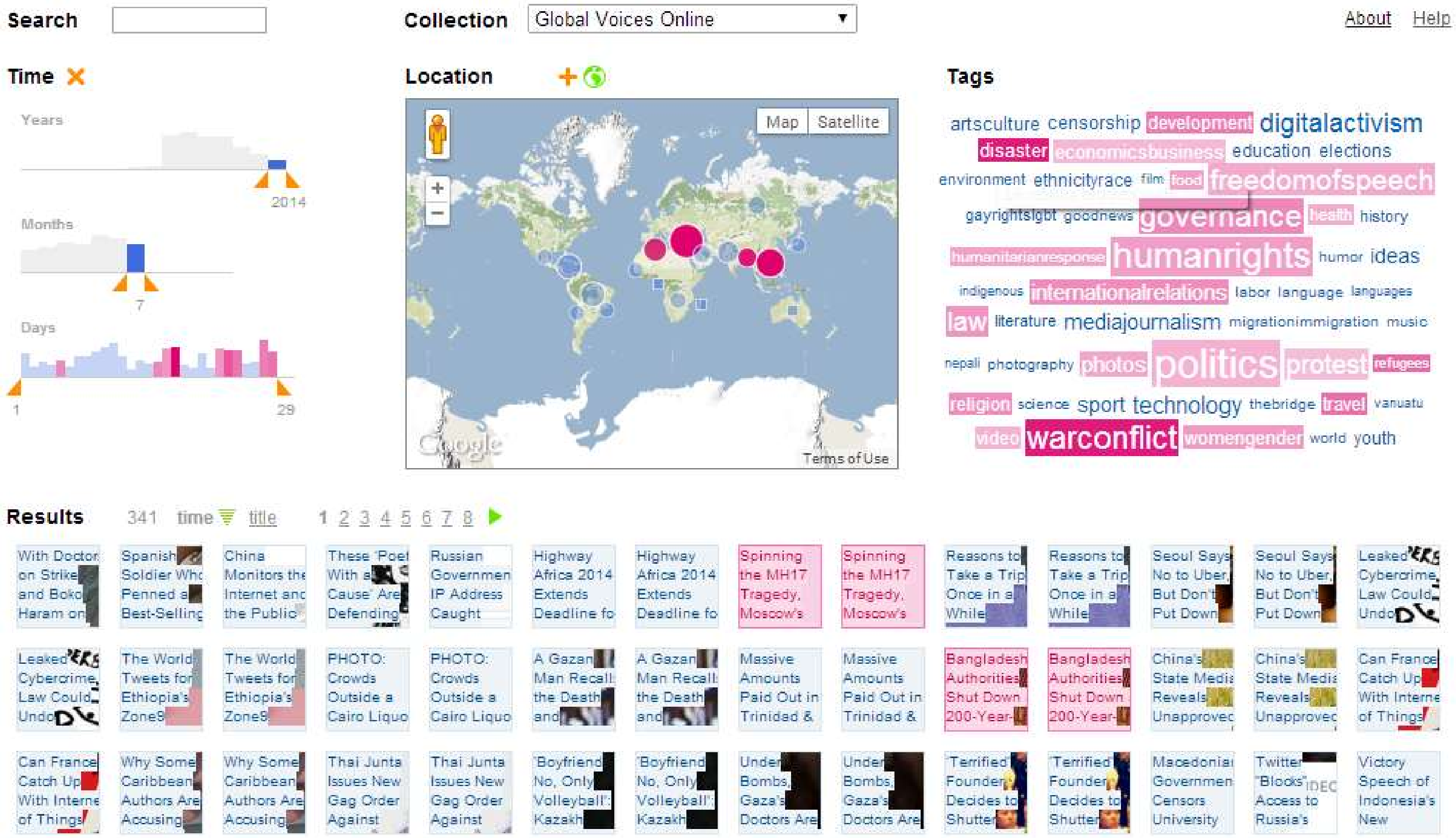
Jigsaw

Topic Models



Parallel Tag Clouds

Themescape



Search

Collection Global Voices Online ▾

About Help

Time ✕

Location

Tags

artsculture censorship development digitalactivism disaster economicsbusiness education elections environment ethnicityrace film food freedomofspeech gayrightslgbt goodnews governance health history humanitarianresponse humanrights humor ideas indigenous internationalrelations labor language languages law literature mediajournalism migrationimmigration music nepali photography photos politics protest refugees religion science sport technology thebridge travel vanuatu video warconflict womengender world youth

Results 341 time title 1 2 3 4 5 6 7 8 ▶

With Doctor on Strike and Boko Haram on | Spanish Soldier Who Penned a Best-Selling | China Monitors the Internet and the Public | These 'Poet With a Cause' Are Defending | Russian Government IP Address Caught | Highway Africa 2014 Extends Deadline fo | Highway Africa 2014 Extends Deadline fo | Spinning the MH17 Tragedy, Moscow's | Spinning the MH17 Tragedy, Moscow's | Reasons to Take a Trip Once in a While | Reasons to Take a Trip Once in a While | Seoul Says No to Uber, But Don't Put Down | Seoul Says No to Uber, But Don't Put Down | Leaked Cybercrime Law Could Undo

Leaked Cybercrime Law Could Undo | The World Tweets for Ethiopia's Zone9 | The World Tweets for Ethiopia's Zone9 | PHOTO: Crowds Outside a Cairo Liquo | PHOTO: Crowds Outside a Cairo Liquo | A Gazan Man Recalls the Death and | A Gazan Man Recalls the Death and | Massive Amounts Paid Out in Trinidad & | Massive Amounts Paid Out in Trinidad & | Bangladesh Authorities Shut Down 200-Year- | Bangladesh Authorities Shut Down 200-Year- | China's State Media Reveals Unapproved | China's State Media Reveals Unapproved | Can France Catch Up With Interne of Things

Can France Catch Up With Interne of Things | Why Some Caribbean Authors Are Accusing | Why Some Caribbean Authors Are Accusing | Thai Junta Issues New Gag Order Against | Thai Junta Issues New Gag Order Against | 'Boyfriend' No, Only Volleyball': Kazakh | 'Boyfriend' No, Only Volleyball': Kazakh | Under Bombs, Gaza's Doctors Are | Under Bombs, Gaza's Doctors Are | Terrified Founder Decides to Shutter | Terrified Founder Decides to Shutter | Macedonian Government Censors University | Twitter 'Blocks' Access to Russia's | Victory Speech of Indonesia's New

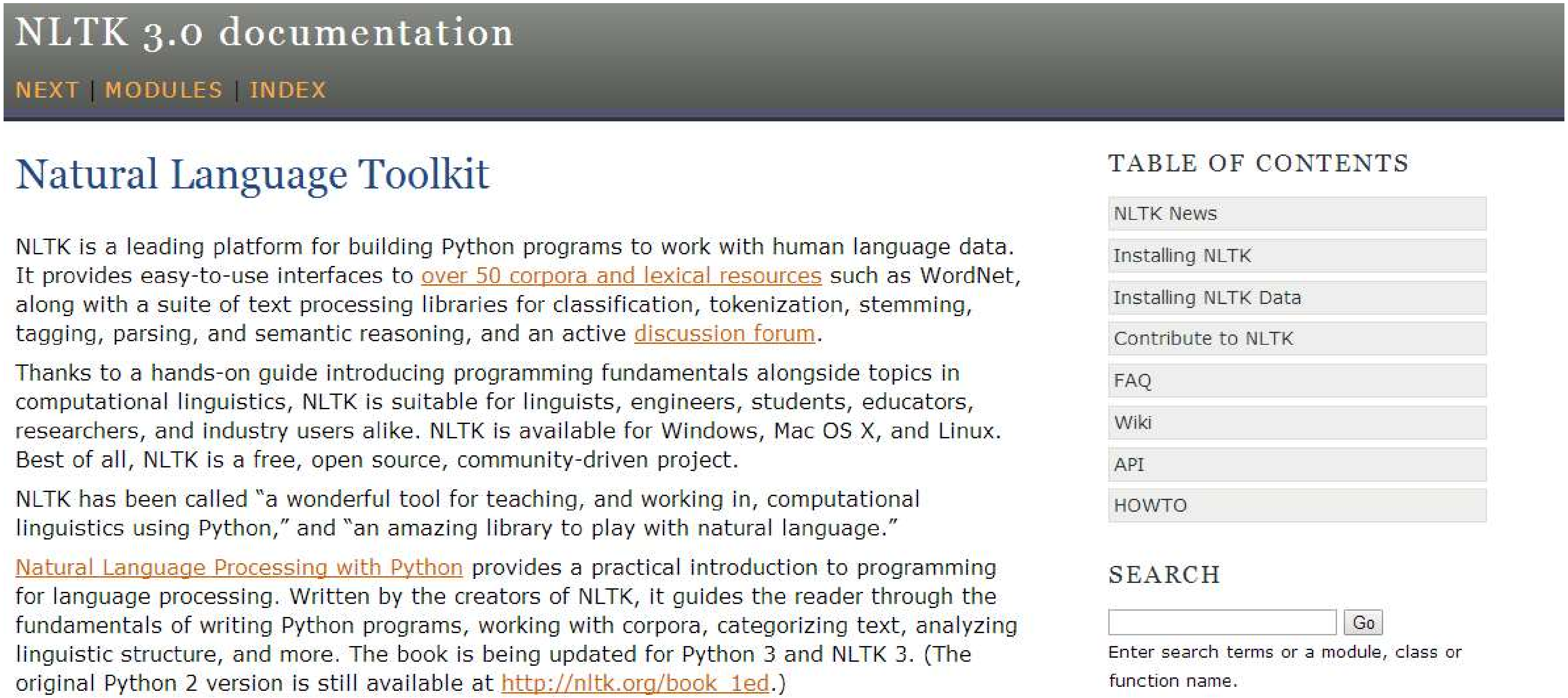
Marian Dörk et al. VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1205-1212, November-December, 2008.

Linguistic Methods

- Word Counting
- Word Scoring
- Stemming
- Stop Word Removal
- Part of Speech Tagging
- Parsing
- Word Sense Disambiguation
- Named Entity Recognition
- Semantic Categorization
- Sentiment Analysis
- Topic Modeling (some caveats)

NLTK: Natural Language Toolkit

- NLTK.org
- Python



NLTK 3.0 documentation

Natural Language Toolkit

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, and an active discussion forum.

Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project.

NLTK has been called "a wonderful tool for teaching, and working in, computational linguistics using Python," and "an amazing library to play with natural language."

Natural Language Processing with Python provides a practical introduction to programming for language processing. Written by the creators of NLTK, it guides the reader through the fundamentals of writing Python programs, working with corpora, categorizing text, analyzing linguistic structure, and more. The book is being updated for Python 3 and NLTK 3. (The original Python 2 version is still available at http://nltk.org/book_1ed.)

TABLE OF CONTENTS

- NLTK News
- Installing NLTK
- Installing NLTK Data
- Contribute to NLTK
- FAQ
- Wiki
- API
- HOWTO

SEARCH

Enter search terms or a module, class or function name.

Stemming

- Reduce words to their ‘stems’ by removing endings (morphology)
    - running -> run
    - runs -> run

- A good way to increase signal and reduce fracturing of the corpus if there aren’t many words.
- Note: Keep the original words somewhere! Also keep the case if you choose to lowercase the word; you never know when you’ll need this data

Stop Word Removal

- Common words such as “and”, “the”, “I” are removed from view to highlight content words
- Domain specific stop words, e.g. in legal domain:
  – Court, attorney, honour, plaintiff, etc.
- Caution! These words have been shown to be useful for stylistic analysis! When working with text corpora, KEEP EVERYTHING.

Part of Speech Tagging

- Assign grammatical roles to words
- Conventional tagsets and representation:
  - The/AT grand/JJ jury/NN commented/VBD on/IN a/AT number/NN of/IN …
- Many words are ambiguous: fly, chair, run, store, table, and more!
  - Fly/NN
  - Fly/VB

Fly/NN

Fly/VB

Term / Concept Ambiguity

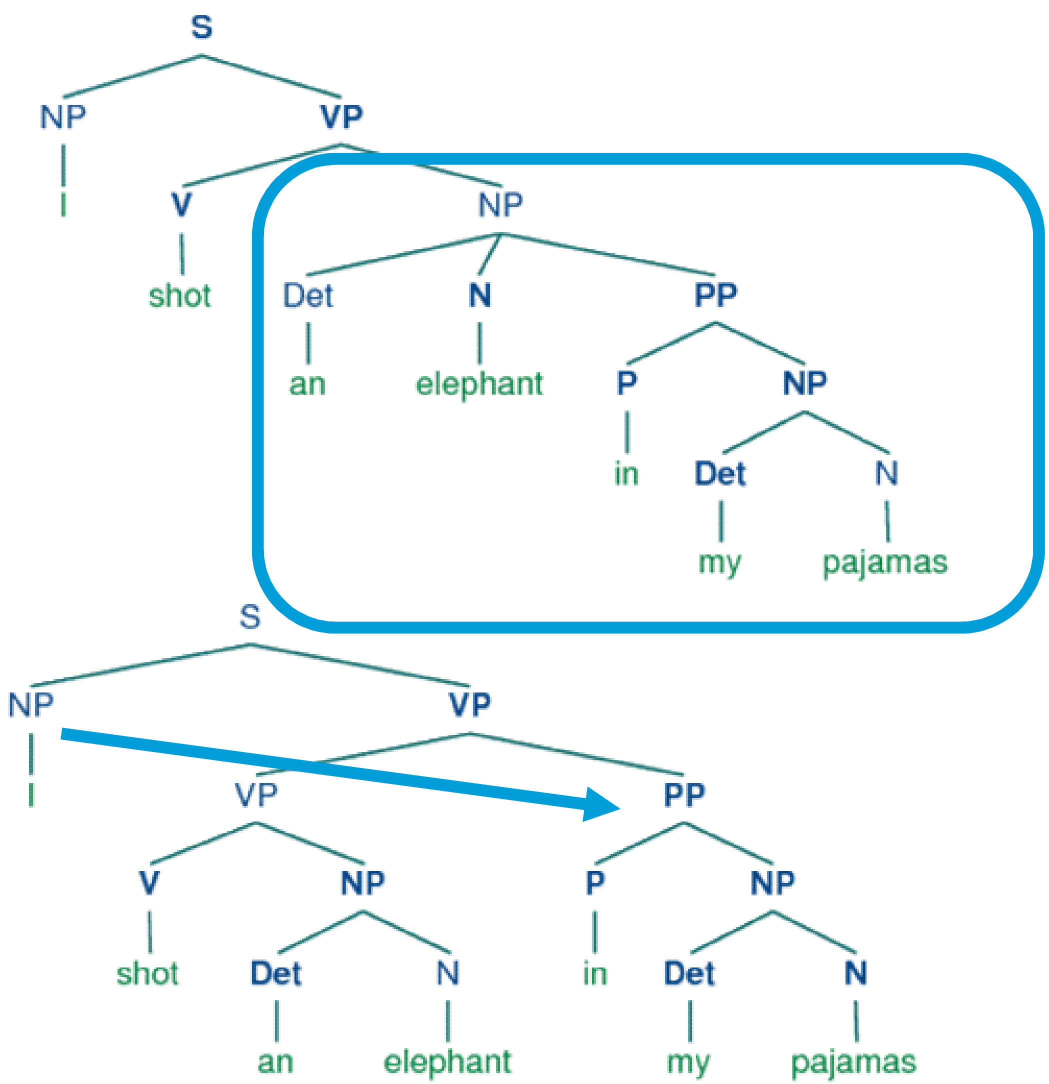
- Most meaning comes from our minds and common understanding.

- “How much is that doggy in the window?”
  - how much: social system of barter and trade (not the size of the dog)
  - “doggy” implies childlike, plaintive, probably cannot do the purchasing on their own
  - “in the window” implies behind a store window, not really inside a window, requires notion of window shopping

(Hearst, 2006)

Parsing

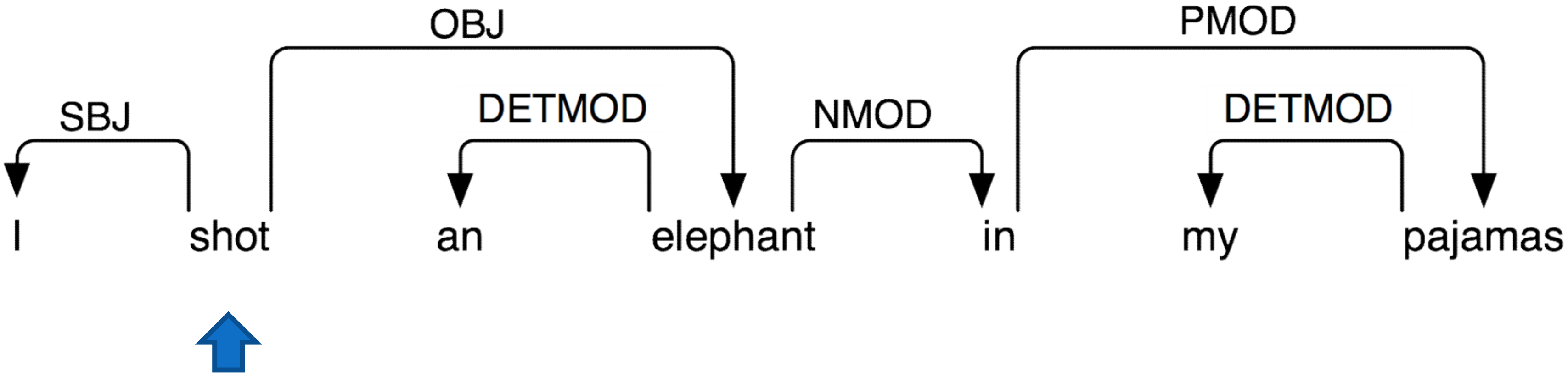
- Determining language structure
- Can reveal word-word relationships
- Useful for processing negation



https://nltk.googlecode.com/svn/trunk/doc/book/ch08.html

Dependency Parsing

- Labelled directed graph
- Arcs represent relationships from heads to dependents



Head of sentence

https://nltk.googlecode.com/svn/trunk/doc/book/ch08.html

Word Sense Disambiguation

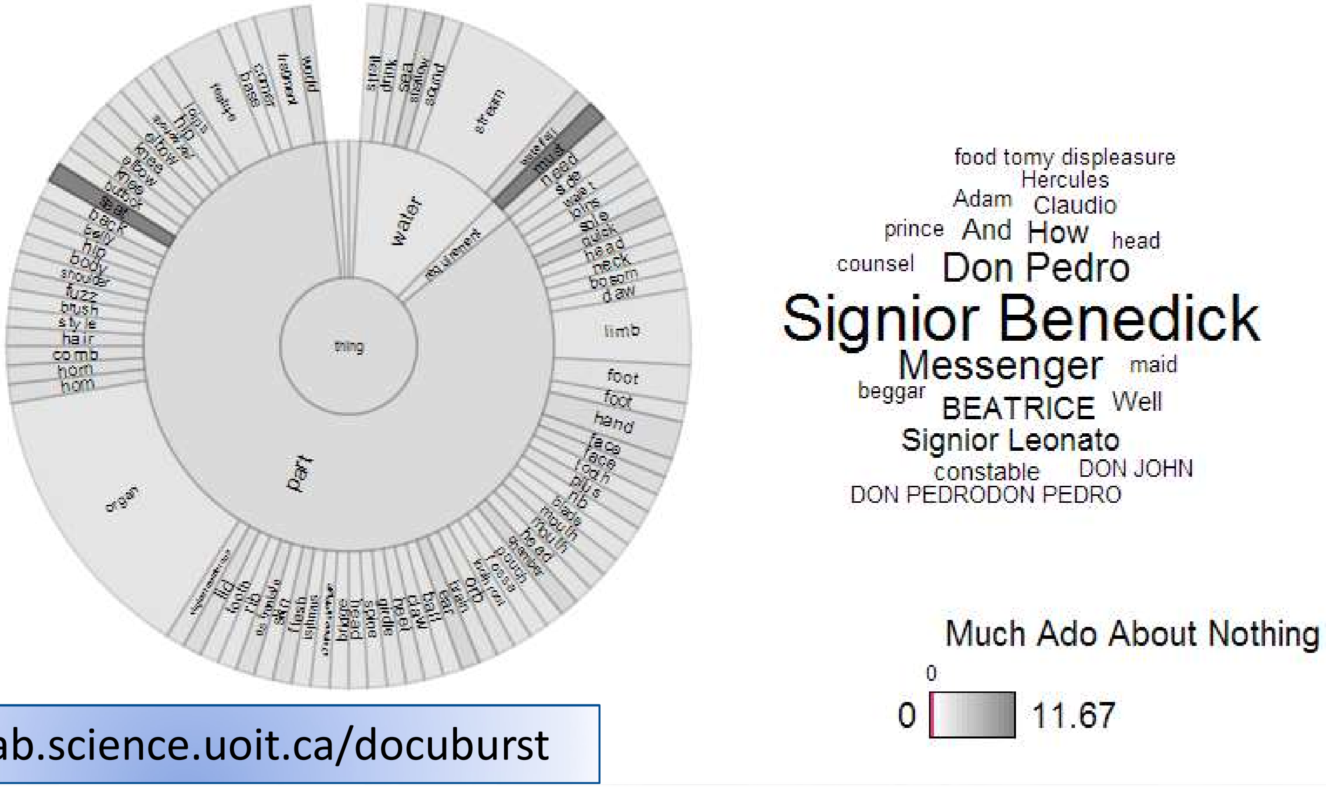
- Susan, the meeting chair, chaired the meeting well from the big chair in the front of the room.
  - Leader of a meeting
  - Action of leading a meeting
  - An object to sit upon

Word Sense Disambiguation

- This is VERY difficult for a computer.
- Contexts are often the same and meanings can be quite fine-grained:
  - bank the financial institution, bank the building in which the financial institution is housed
- Annual contest: SENSEVAL
- My method: assume the most common sense

Named Entity Recognition

- What are the people, places in the text?
- Use NLTK – it's very good at this.



http://vialab.science.uoit.ca/docuburst

Semantic Categorization

- Placing a word into an ontology or sense thesaurus based on *meaning.*

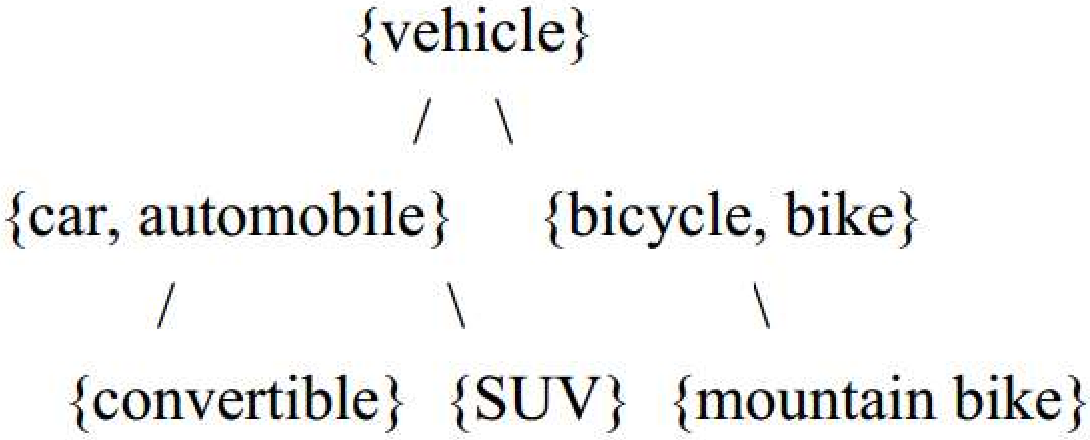
- Common resources include:
  - WordNet
  - Roget’s Thesaurus

WordNet

- A large lexical database, or “digital dictionary”
- Covers most English nouns, verbs, adjectives, adverbs
- Organizes *synsets* by *meaning*
- Words are related to one another through many different relationship types:
- X is a kind of Y, X has part Y, an X Ys, X is Y/has property Y

Hyponymy

- The “IS-A” relation for nouns



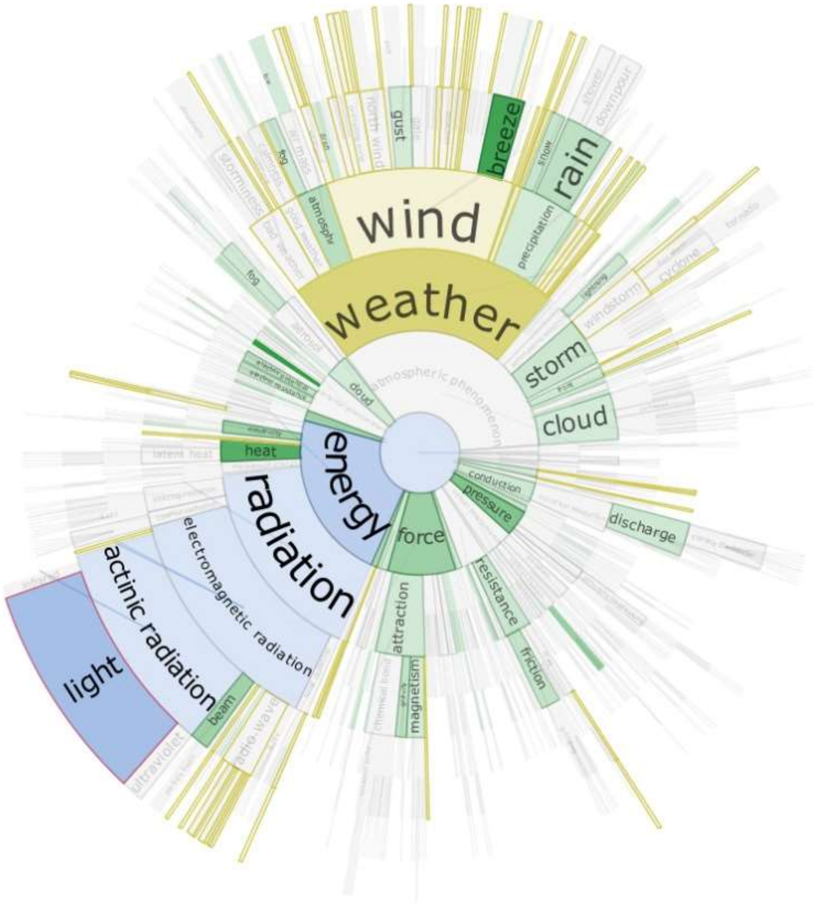
{vehicle}

{car, automobile} {bicycle, bike}

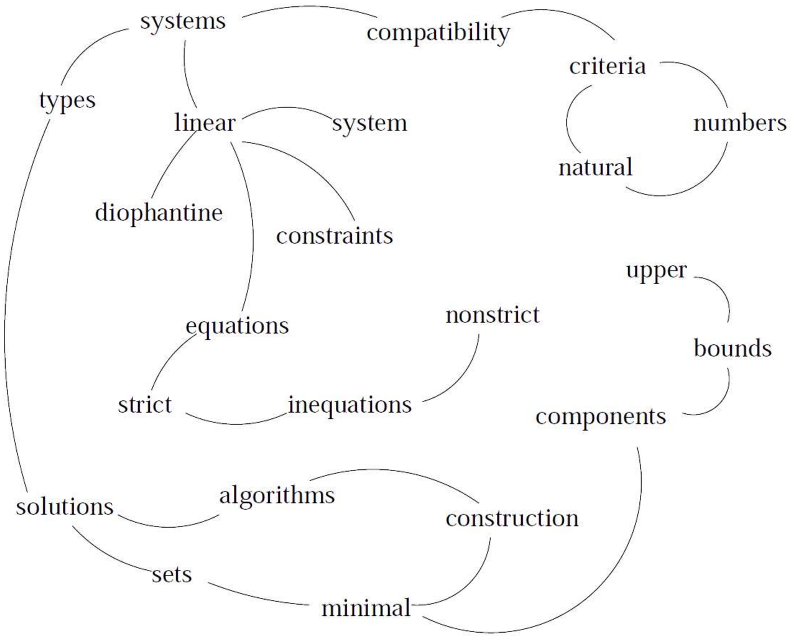
{convertible} {SUV} {mountain bike}

SEMANTIC VISUALIZATIONS

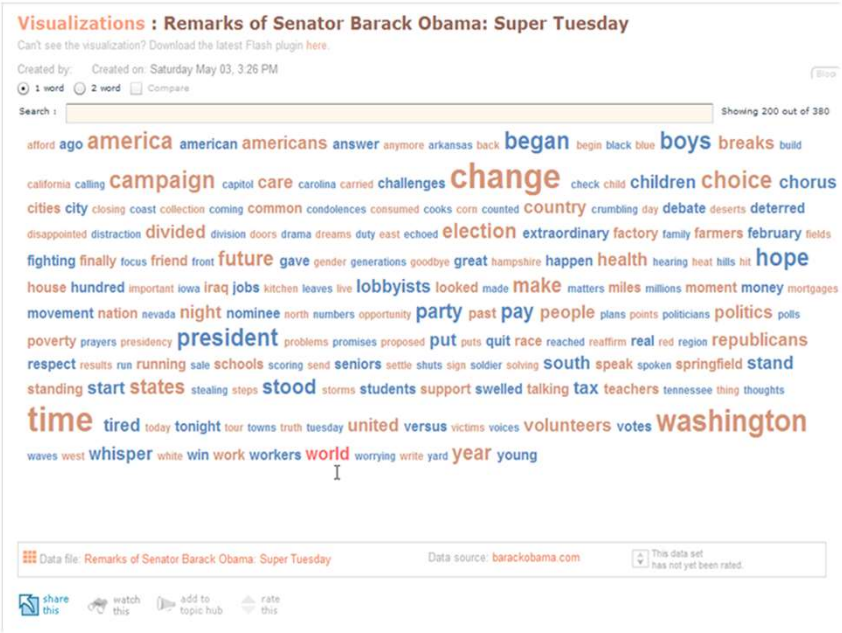
DocuBurst



Collins, C.; Carpendale, S.; Penn, G. DocuBurst: Visualizing Document Content using Language Structure. Proceedings of Eurographics/IEEE VGTC Symposium on Visualization, June, 2009.



Mihalcea and Tarau, 2004



Visualizations : Remarks of Senator Barack Obama: Super Tuesday

Can't see the visualization? Download the latest Flash plugin here.

Created by: Created on: Saturday May 03, 3:26 PM

● 1 word ○ 2 word □ Compare

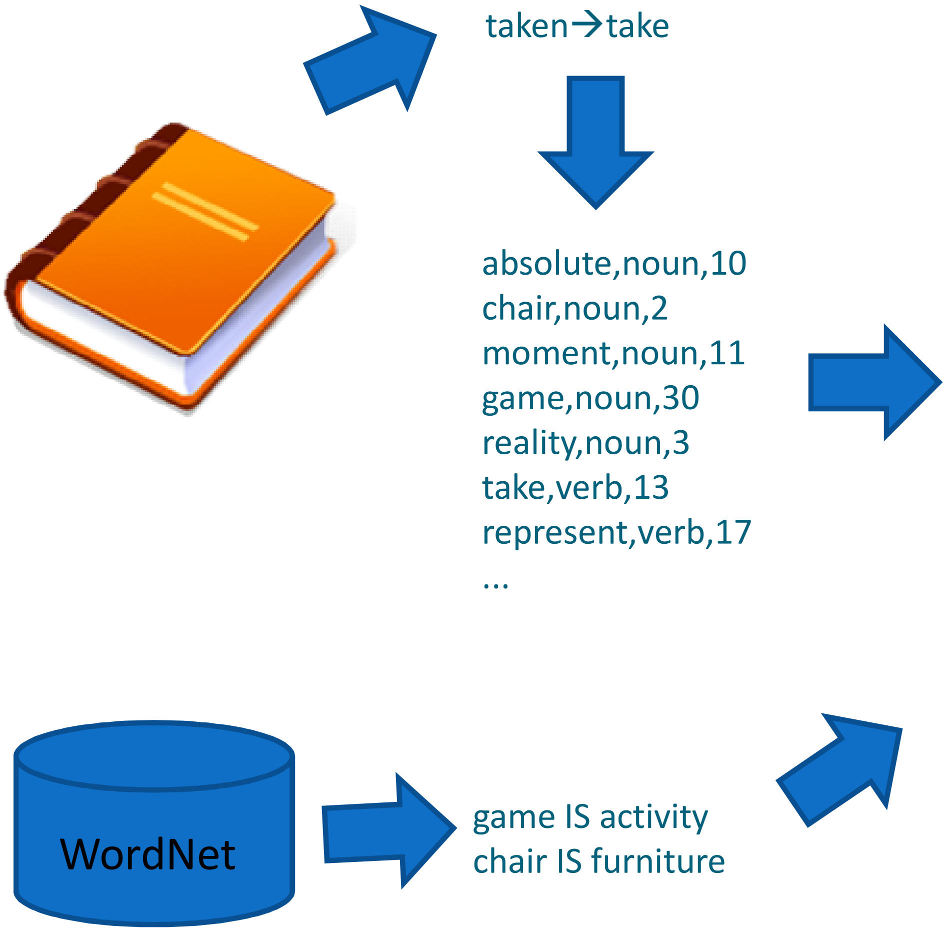
Search : Showing 200 out of 380

afford ago america american americans answer anymore arkansas back began begin black blue boys breaks build
california calling campaign capitol care carolina carried challenges change check child children choice chorus
cities city closing coast collection coming common condolences consumed cooks corn counted country crumbling day debate deserts deterred
disappointed distraction divided division doors drama dreams duty east echoed election extraordinary factory family farmers february fields
fighting finally focus friend front future gave gender generations goodbye great hampshire happen health hearing heat hills hit hope
house hundred important iowa iraq jobs kitchen leaves live lobbyists looked made make matters miles millions moment money mortgages
movement nation nevada night nominee north numbers opportunity party past pay people plans points politicians politics polls
poverty prayers presidency president problems promises proposed put puts quit race reached reaffirm real red region republicans
respect results run running sale schools scoring send seniors settle shuts sign soldier solving south speak spoken springfield stand
standing start states stealing steps stood storms students support swelled talking tax teachers tennessee thing thoughts
time tired today tonight tour towns truth tuesday united versus victims voices volunteers votes washington
waves west whisper white win work workers world worrying write yard year young

Data file: Remarks of Senator Barack Obama: Super Tuesday Data source: barackobama.com This data set has not yet been rated.

Wattenberg et al., 2008

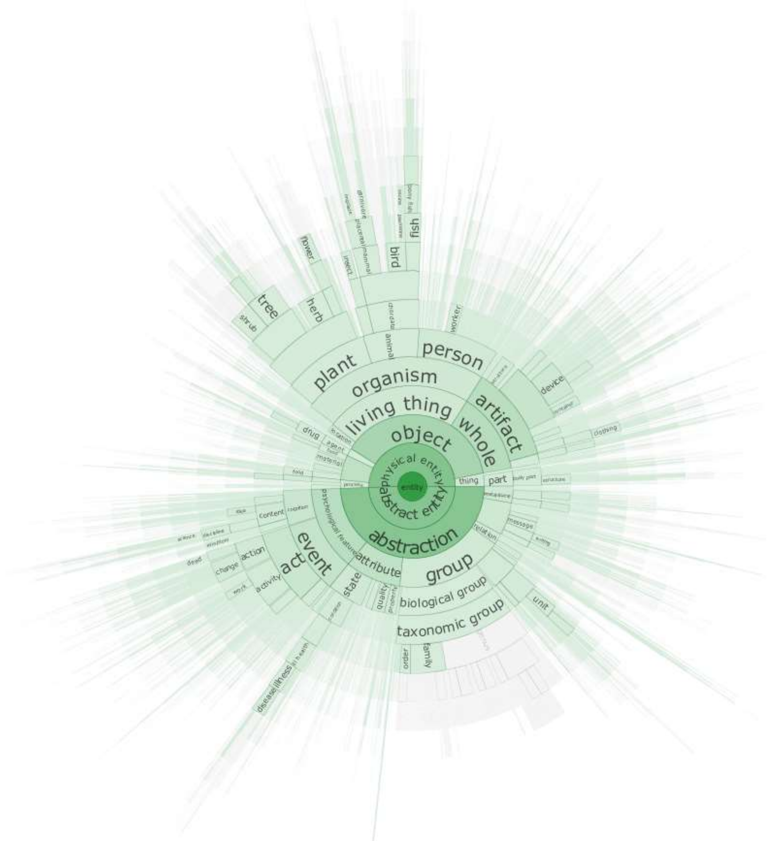
DocuBurst

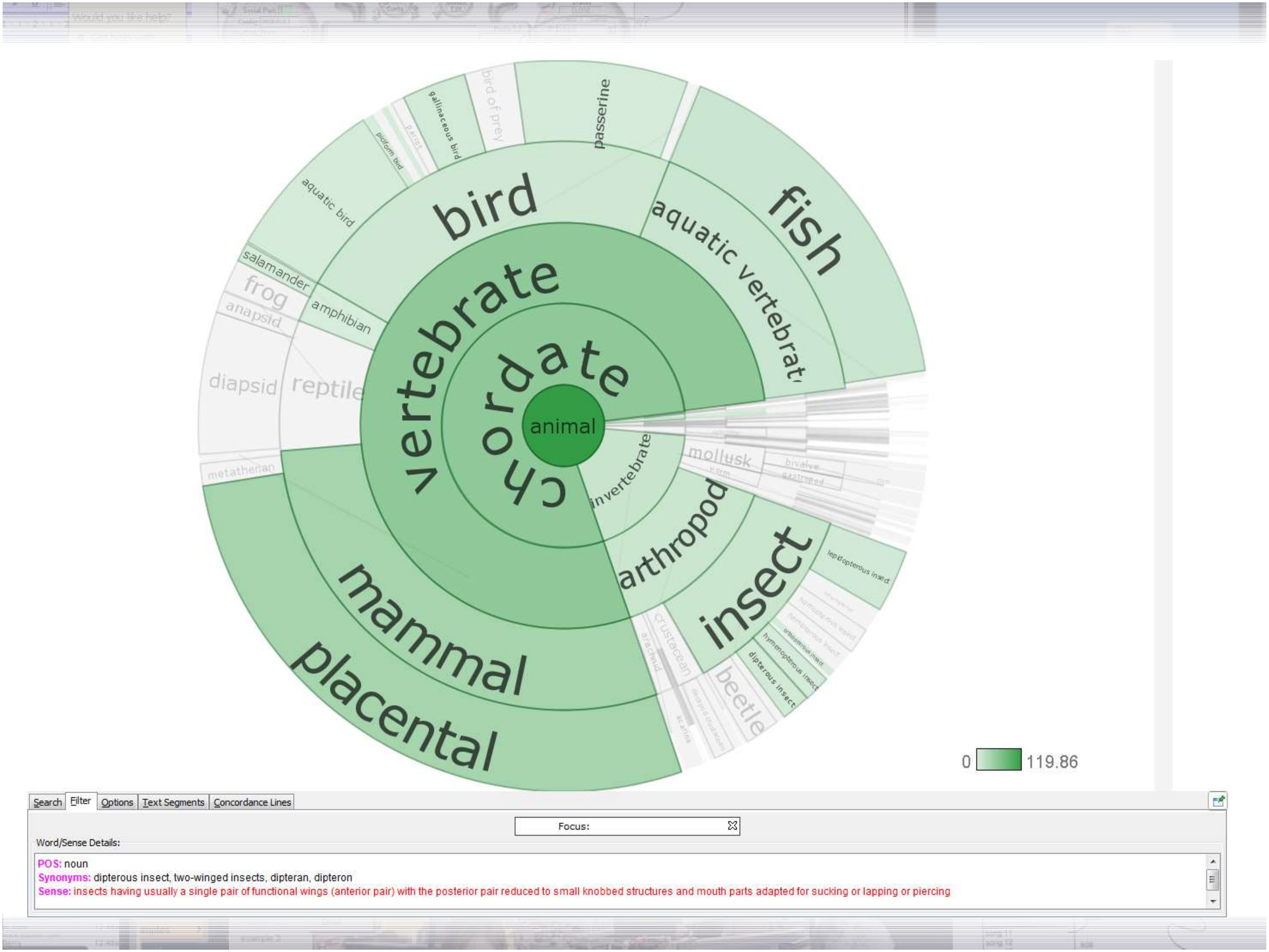


games→game
taken→take

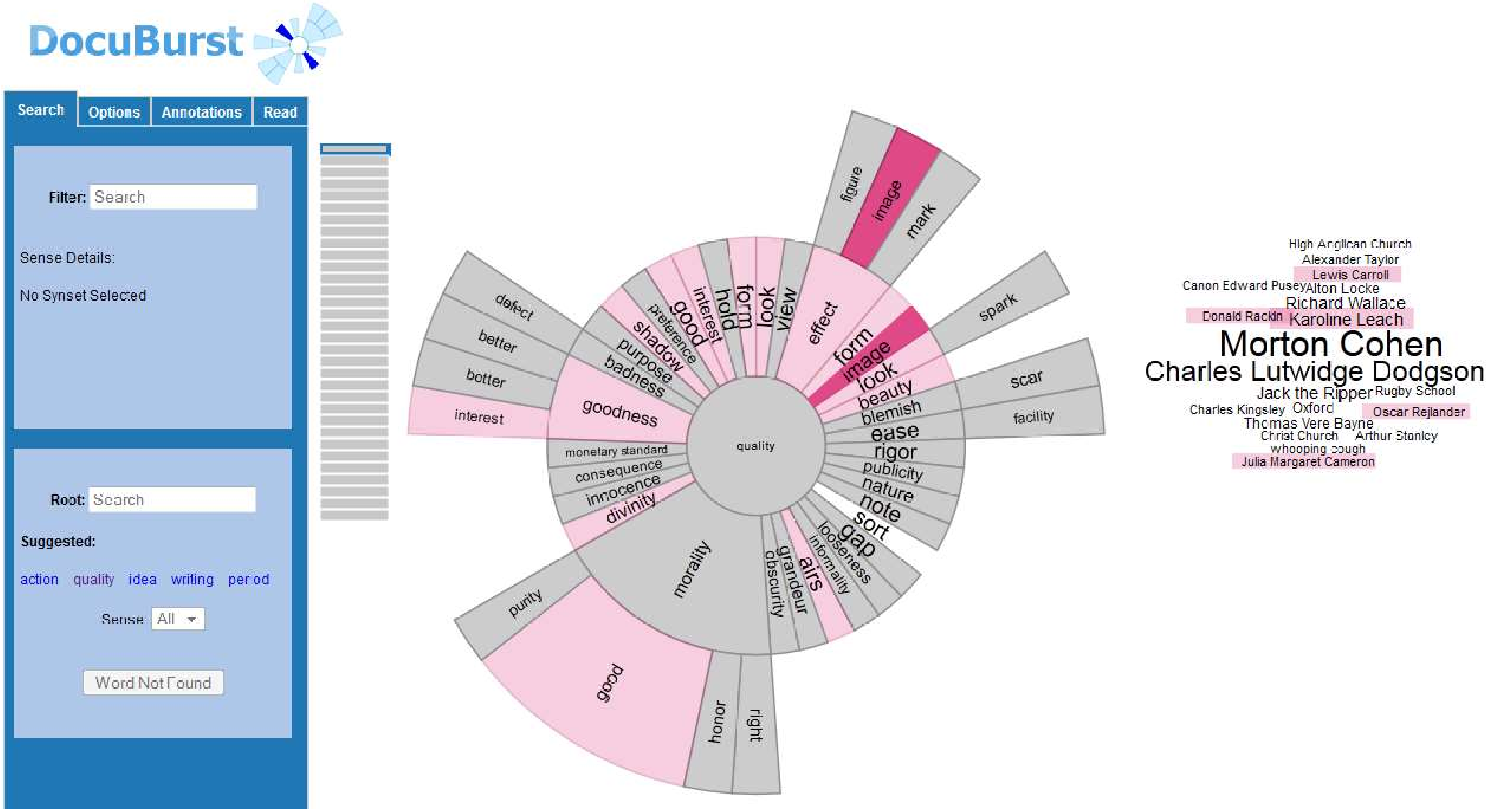
absolute,noun,10
chair,noun,2
moment,noun,11
game,noun,30
reality,noun,3
take,verb,13
represent,verb,17
...

game IS activity
chair IS furniture





POS: noun
Synonyms: dipterous insect, two-winged insects, dipteran, dipteron
Sense: insects having usually a single pair of functional wings (anterior pair) with the posterior pair reduced to small knobbed structures and mouth parts adapted for sucking or lapping or piercing



Filter: Search

Root: Search

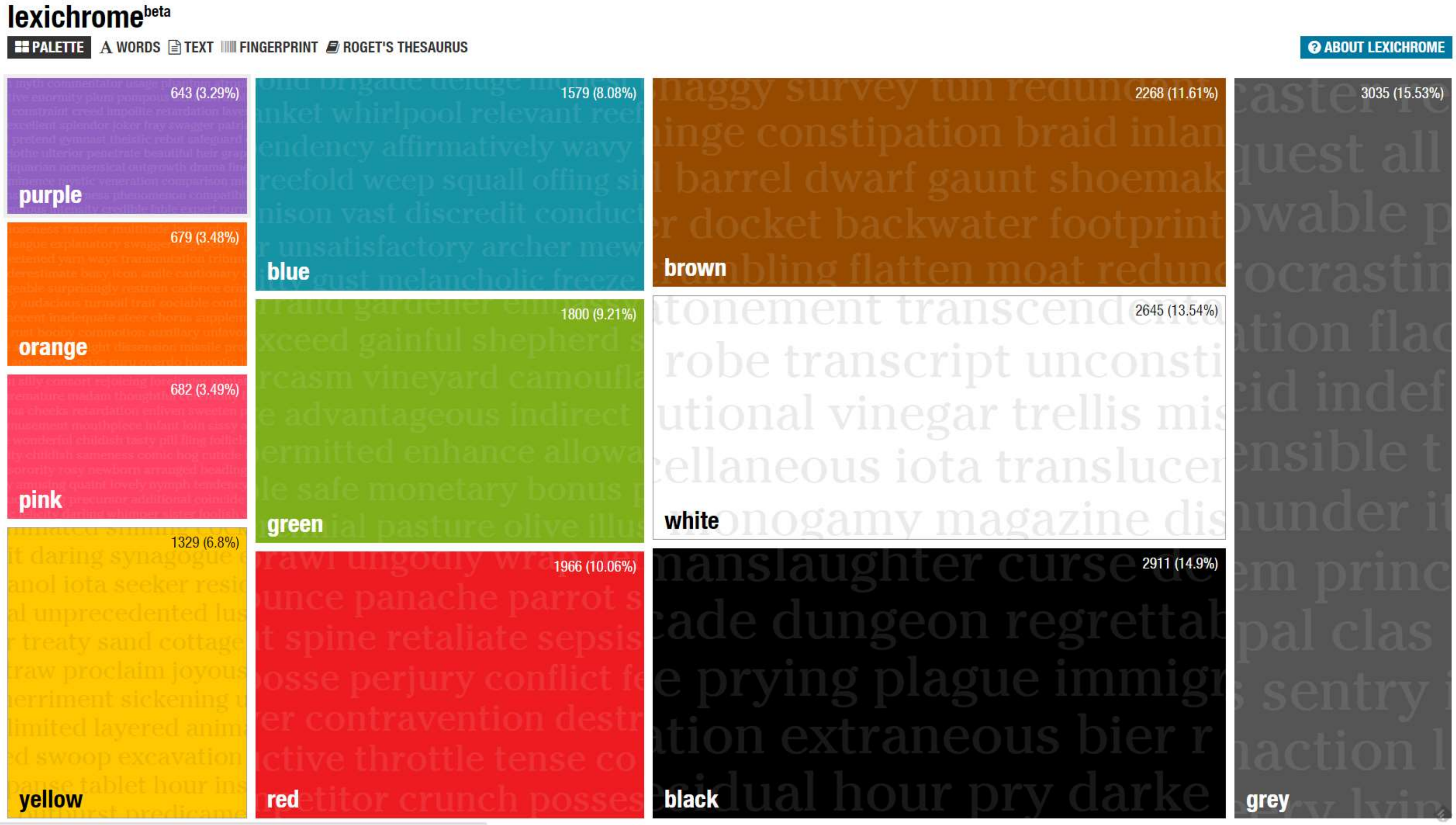
Suggested:

action quality idea writing period

Sense: All

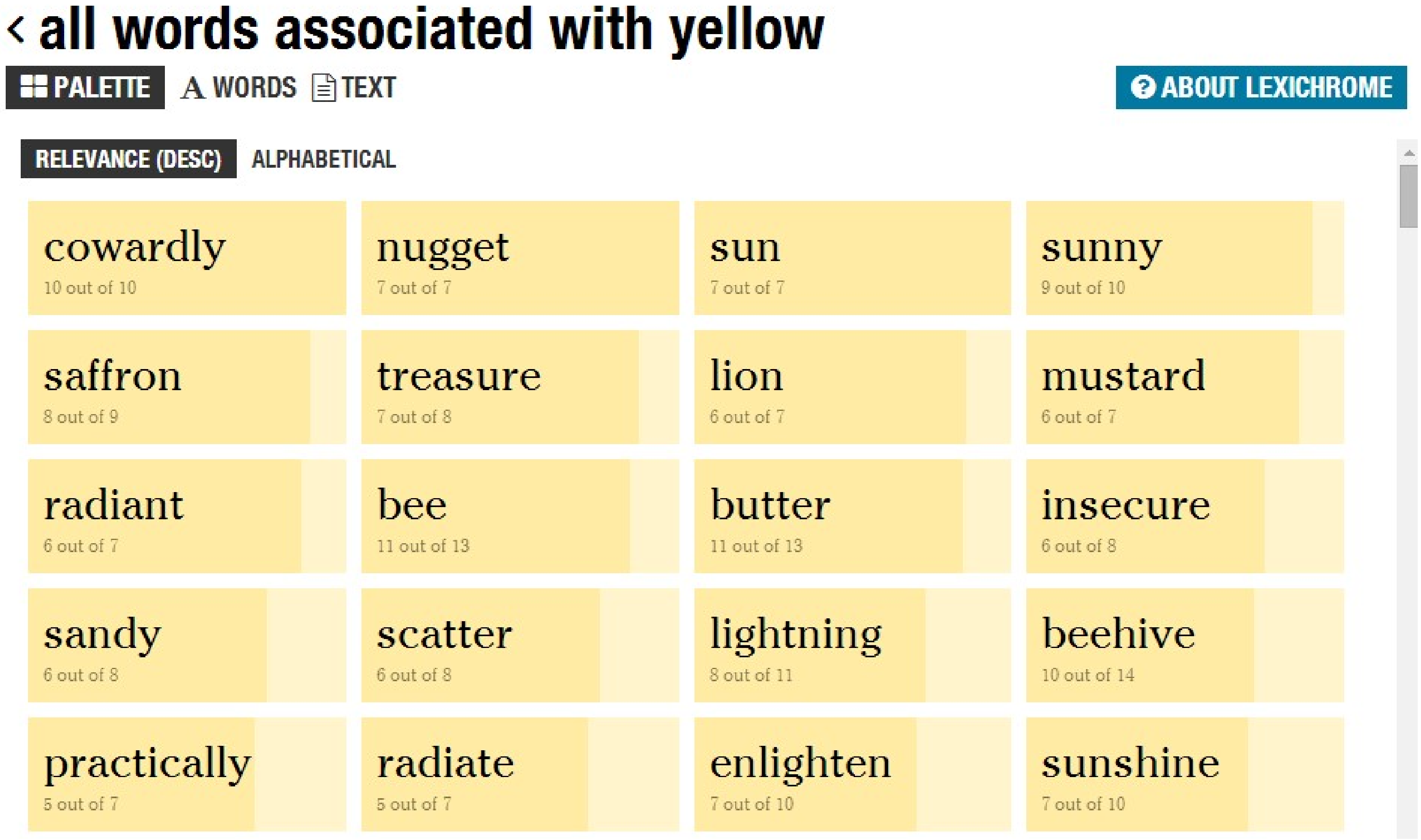
Try it! http://vialab.science.uoit.ca/docuburst

Lexichrome



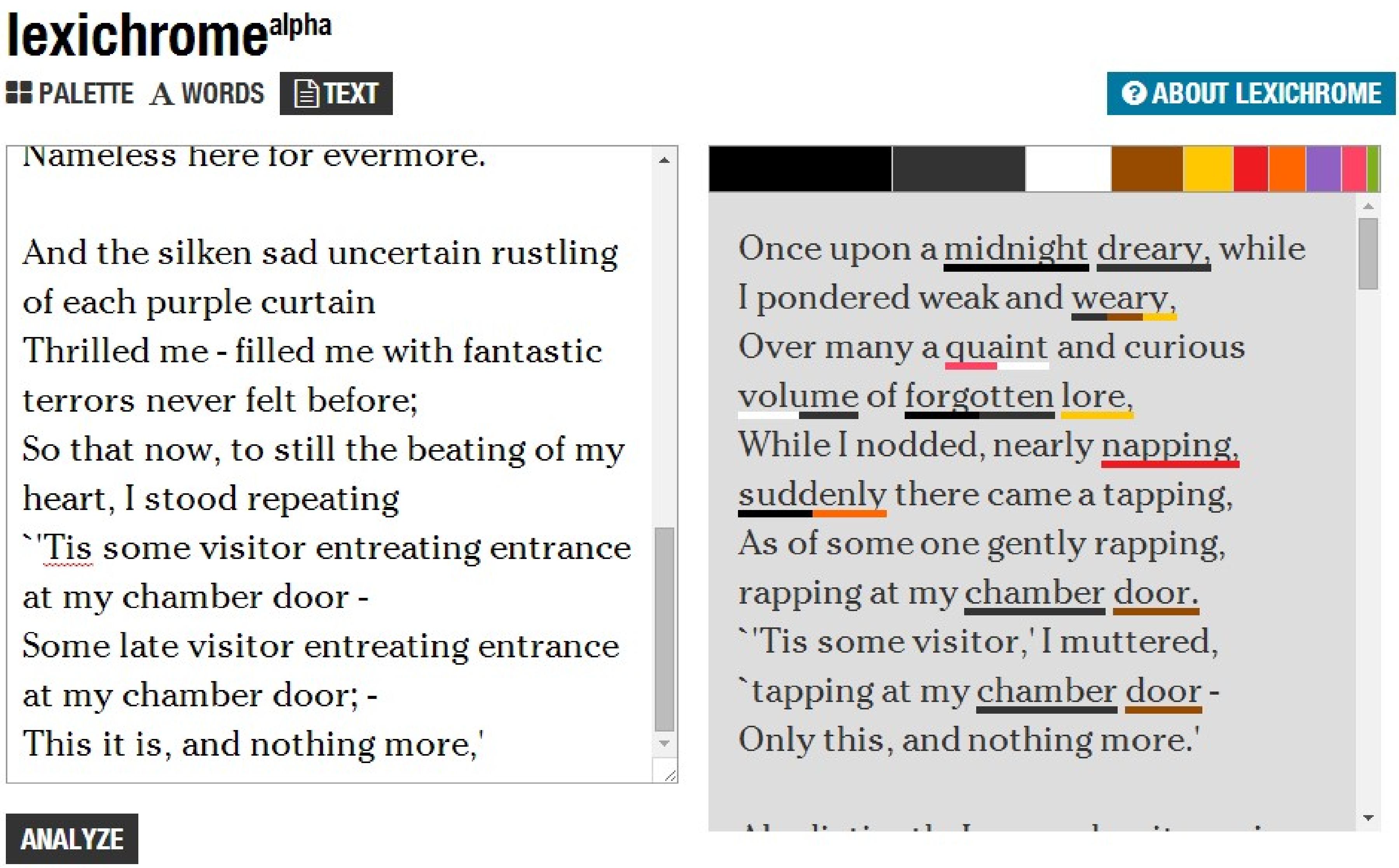
http://lexichrome.com

Work in Progress with Chris Kim and Saif Mohammed



‹ all words associated with yellow

cowardly 10 out of 10	nugget 7 out of 7	sun 7 out of 7	sunny 9 out of 10
saffron 8 out of 9	treasure 7 out of 8	lion 6 out of 7	mustard 6 out of 7
radiant 6 out of 7	bee 11 out of 13	butter 11 out of 13	insecure 6 out of 8
sandy 6 out of 8	scatter 6 out of 8	lightning 8 out of 11	beehive 10 out of 14
practically 5 out of 7	radiate 5 out of 7	enlighten 7 out of 10	sunshine 7 out of 10

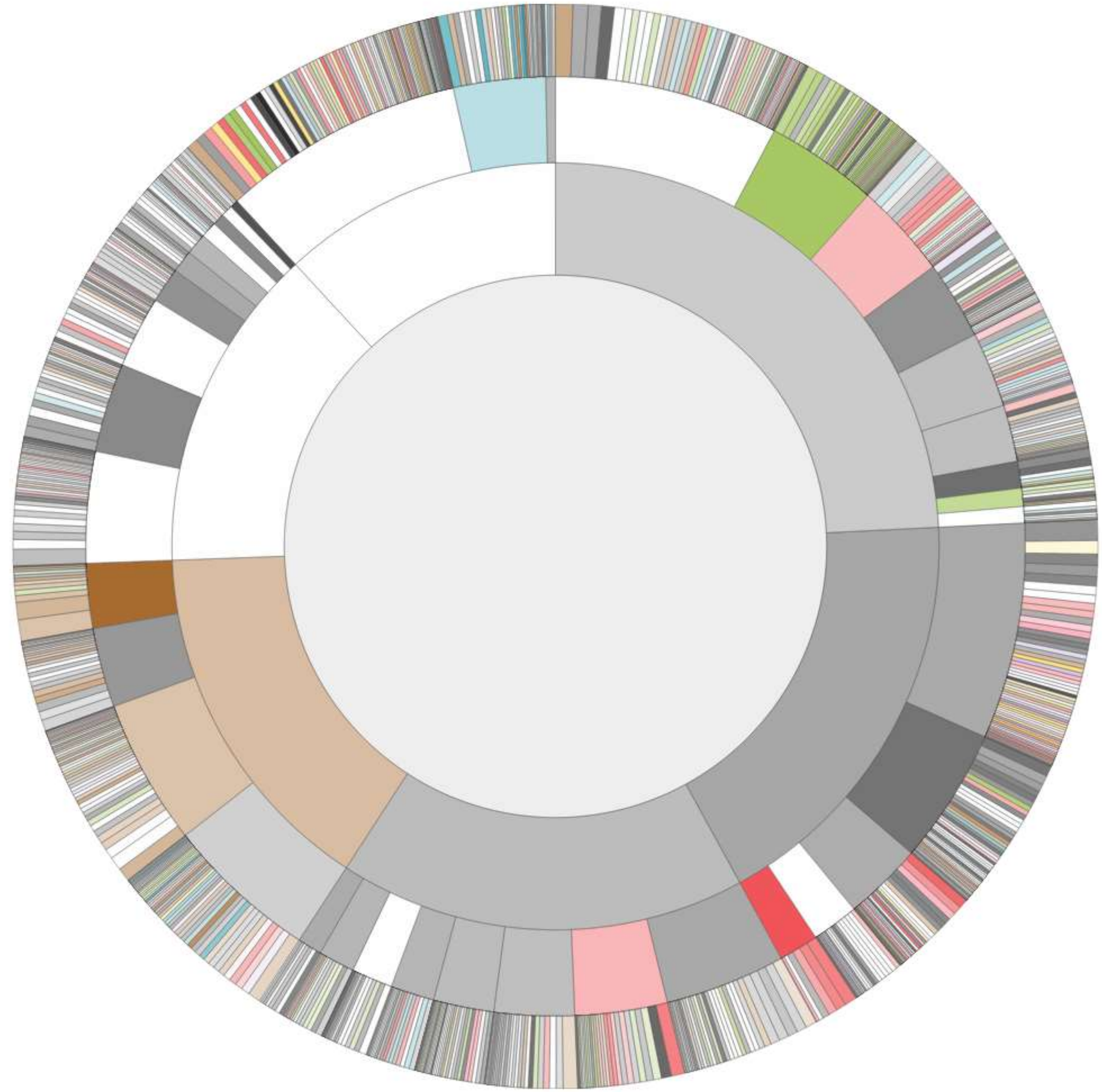


Nameless here for evermore.

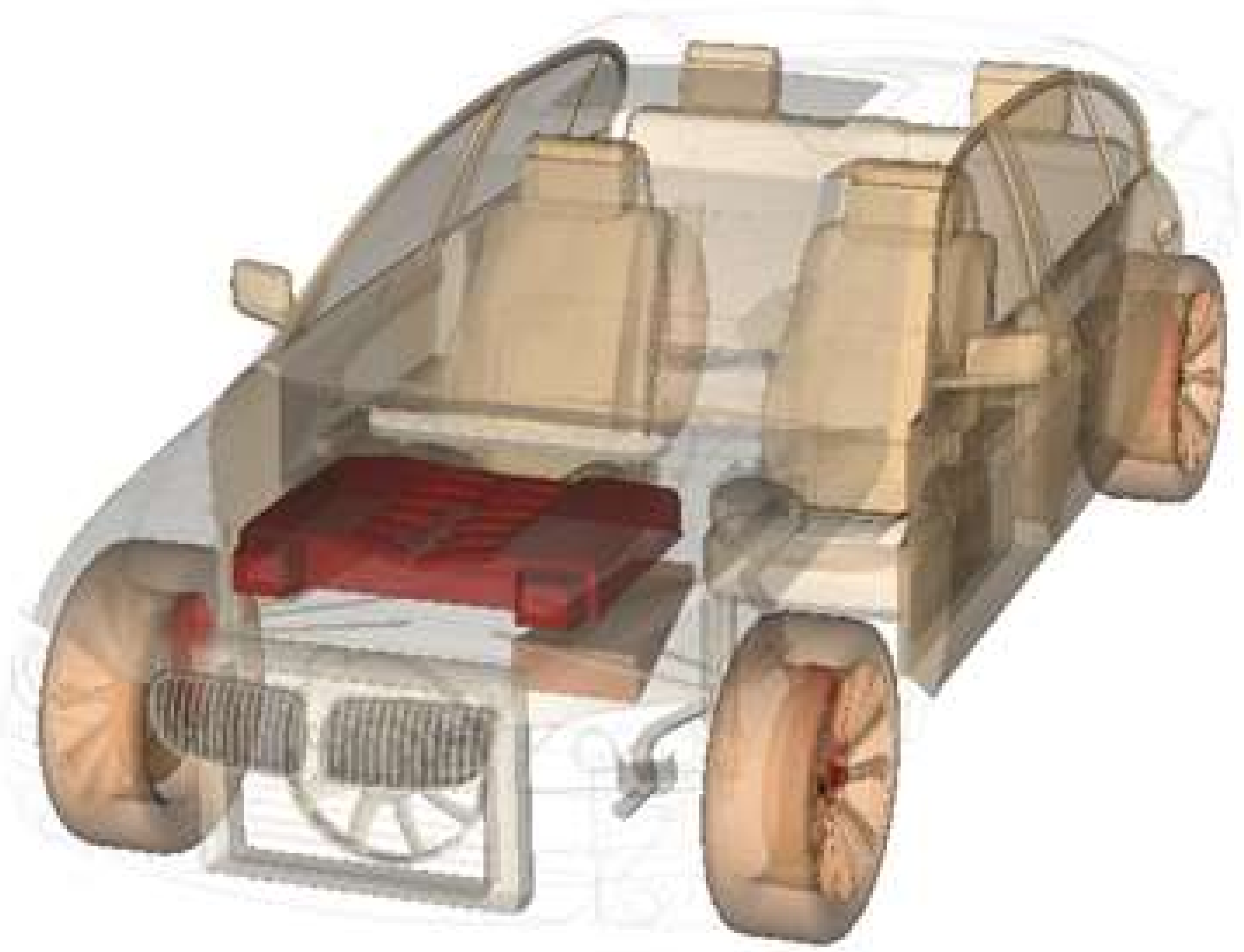
And the silken sad uncertain rustling of each purple curtain
Thrilled me - filled me with fantastic terrors never felt before;
So that now, to still the beating of my heart, I stood repeating
`'Tis some visitor entreating entrance at my chamber door -
Some late visitor entreating entrance at my chamber door; -
This it is, and nothing more,'

Once upon a midnight dreary, while I pondered weak and weary,
Over many a quaint and curious volume of forgotten lore,
While I nodded, nearly napping, suddenly there came a tapping,
As of some one gently rapping, rapping at my chamber door.
`'Tis some visitor,' I muttered, `tapping at my chamber door -
Only this, and nothing more.'

ROGET'S THESAURUS

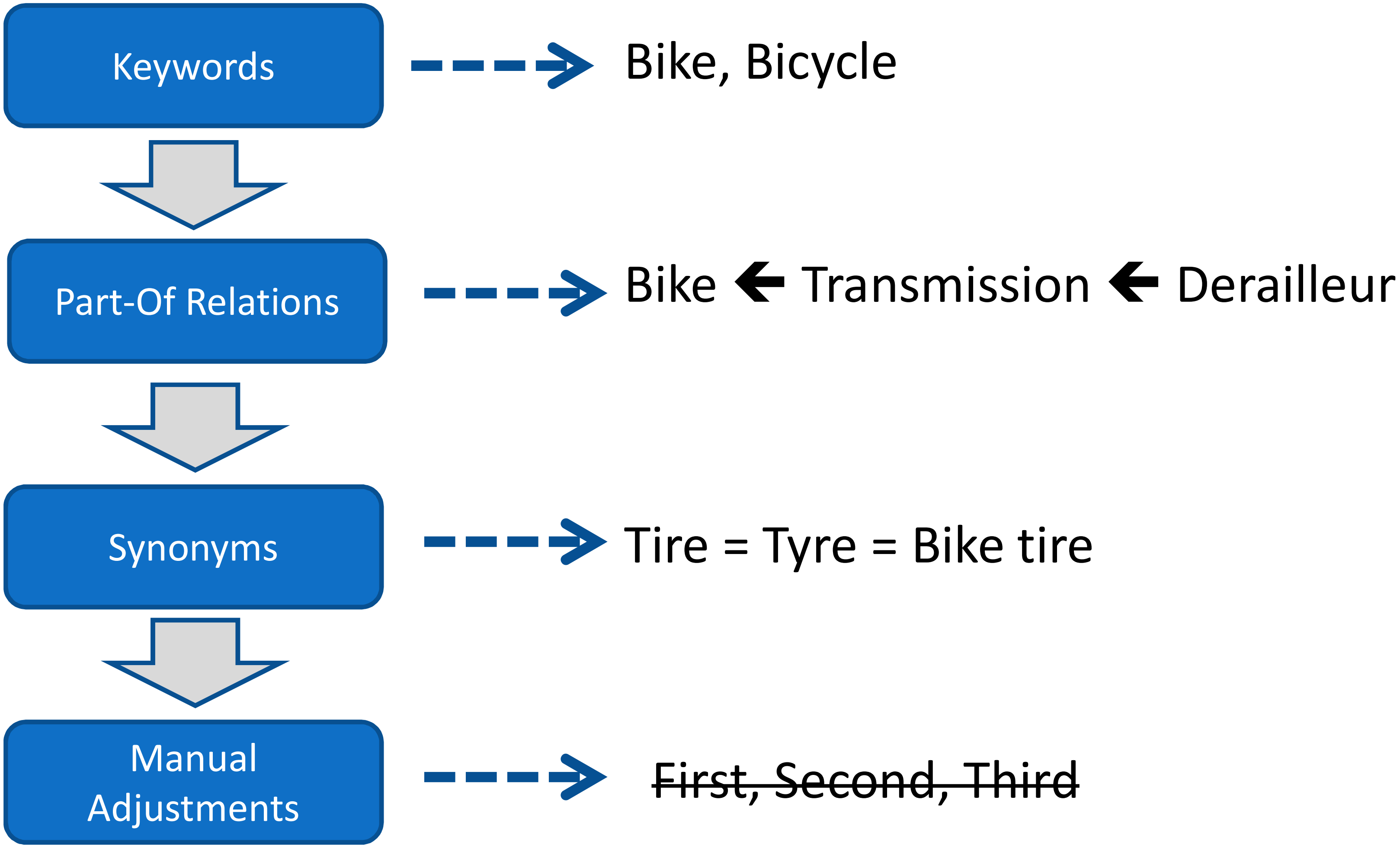


Descriptive Non-Photorealistic Rendering



M. Chang and C. Collins, “Exploring Entities in Text with Descriptive Non-photorealistic Rendering,” in *Proc. of the 2013 IEEE Pacific Visualization Symposium (PACIFICVIS ’13)*, 2013.

Ontology Generation



Keywords → Bike, Bicycle

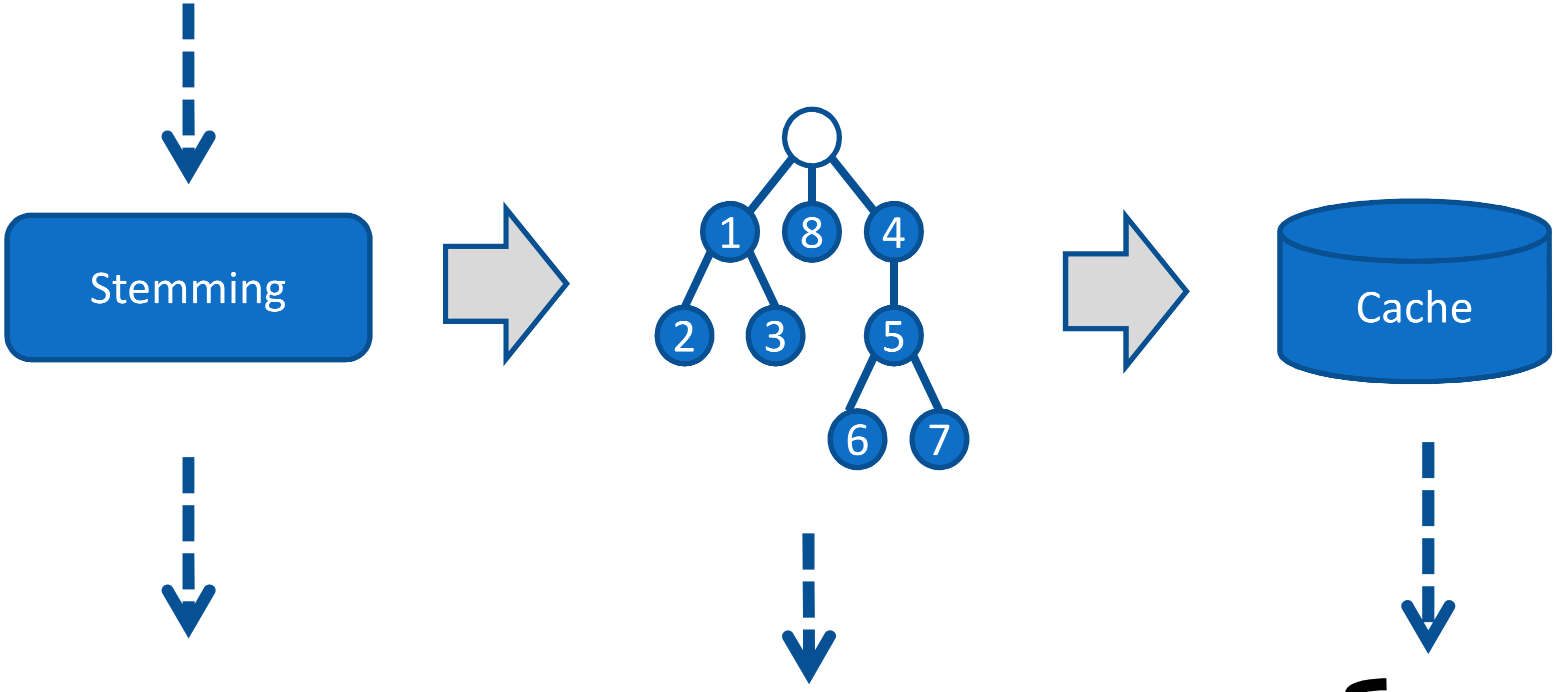
Part-Of Relations → Bike ← Transmission ← Derailleur

Synonyms → Tire = Tyre = Bike tire

Manual Adjustments → ~~First, Second, Third~~

Entity Extraction

“**Brakes** failed going at 35 mph.”

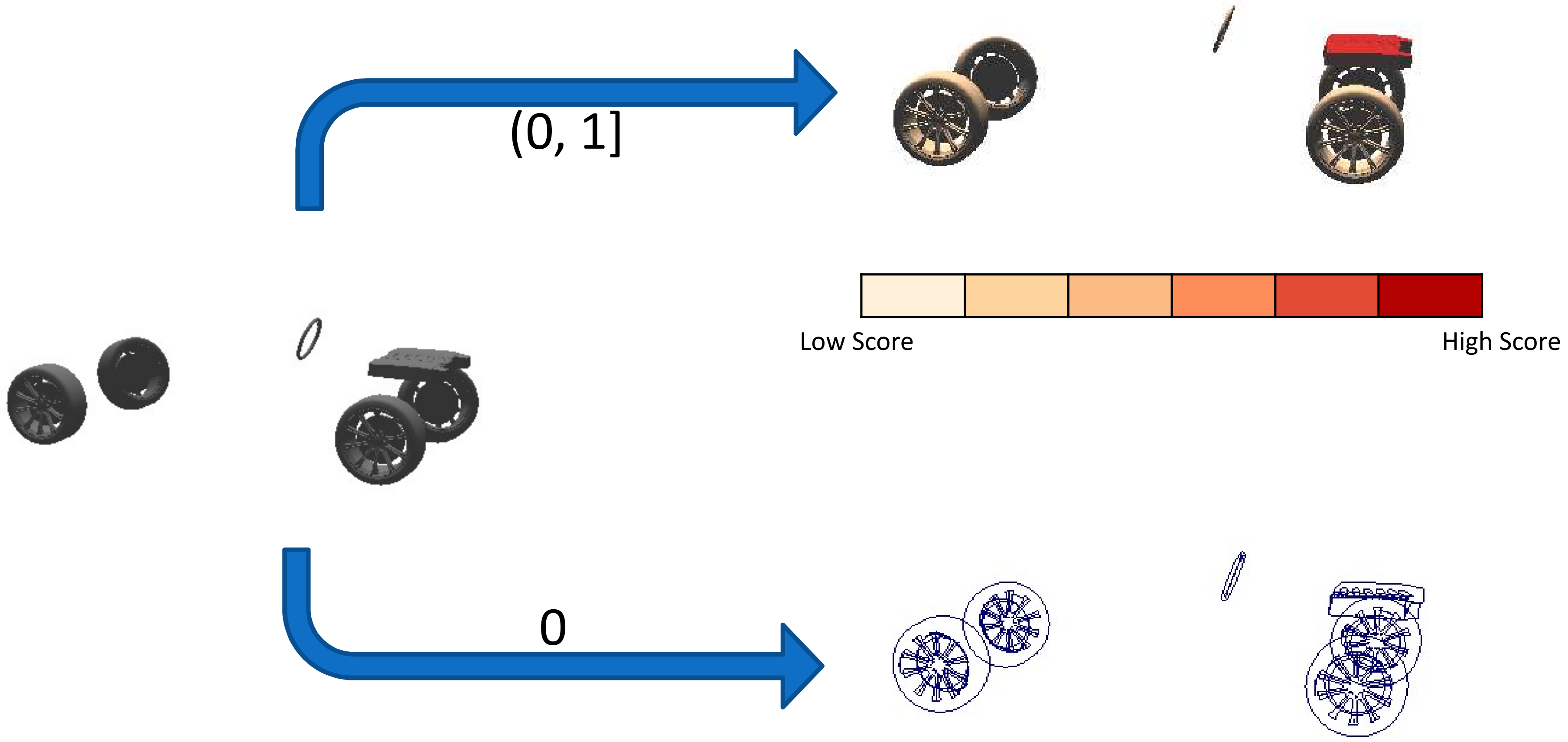


“brak”

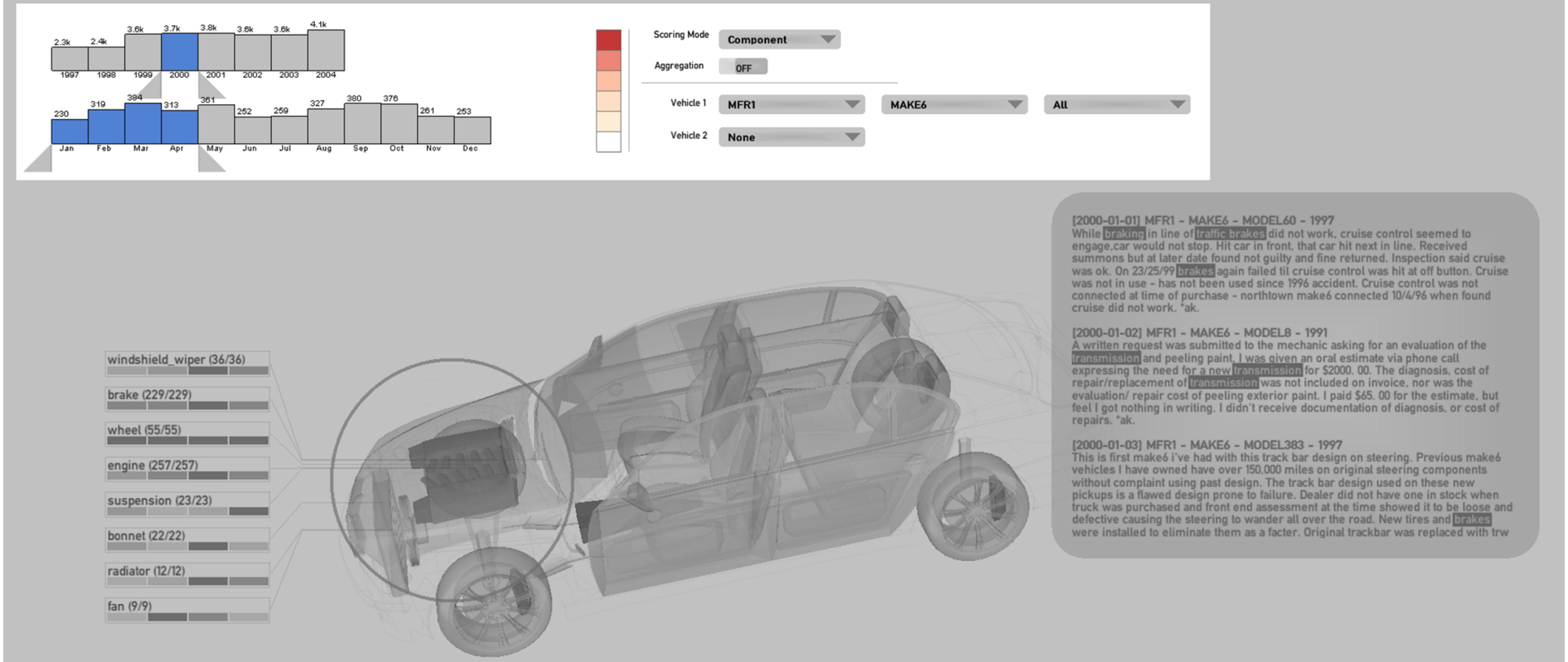
6: {brak, brak system…}

Document: 123
Entity: 6
Location: 0

Visual Representation



Main Interface



Scoring Mode: Component

Aggregation: OFF

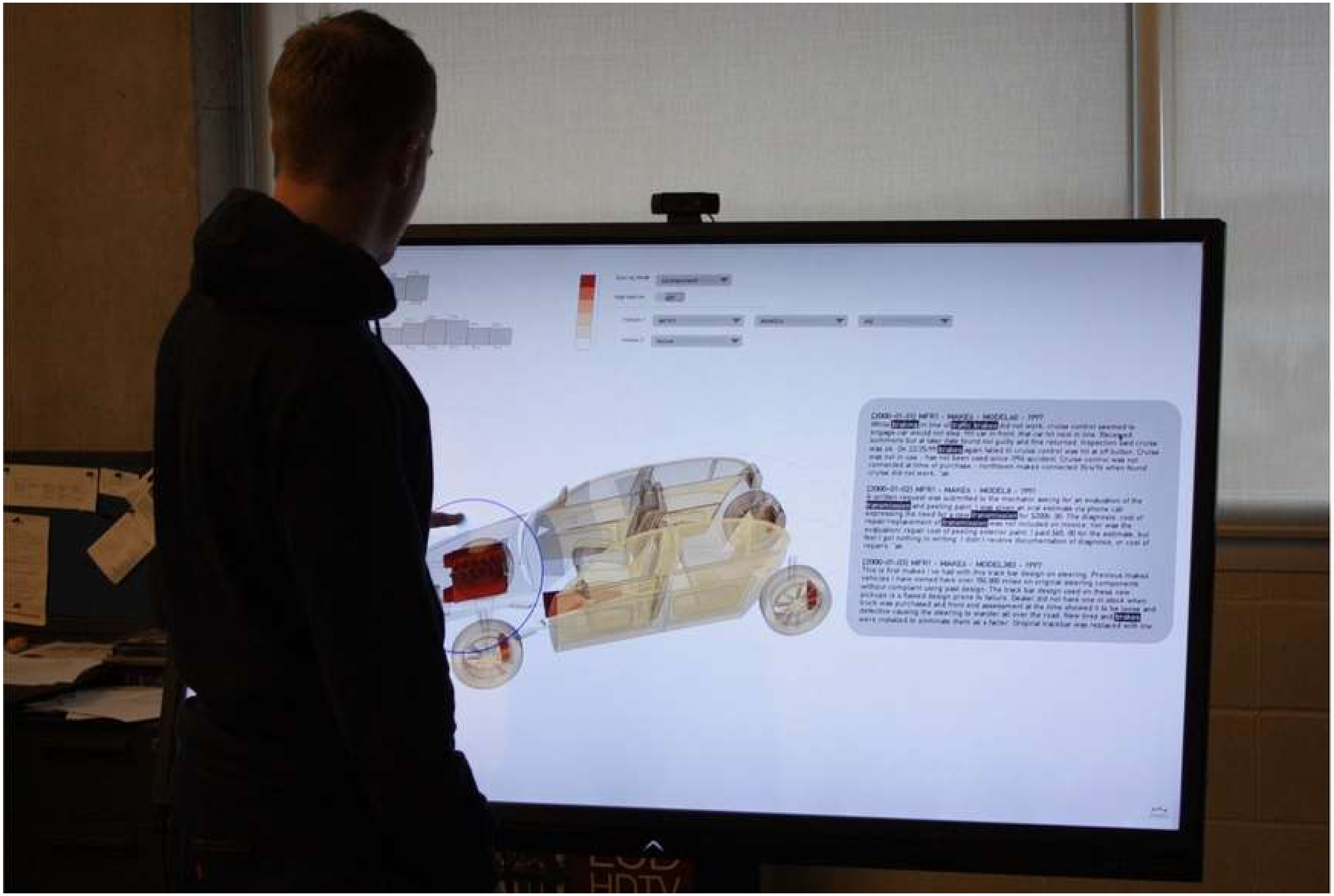
Vehicle 1: MFR1 | MAKE6 | All

Vehicle 2: None

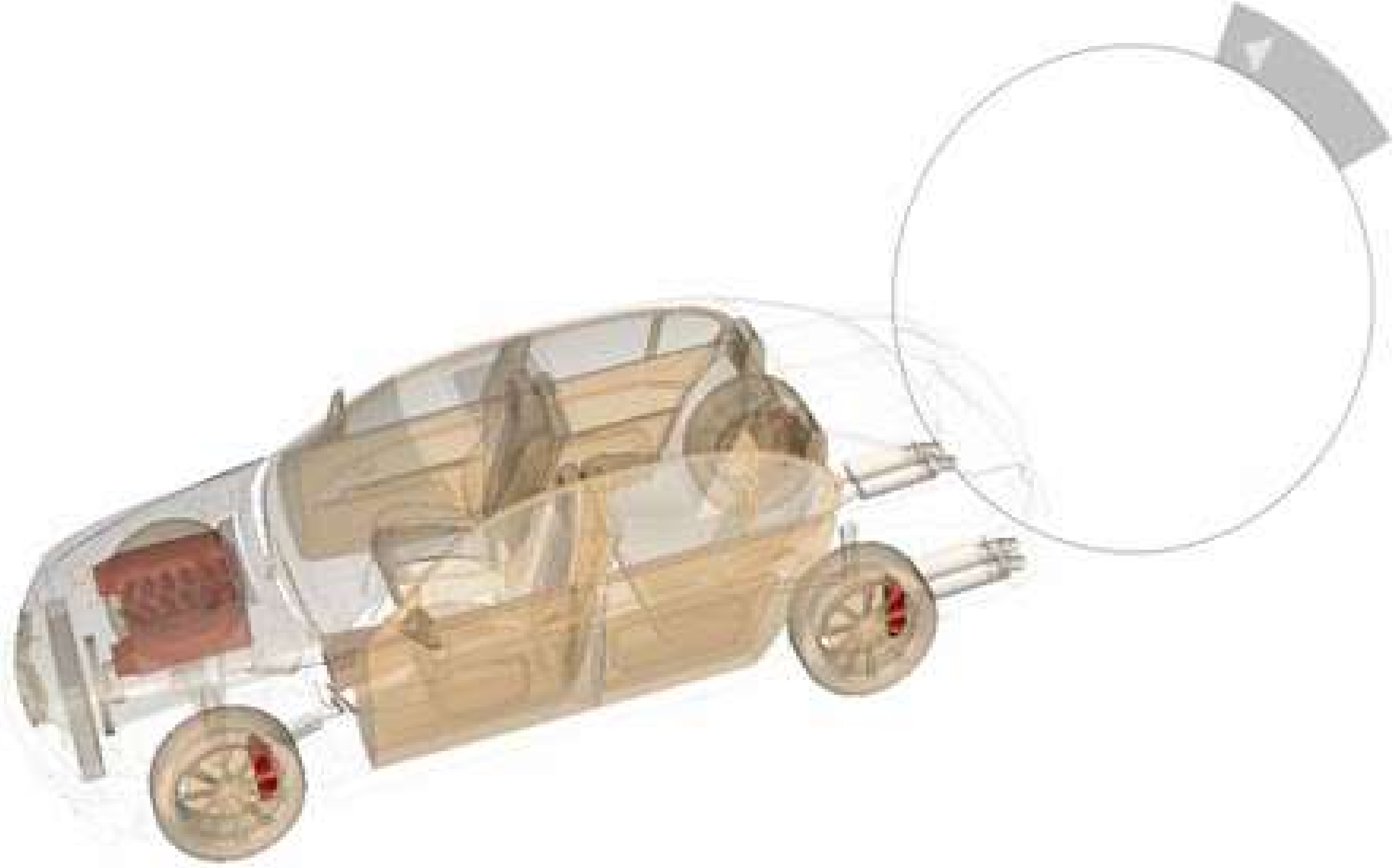
[2000-01-01] MFR1 - MAKE6 - MODEL60 - 1997
While braking in line of traffic brakes did not work, cruise control seemed to engage,car would not stop. Hit car in front, that car hit next in line. Received summons but at later date found not guilty and fine returned. Inspection said cruise was ok. On 23/25/99 brakes again failed til cruise control was hit at off button. Cruise was not in use - has not been used since 1996 accident. Cruise control was not connected at time of purchase - northtown make6 connected 10/4/96 when found cruise did not work. *ak.

[2000-01-02] MFR1 - MAKE6 - MODEL8 - 1991
A written request was submitted to the mechanic asking for an evaluation of the transmission and peeling paint. I was given an oral estimate via phone call expressing the need for a new transmission for $2000. 00. The diagnosis, cost of repair/replacement of transmission was not included on invoice, nor was the evaluation/ repair cost of peeling exterior paint. I paid $65. 00 for the estimate, but feel I got nothing in writing. I didn't receive documentation of diagnosis, or cost of repairs. *ak.

[2000-01-03] MFR1 - MAKE6 - MODEL383 - 1997
This is first make6 i've had with this track bar design on steering. Previous make6 vehicles I have owned have over 150,000 miles on original steering components without complaint using past design. The track bar design used on these new pickups is a flawed design prone to failure. Dealer did not have one in stock when truck was purchased and front end assessment at the time showed it to be loose and defective causing the steering to wander all over the road. New tires and brakes were installed to eliminate them as a facter. Original trackbar was replaced with trw



Exploration with Lens

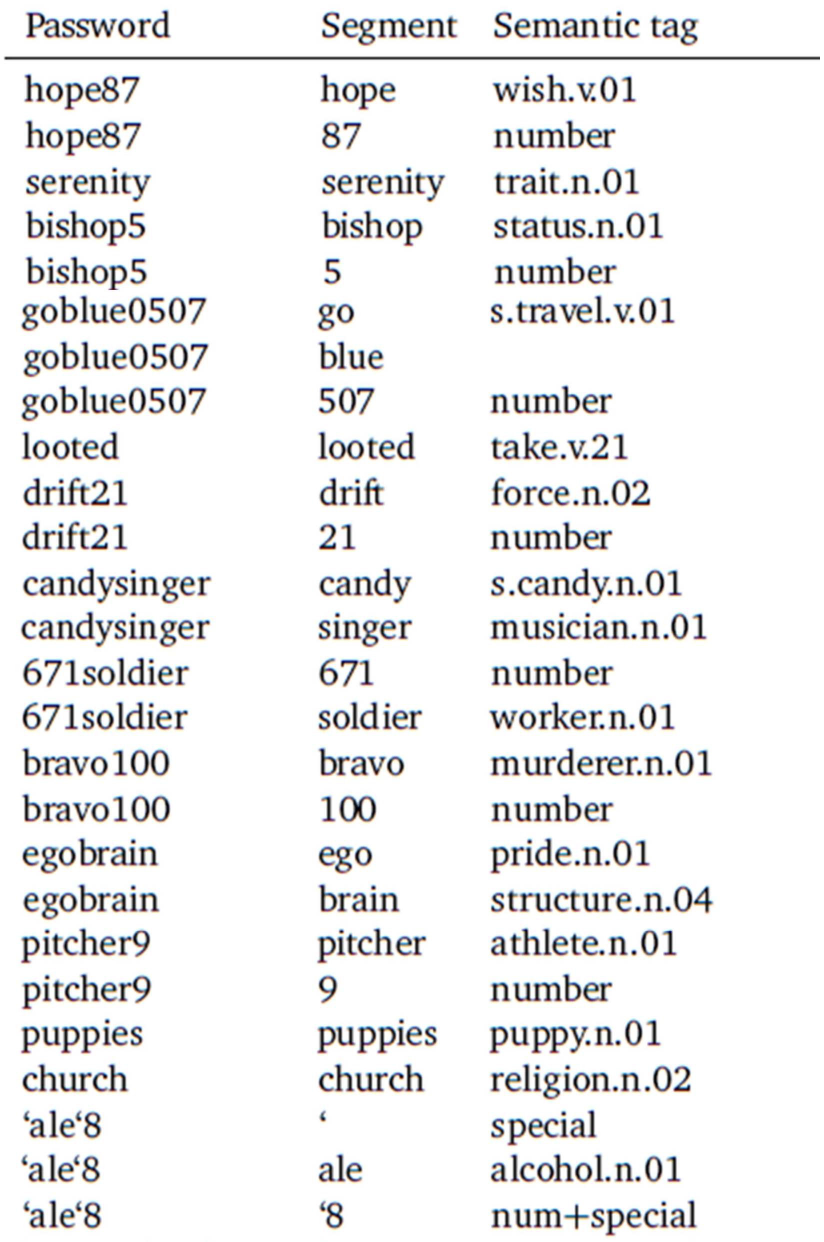


Semantic Password Analysis

- What types of words do people use in their passwords?
- Do the patterns of word use represent security vulnerabilities?

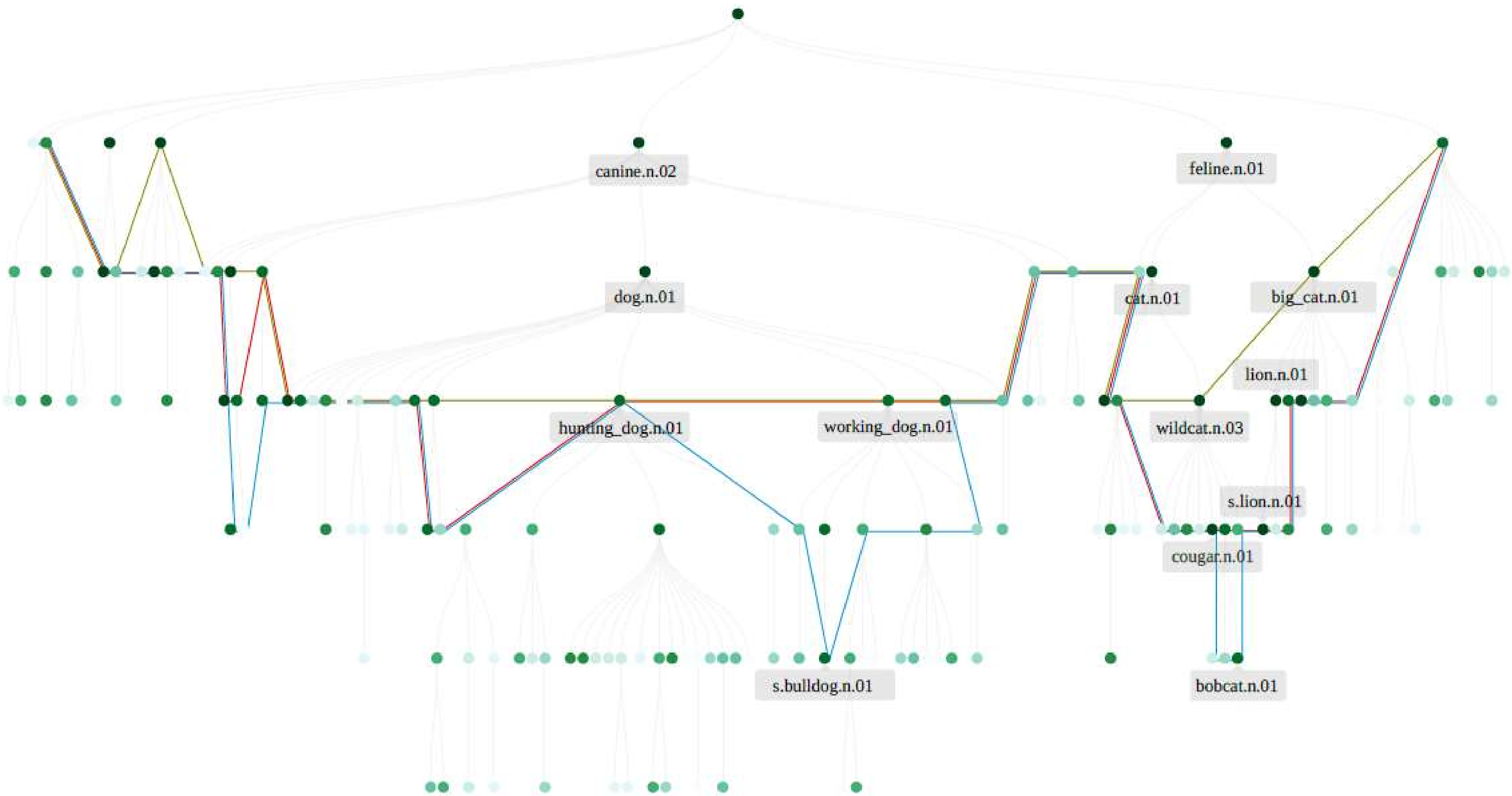
R. Veras, C. Collins, and J. Thorpe, "On Semantic Patterns of Passwords and their Security Impact," *In Proceeding of the Network and Distributed System Security Symposium (NDSS'14)*, 2014.

- Extract words from 32 million passwords
- Categorize them
- Parse the results to find structure
- Create a password guessing system based on the model



Password	Segment	Semantic tag
hope87	hope	wish.v.01
hope87	87	number
serenity	serenity	trait.n.01
bishop5	bishop	status.n.01
bishop5	5	number
goblue0507	go	s.travel.v.01
goblue0507	blue	
goblue0507	507	number
looted	looted	take.v.21
drift21	drift	force.n.02
drift21	21	number
candysinger	candy	s.candy.n.01
candysinger	singer	musician.n.01
671soldier	671	number
671soldier	soldier	worker.n.01
bravo100	bravo	murderer.n.01
bravo100	100	number
egobrain	ego	pride.n.01
egobrain	brain	structure.n.04
pitcher9	pitcher	athlete.n.01
pitcher9	9	number
puppies	puppies	puppy.n.01
church	church	religion.n.02
‘ale‘8	‘	special
‘ale‘8	ale	alcohol.n.01
‘ale‘8	‘8	num+special

Appropriate Levels of Detail

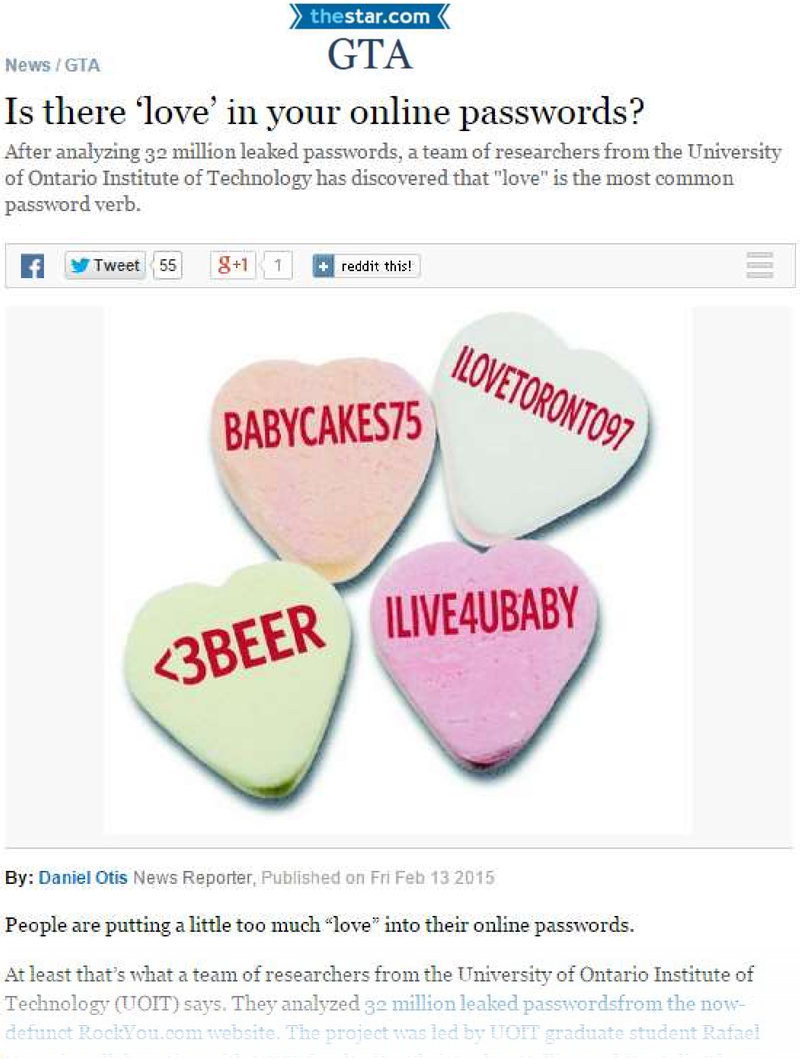


Results

- Created best cracker on several measures, including percent correct guesses
- Designing strategies to help people make passwords more *semantically* secure – keep the meaning but lower the probability

Results

- Created best cracker on measure of % correct guesses
- Place names, male names very popular
- “Cute” animals more common:
  - Monkey, dogs, cats, dolphins
- Emotional verbs like “love” are common
  - People “love” male names 4x more often than female!
- Profanity is very common



thestar.com

GTA

News / GTA

Is there 'love' in your online passwords?

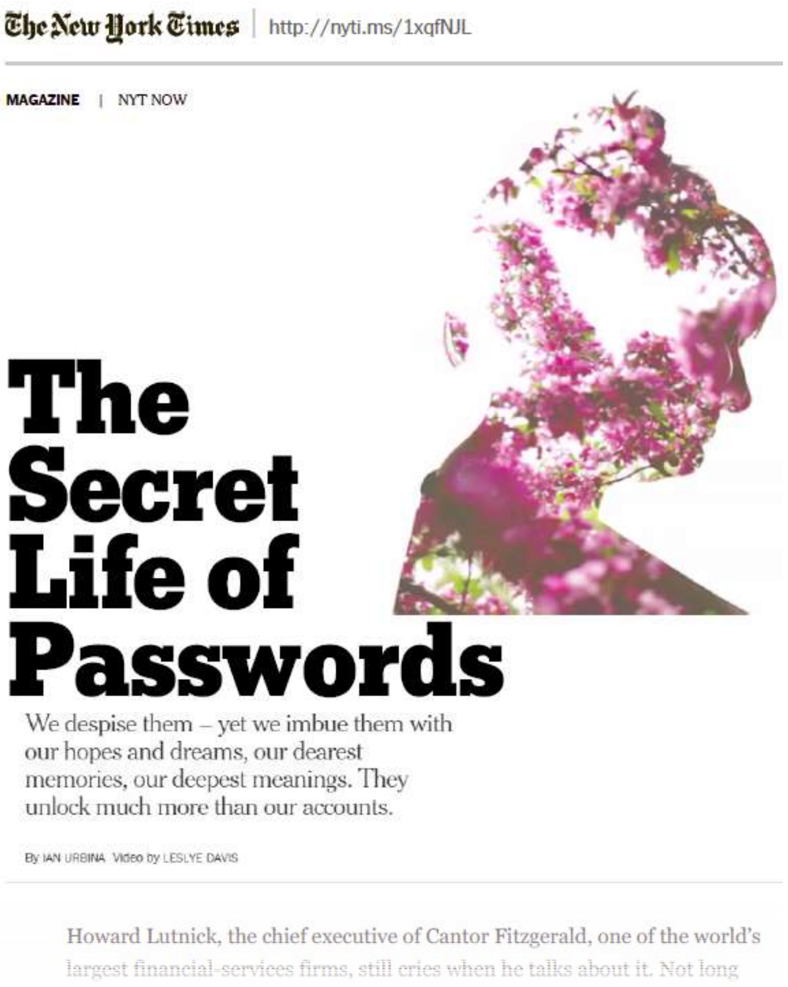
After analyzing 32 million leaked passwords, a team of researchers from the University of Ontario Institute of Technology has discovered that "love" is the most common password verb.

Tweet 55 g+1 1 reddit this!

By: Daniel Otis News Reporter, Published on Fri Feb 13 2015

People are putting a little too much "love" into their online passwords.

At least that's what a team of researchers from the University of Ontario Institute of Technology (UOIT) says. They analyzed 32 million leaked passwordsfrom the now-defunct RockYou.com website. The project was led by UOIT graduate student Rafael



The New York Times | http://nyti.ms/1xqfNJL

MAGAZINE | NYT NOW

The Secret Life of Passwords

We despise them – yet we imbue them with our hopes and dreams, our dearest memories, our deepest meanings. They unlock much more than our accounts.

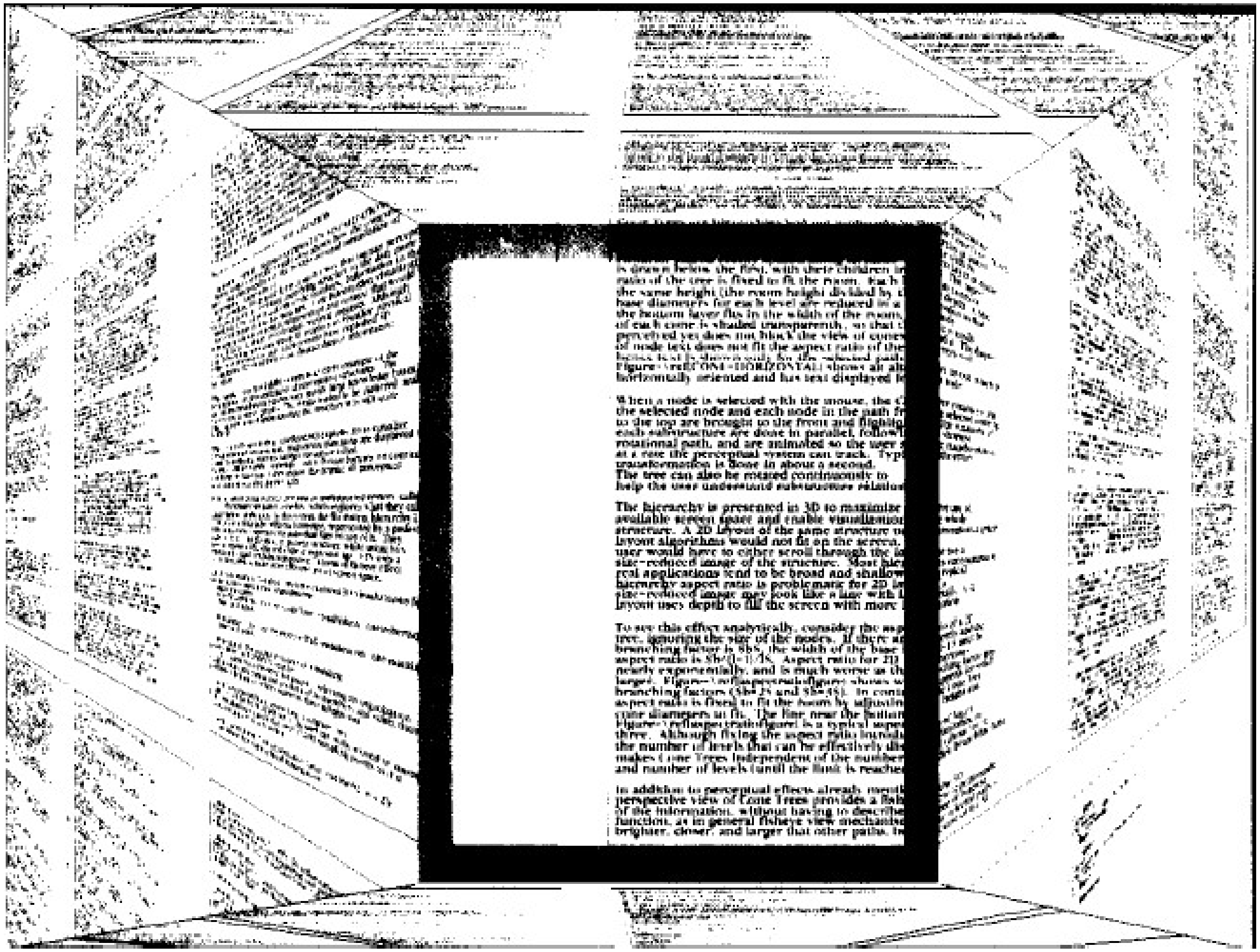
By IAN URBINA Video by LESLYE DAVIS

Howard Lutnick, the chief executive of Cantor Fitzgerald, one of the world's largest financial-services firms, still cries when he talks about it. Not long

Text Visualization

LITERARY ANALYSIS

Document Lens



is drawn below the first, with their children
ratio of the tree is fixed to fit the room. Each
the same height (the room height divided by
base diameters for each level are reduced in a
the bottom layer fits in the width of the room.
of each cone is shaded transparently, so that
perceived yet does not block the view of cones
of node text does not fit the aspect ratio of the
hence text is drawn only for the selected path.
Figure~\ref{ONE-HORIZONTAL} shows an al
horizontally oriented and has text displayed

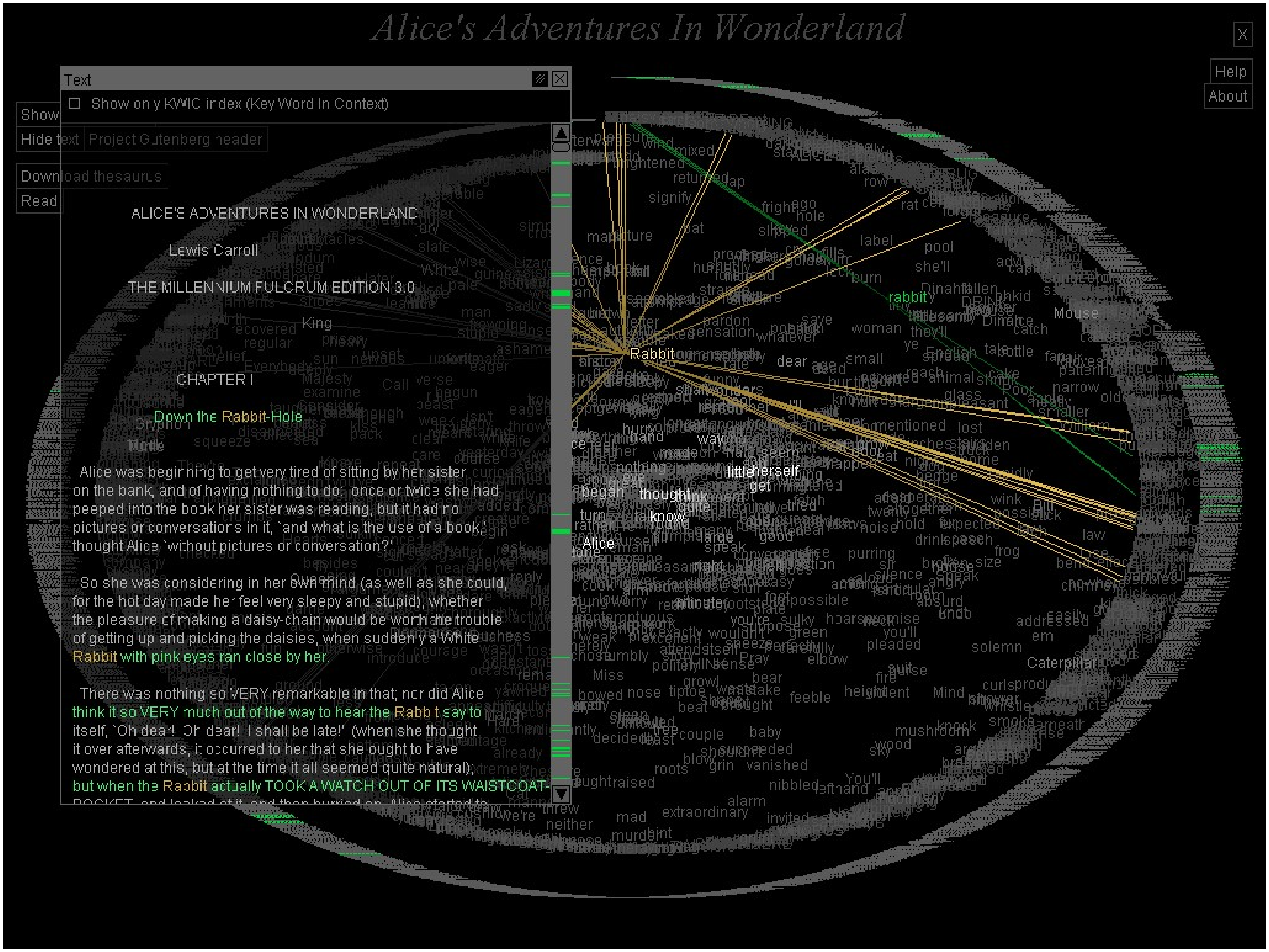
When a node is selected with the mouse, the
the selected node and each node in the path fr
to the top are brought to the front and highlig
each substructure are done in parallel, follow
rotational path, and are animated so the user
at a rate the perceptual system can track. Typ
transformation is done in about a second.
The tree can also be rotated continuously to
help the user understand substructure relation

The hierarchy is presented in 3D to maximize
available screen space and enable visualization
structure. A 2D layout of the same structure u
layout algorithms would not fit on the screen.
user would have to either scroll through the la
size-reduced image of the structure. Most hier
real applications tend to be broad and shallow
hierarchy aspect ratio is problematic for 2D la
size-reduced image may look like a line with l
layout uses depth to fill the screen with more

To see this effect analytically, consider the asp
tree, ignoring the size of the nodes. If there ar
branching factor is b, the width of the base
aspect ratio is b^{d-1}/d. Aspect ratio for 2D
nearly exponentially, and is much worse as th
larger. Figure~\ref{aspectratiofigure} shows w
branching factors (b=3 and b=5). In cont
aspect ratio is fixed to fit the room by adjustin
cone diameters to fit. The line near the bottom
Figure~\ref{aspectratiofigure} is a typical aspe
three. Although fixing the aspect ratio limits
the number of levels that can be effectively di
makes Cone Trees independent of the number
and number of levels (until the limit is reache

In addition to perceptual effects already menti
perspective view of Cone Trees provides a fish
of the information, without having to describe
function, as in general fisheye view mechanism
brighter, closer, and larger that other paths. In

TextArc



☐ Show only KWIC index (Key Word In Context)

ALICE'S ADVENTURES IN WONDERLAND

Lewis Carroll

THE MILLENNIUM FULCRUM EDITION 3.0

CHAPTER I

Down the Rabbit-Hole

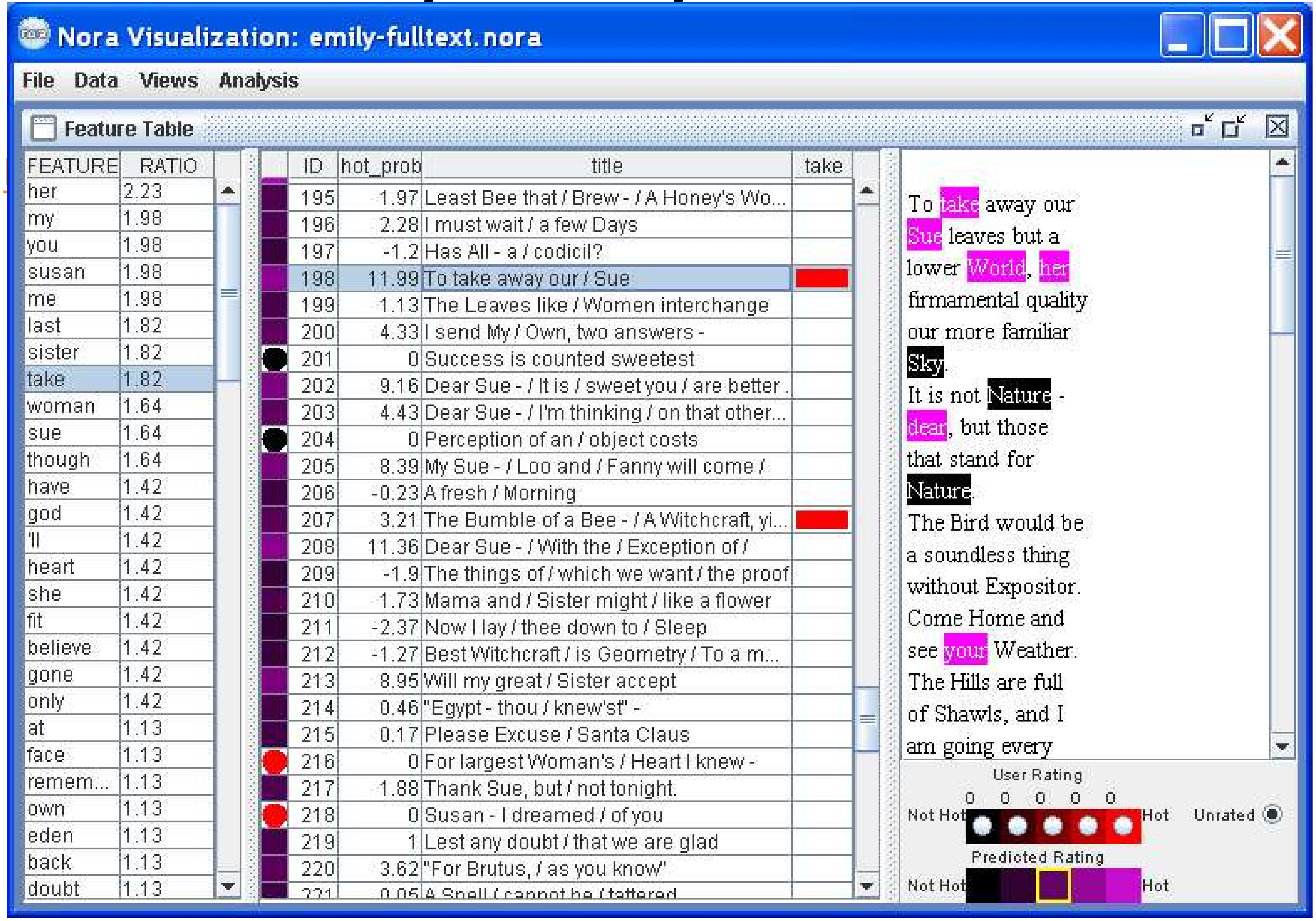
Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, `and what is the use of a book,' thought Alice `without pictures or conversation?'

So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.

There was nothing so VERY remarkable in that; nor did Alice think it so VERY much out of the way to hear the Rabbit say to itself, `Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually TOOK A WATCH OUT OF ITS WAISTCOAT-POCKET, and looked at it, and then hurried on, Alice started to

- http://www.textarc.org/Hamlet2.html

Literary Analysis: Semantics



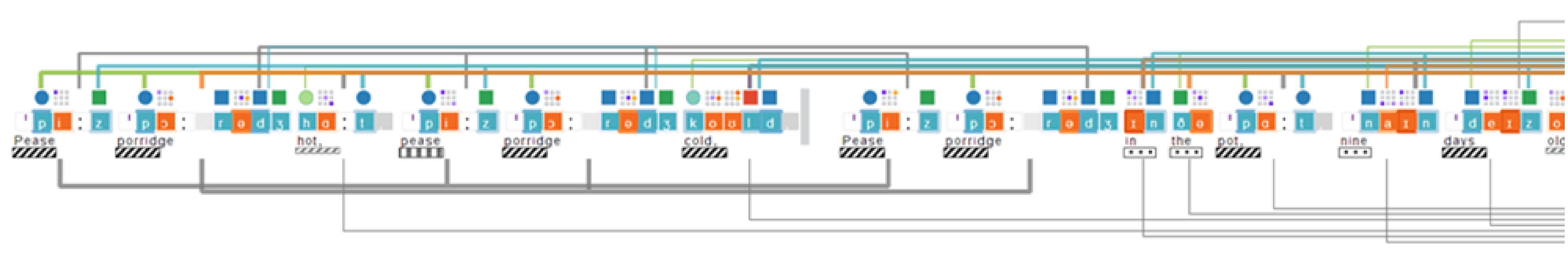
FEATURE	RATIO
her	2.23
my	1.98
you	1.98
susan	1.98
me	1.98
last	1.82
sister	1.82
take	1.82
woman	1.64
sue	1.64
though	1.64
have	1.42
god	1.42
'll	1.42
heart	1.42
she	1.42
fit	1.42
believe	1.42
gone	1.42
only	1.42
at	1.13
face	1.13
remem...	1.13
own	1.13
eden	1.13
back	1.13
doubt	1.13

	ID	hot_prob	title	take
	195	1.97	Least Bee that / Brew - / A Honey's Wo...	
	196	2.28	I must wait / a few Days	
	197	-1.2	Has All - a / codicil?	
	198	11.99	To take away our / Sue	🟥
	199	1.13	The Leaves like / Women interchange	
	200	4.33	I send My / Own, two answers -	
●	201	0	Success is counted sweetest	
	202	9.16	Dear Sue - / It is / sweet you / are better .	
	203	4.43	Dear Sue - / I'm thinking / on that other...	
●	204	0	Perception of an / object costs	
	205	8.39	My Sue - / Loo and / Fanny will come /	
	206	-0.23	A fresh / Morning	
	207	3.21	The Bumble of a Bee - / A Witchcraft, yi...	🟥
	208	11.36	Dear Sue - / With the / Exception of /	
	209	-1.9	The things of / which we want / the proof	
	210	1.73	Mama and / Sister might / like a flower	
	211	-2.37	Now I lay / thee down to / Sleep	
	212	-1.27	Best Witchcraft / is Geometry / To a m...	
	213	8.95	Will my great / Sister accept	
	214	0.46	"Egypt - thou / knew'st" -	
	215	0.17	Please Excuse / Santa Claus	
●	216	0	For largest Woman's / Heart I knew -	
	217	1.88	Thank Sue, but / not tonight.	
●	218	0	Susan - I dreamed / of you	
	219	1	Lest any doubt / that we are glad	
	220	3.62	"For Brutus, / as you know"	
	221	0.05	A Spell / cannot be / tattered	

To take away our
Sue leaves but a
lower World, her
firmamental quality
our more familiar
Sky.
It is not Nature -
dear, but those
that stand for
Nature.
The Bird would be
a soundless thing
without Expositor.
Come Home and
see your Weather.
The Hills are full
of Shawls, and I
am going every

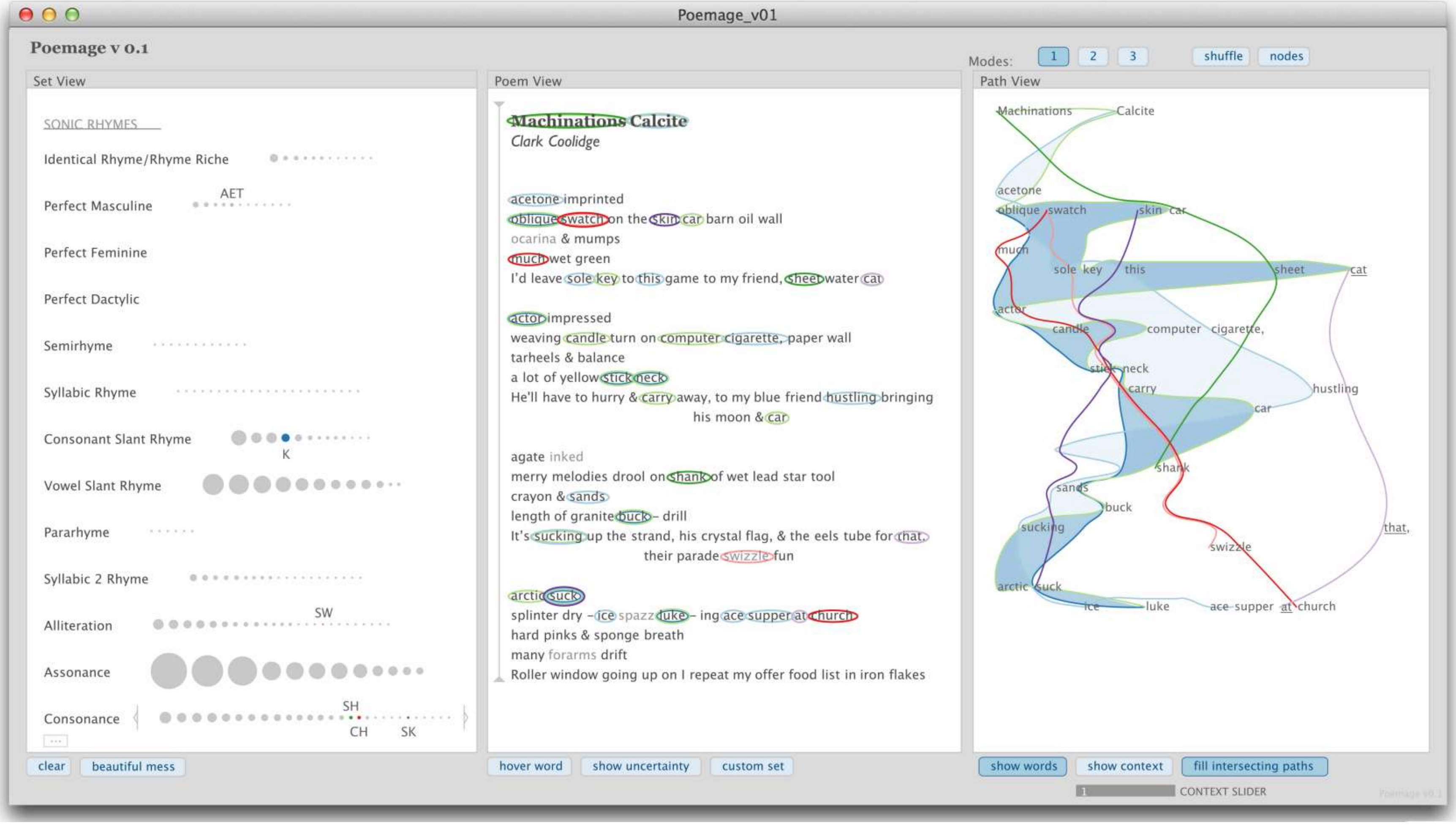
Literary Analysis: PoemViewer

- Phonetics, repetition



http://ovii.oerc.ox.ac.uk/PoemVis/

Literary Analysis: Poemage



Machinations Calcite
Clark Coolidge

acetone imprinted
oblique swatch on the skin car barn oil wall
ocarina & mumps
much wet green
I'd leave sole key to this game to my friend, sheet water cat

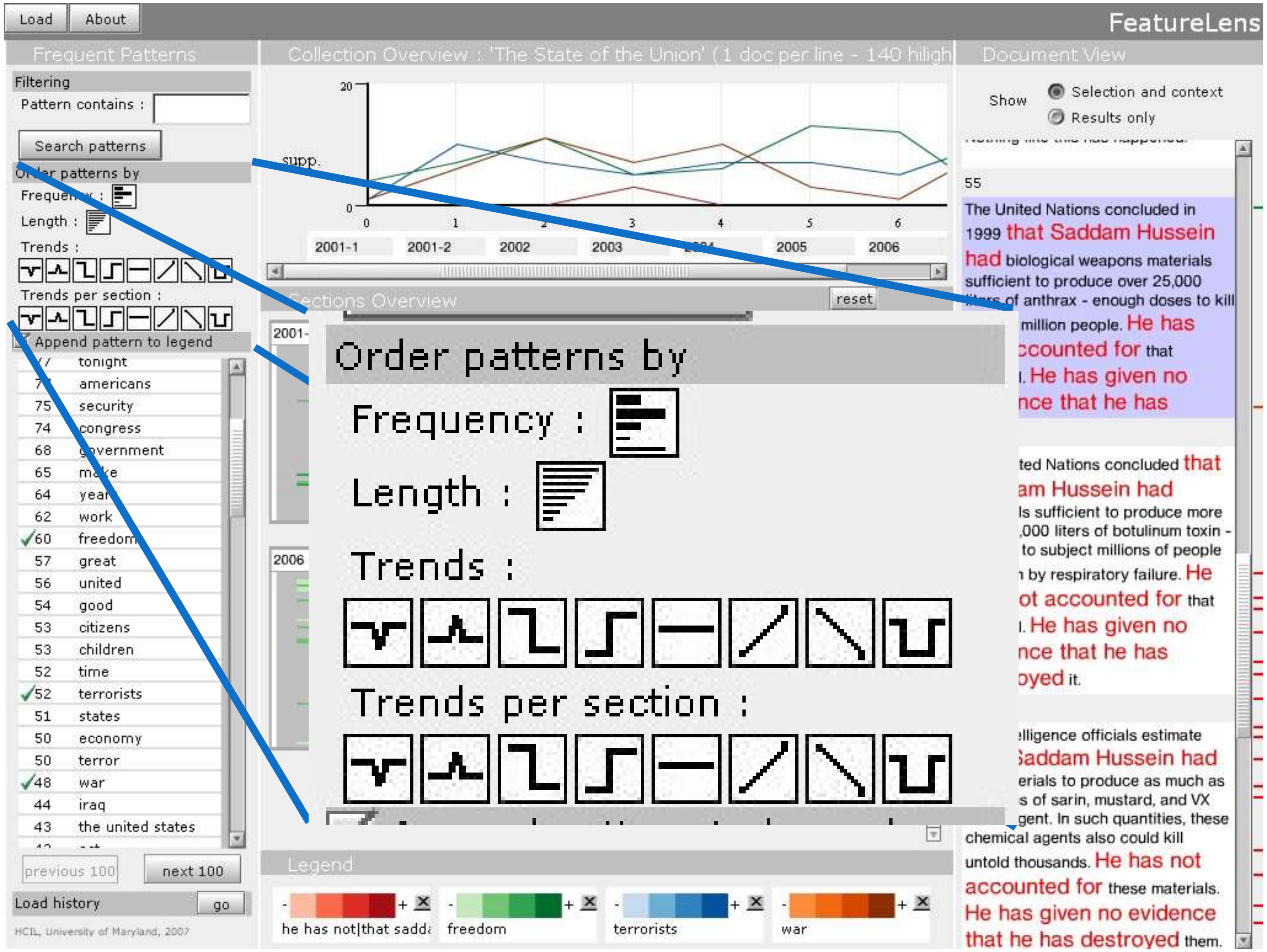
actor impressed
weaving candle turn on computer cigarette, paper wall
tarheels & balance
a lot of yellow stick neck
He'll have to hurry & carry away, to my blue friend hustling bringing
his moon & car

agate inked
merry melodies drool on shank of wet lead star tool
crayon & sands
length of granite buck – drill
It's sucking up the strand, his crystal flag, & the eels tube for that,
their parade swizzle fun

arctic suck
splinter dry – ice spazz luke – ing ace supper at church
hard pinks & sponge breath
many forarms drift
Roller window going up on I repeat my offer food list in iron flakes

http://www.sci.utah.edu/~nmccurdy/Poemage/

Literary Analysis: Patterns



77	tonight
7?	americans
75	security
74	congress
68	government
65	make
64	year
62	work
✓60	freedom
57	great
56	united
54	good
53	citizens
53	children
52	time
✓52	terrorists
51	states
50	economy
50	terror
✓48	war
44	iraq
43	the united states

Nothing like this has happened.

55

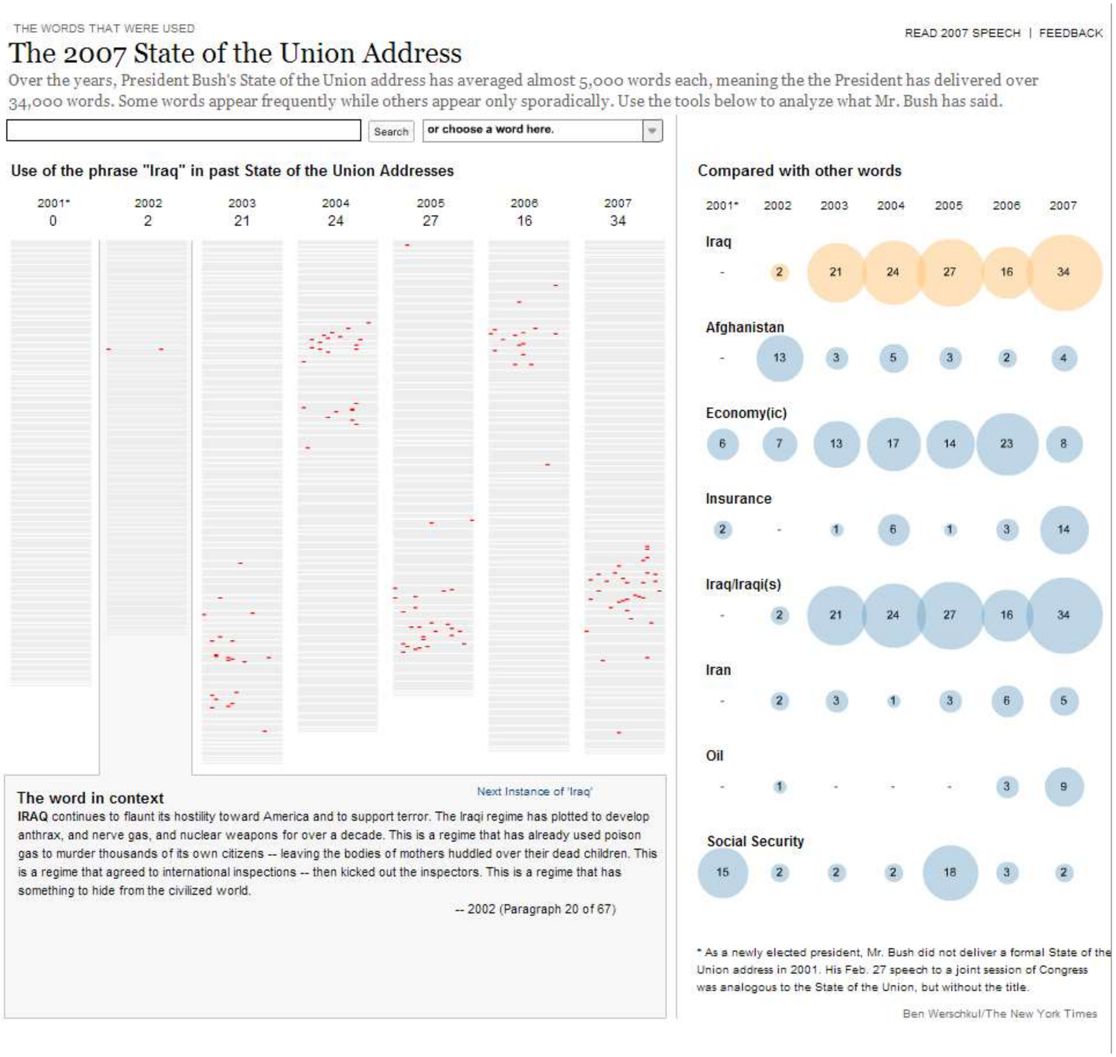
The United Nations concluded in 1999 that Saddam Hussein had biological weapons materials sufficient to produce over 25,000 liters of anthrax - enough doses to kill ... million people. He has ...ccounted for that ...l. He has given no ...nce that he has

...ted Nations concluded that ...am Hussein had ...ls sufficient to produce more ...000 liters of botulinum toxin - ... to subject millions of people ... by respiratory failure. He ...ot accounted for that ...l. He has given no ...nce that he has ...oyed it.

...lligence officials estimate ...Saddam Hussein had ...erials to produce as much as ...s of sarin, mustard, and VX ...gent. In such quantities, these chemical agents also could kill untold thousands. He has not accounted for these materials. He has given no evidence that he has destroyed them.

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

Literary Analysis: Patterns



THE WORDS THAT WERE USED

The 2007 State of the Union Address

READ 2007 SPEECH | FEEDBACK

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 "Iraq" in past State of the Union Addresses

2001*	2002	2003	2004	2005	2006	2007
0	2	21	24	27	16	34

The word in context

IRAQ continues to flaunt its hostility toward America and to support terror. The Iraqi regime has plotted to develop anthrax, and nerve gas, and nuclear weapons for over a decade. This is a regime that has already used poison gas to murder thousands of its own citizens -- leaving the bodies of mothers huddled over their dead children. This is a regime that agreed to international inspections -- then kicked out the inspectors. This is a regime that has something to hide from the civilized world.

-- 2002 (Paragraph 20 of 67)

Compared with other words

	2001*	2002	2003	2004	2005	2006	2007
Iraq	-	2	21	24	27	16	34
Afghanistan	-	13	3	5	3	2	4
Economy(ic)	6	7	13	17	14	23	8
Insurance	2	-	1	6	1	3	14
Iraq/Iraqi(s)	-	2	21	24	27	16	34
Iran	-	2	3	1	3	6	5
Oil	-	1	-	-	-	3	9
Social Security	15	2	2	2	18	3	2

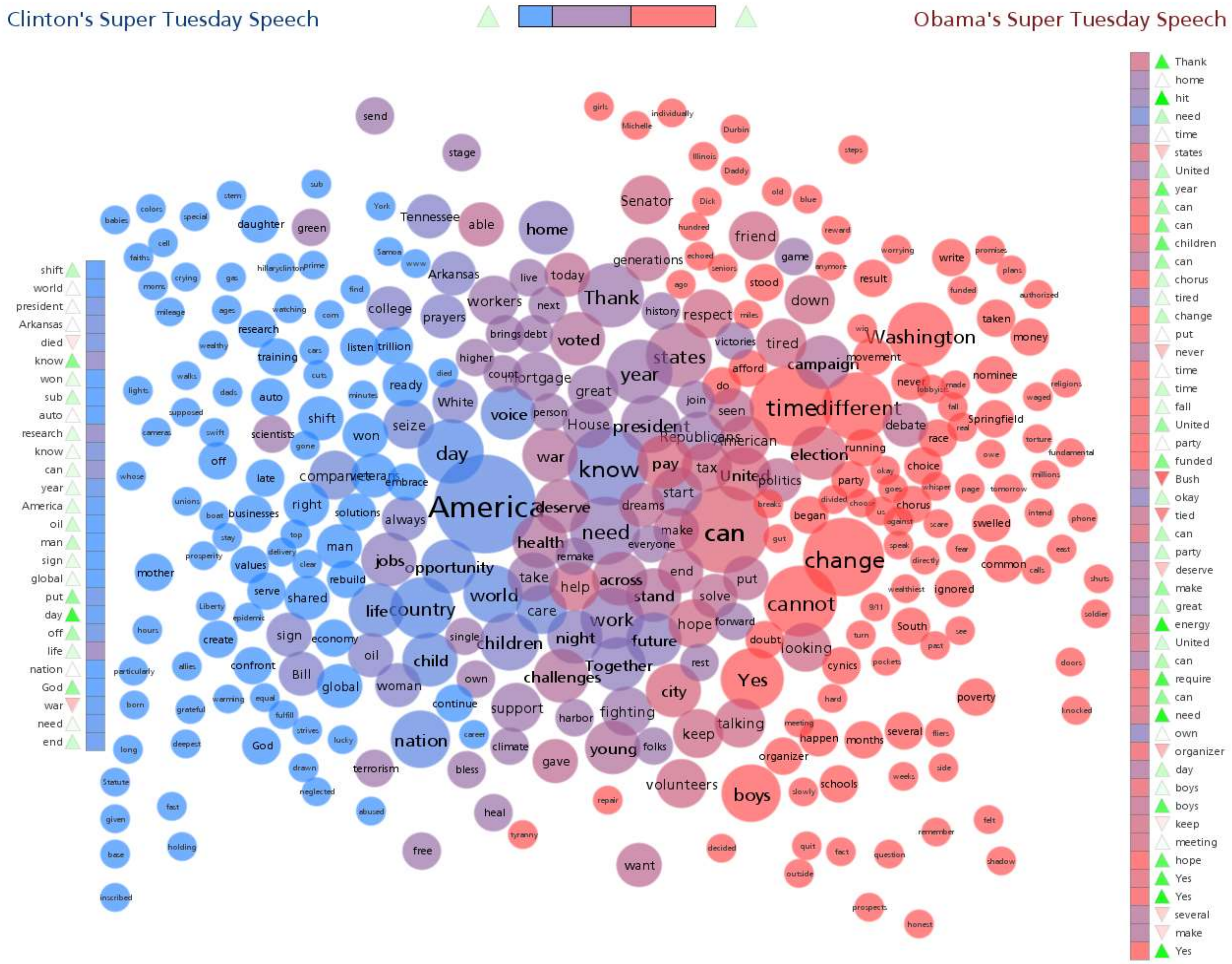
* As a newly elected president, Mr. Bush did not deliver a formal State of the Union address in 2001. His Feb. 27 speech to a joint session of Congress was analogous to the State of the Union, but without the title.

Ben Werschkul/The New York Times

NY Times
(no longer working)

(Werschkul, 2007)

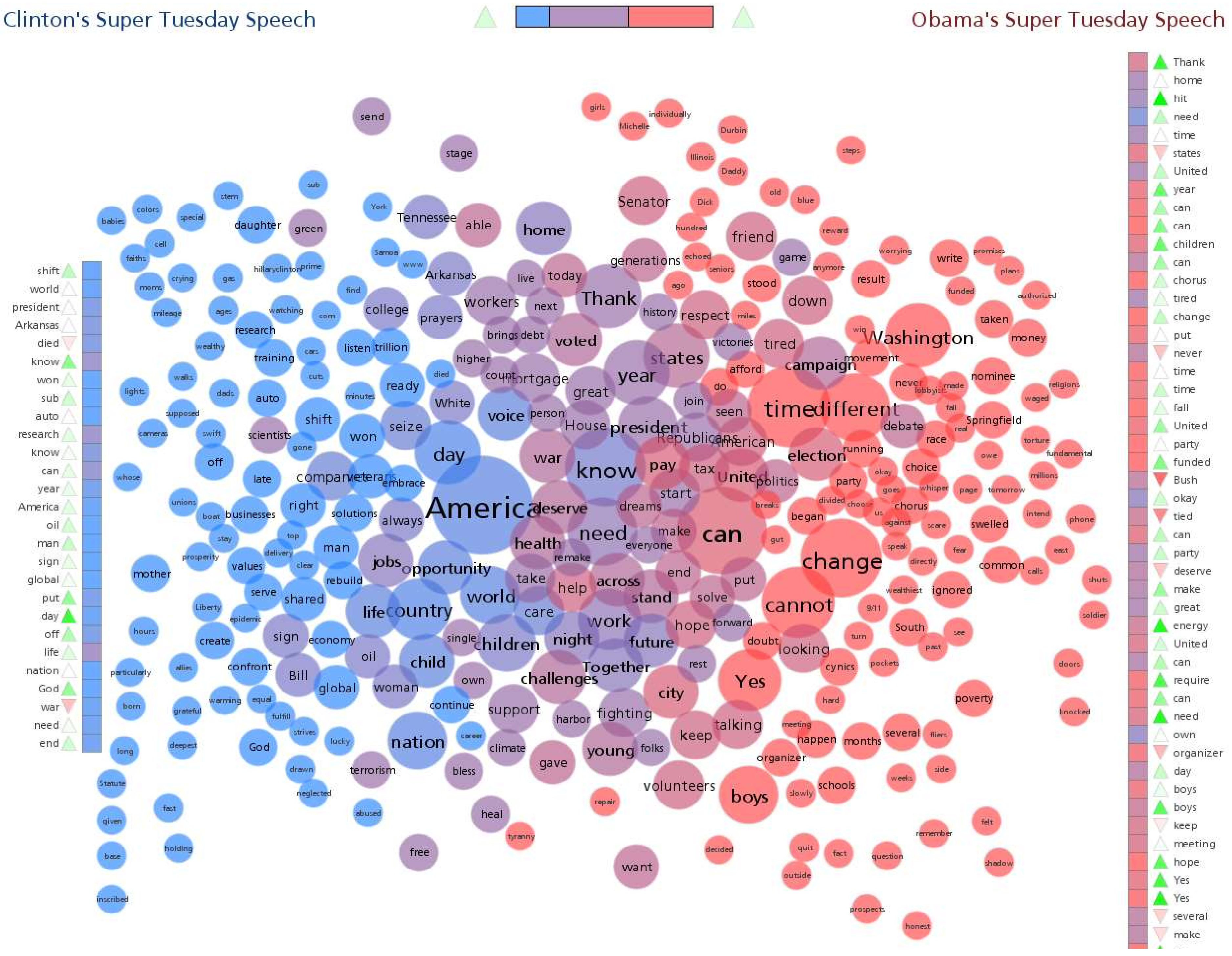
Twitter Contrast Diagrams



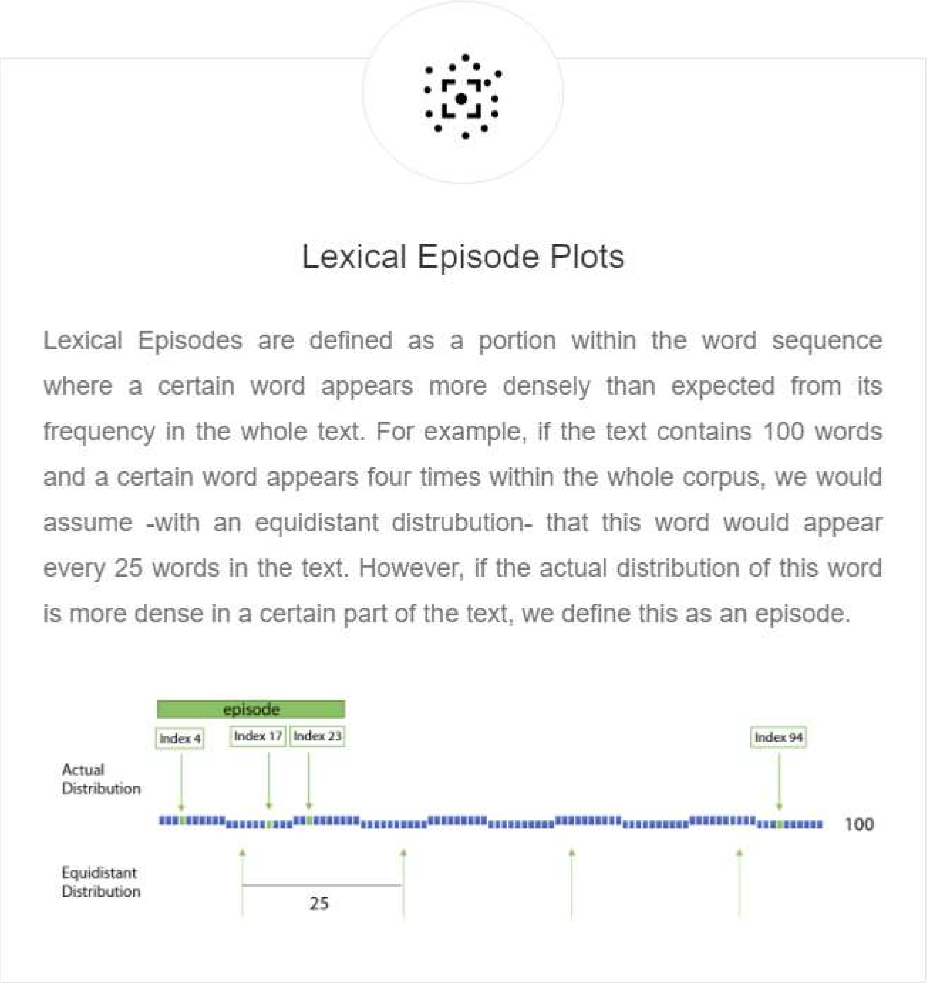
(Clark, 2008)

Clinton's Super Tuesday Speech

Obama's Super Tuesday Speech

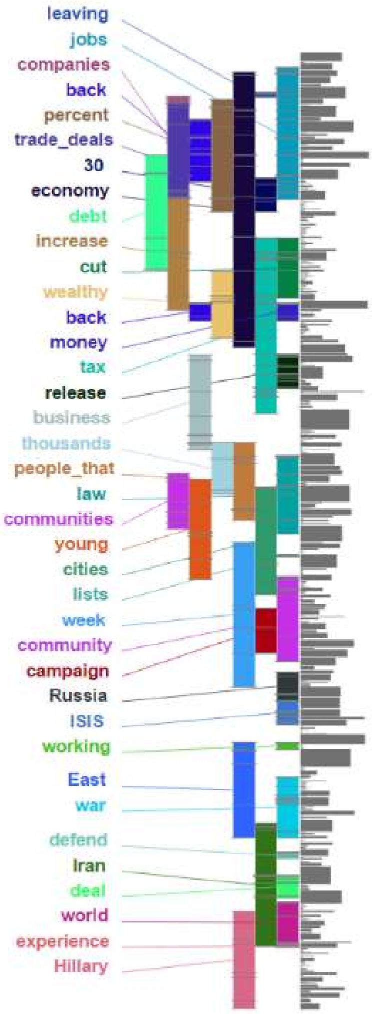


VisArgue Project



Lexical Episode Plots

Lexical Episodes are defined as a portion within the word sequence where a certain word appears more densely than expected from its frequency in the whole text. For example, if the text contains 100 words and a certain word appears four times within the whole corpus, we would assume -with an equidistant distrubution- that this word would appear every 25 words in the text. However, if the actual distribution of this word is more dense in a certain part of the text, we define this as an episode.



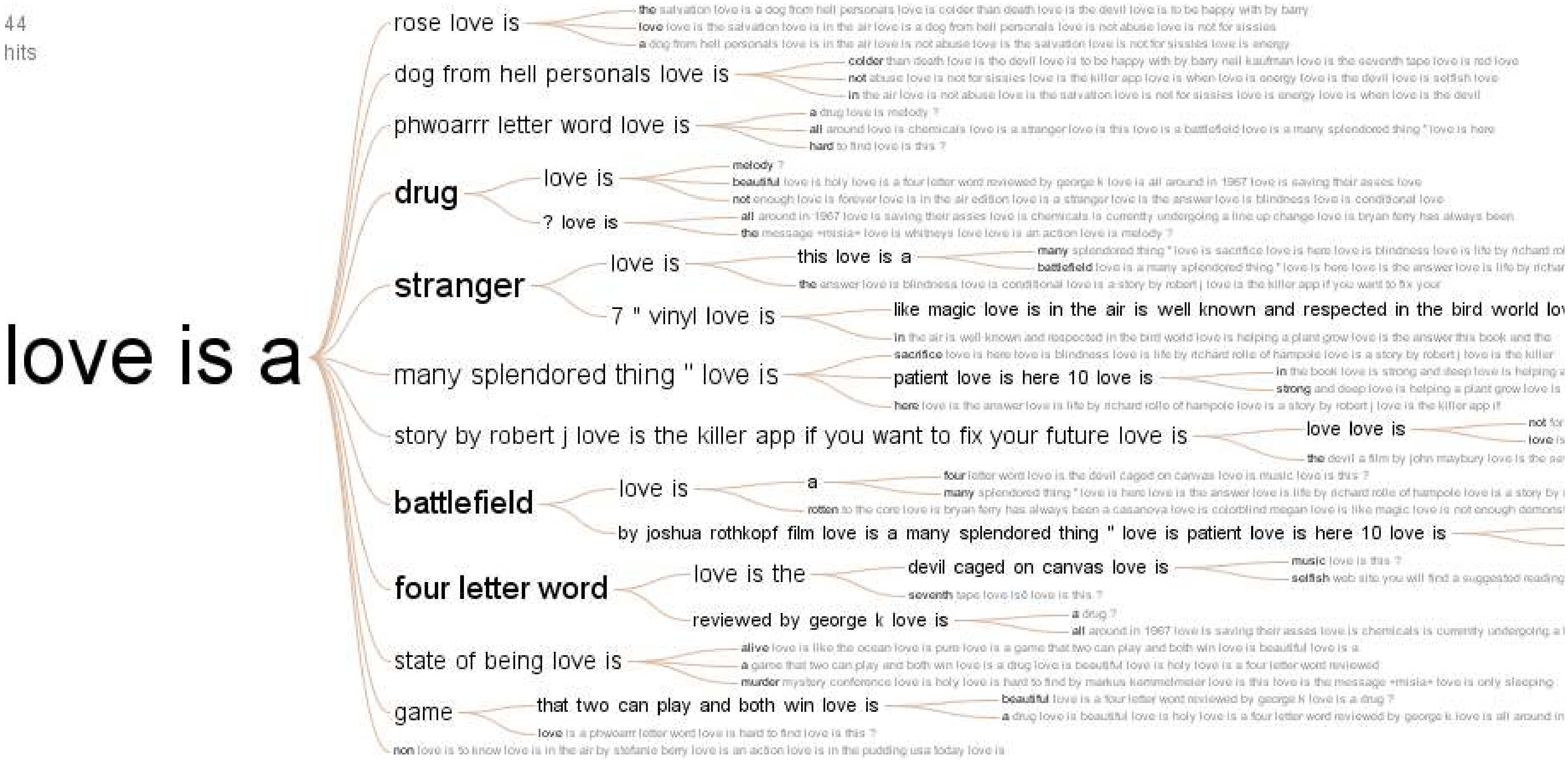
Mennatallah El-Assady, http://presidential-debates.dbvis.de/

Many Eyes

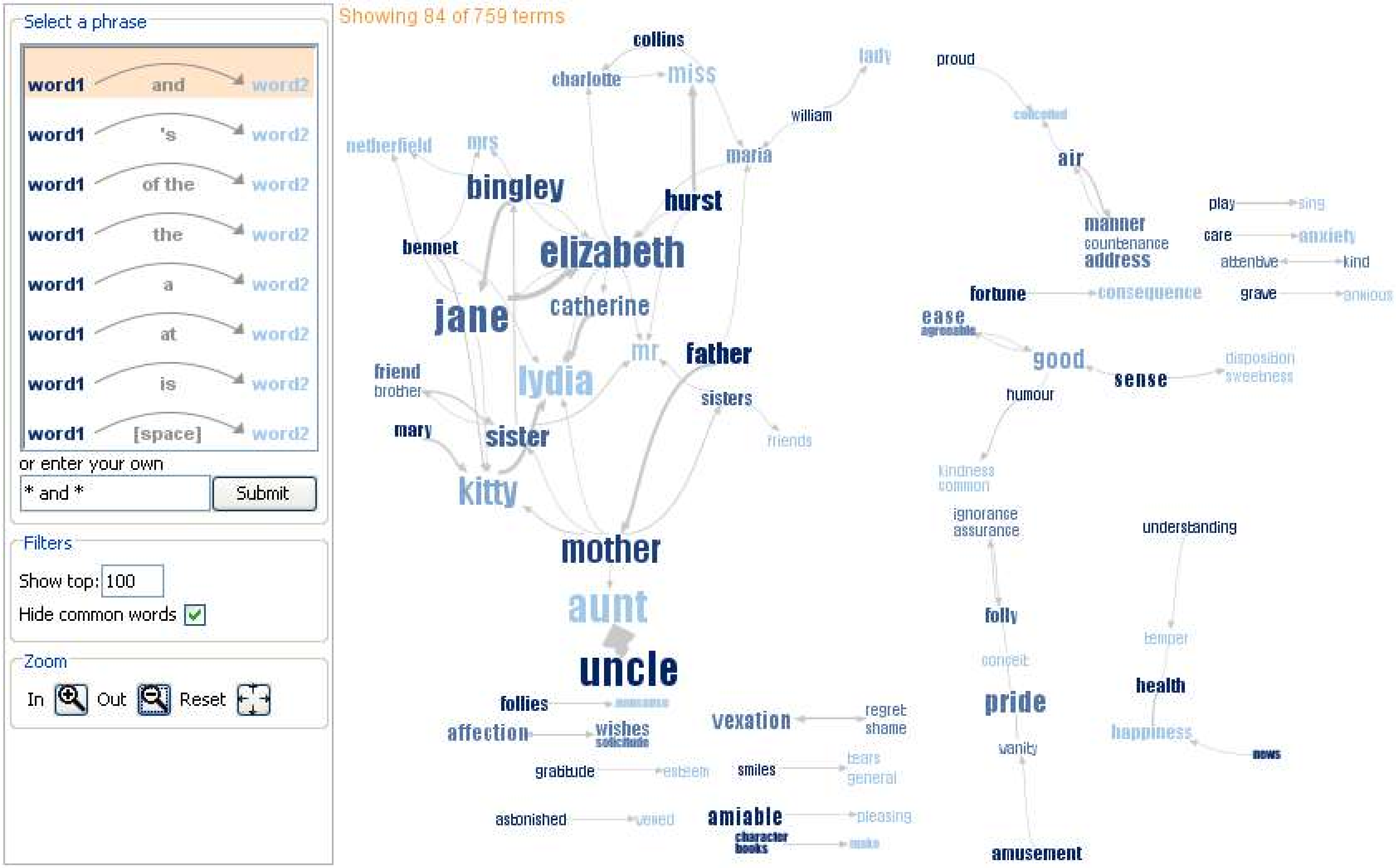
- IBM system for uploading your data and visualizing it
- You can share your visualizations (recall: empowerment aspect of Critical Visualization)

- http://www-969.ibm.com/software/analytics/manyeyes/

Literary Analysis: Repetition


44
hits
love is a
rose love is
the salvation love is a dog from hell personals love is colder than death love is the devil love is to be happy with by barry
love love is the salvation love is in the air love is a dog from hell personals love is not abuse love is not for sissies
a dog from hell personals love is in the air love is not abuse love is the salvation love is not for sissies love is energy
dog from hell personals love is
colder than death love is the devil love is to be happy with by barry neil kaufman love is the seventh tape love is red love
not abuse love is not for sissies love is the killer app love is when love is energy love is the devil love is selfish love
in the air love is not abuse love is the salvation love is not for sissies love is energy love is when love is the devil
phwoarrr letter word love is
a drug love is melody ?
all around love is chemicals love is a stranger love is this love is a battlefield love is a many splendored thing " love is here
hard to find love is this ?
drug
love is
melody ?
beautiful love is holy love is a four letter word reviewed by george k love is all around in 1967 love is saving their asses love
not enough love is forever love is in the air edition love is a stranger love is the answer love is blindness love is conditional love
? love is
all around in 1967 love is saving their asses love is chemicals is currently undergoing a line up change love is bryan ferry has always been
the message +misia+ love is whitneys love love is an action love is melody ?
stranger
love is
this love is a
many splendored thing " love is sacrifice love is here love is blindness love is life by richard rol
battlefield love is a many splendored thing " love is here love is the answer love is life by richar
the answer love is blindness love is conditional love is a story by robert j love is the killer app if you want to fix your
7 " vinyl love is
like magic love is in the air is well known and respected in the bird world lov
in the air is well known and respected in the bird world love is helping a plant grow love is the answer this book and the
many splendored thing " love is
sacrifice love is here love is blindness love is life by richard rolle of hampole love is a story by robert j love is the killer
patient love is here 10 love is
in the book love is strong and deep love is helping a
strong and deep love is helping a plant grow love is
here love is the answer love is life by richard rolle of hampole love is a story by robert j love is the killer app if
story by robert j love is the killer app if you want to fix your future love is
love love is
not for
love is
the devil a film by john maybury love is the se
battlefield
love is
a
four letter word love is the devil caged on canvas love is music love is this ?
many splendored thing " love is here love is the answer love is life by richard rolle of hampole love is a story by
rotten to the core love is bryan ferry has always been a casanova love is colorblind megan love is like magic love is not enough demonst
by joshua rothkopf film love is a many splendored thing " love is patient love is here 10 love is
four letter word
love is the
devil caged on canvas love is
music love is this ?
selfish web site you will find a suggested reading
seventh tape love isĉ love is this ?
reviewed by george k love is
a drug ?
all around in 1967 love is saving their asses love is chemicals is currently undergoing a l
state of being love is
alive love is like the ocean love is pure love is a game that two can play and both win love is beautiful love is a
a game that two can play and both win love is a drug love is beautiful love is holy love is a four letter word reviewed
murder mystery conference love is holy love is hard to find by markus kemmelmeier love is this love is the message +misia+ love is only sleeping
game
that two can play and both win love is
beautiful love is a four letter word reviewed by george k love is a drug ?
a drug love is beautiful love is holy love is a four letter word reviewed by george k love is all around in
love is a phwoarrr letter word love is hard to find love is this ?
non love is to know love is in the air by stefanie berry love is an action love is in the pudding usa today love is

Literary Analysis: Repetition



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

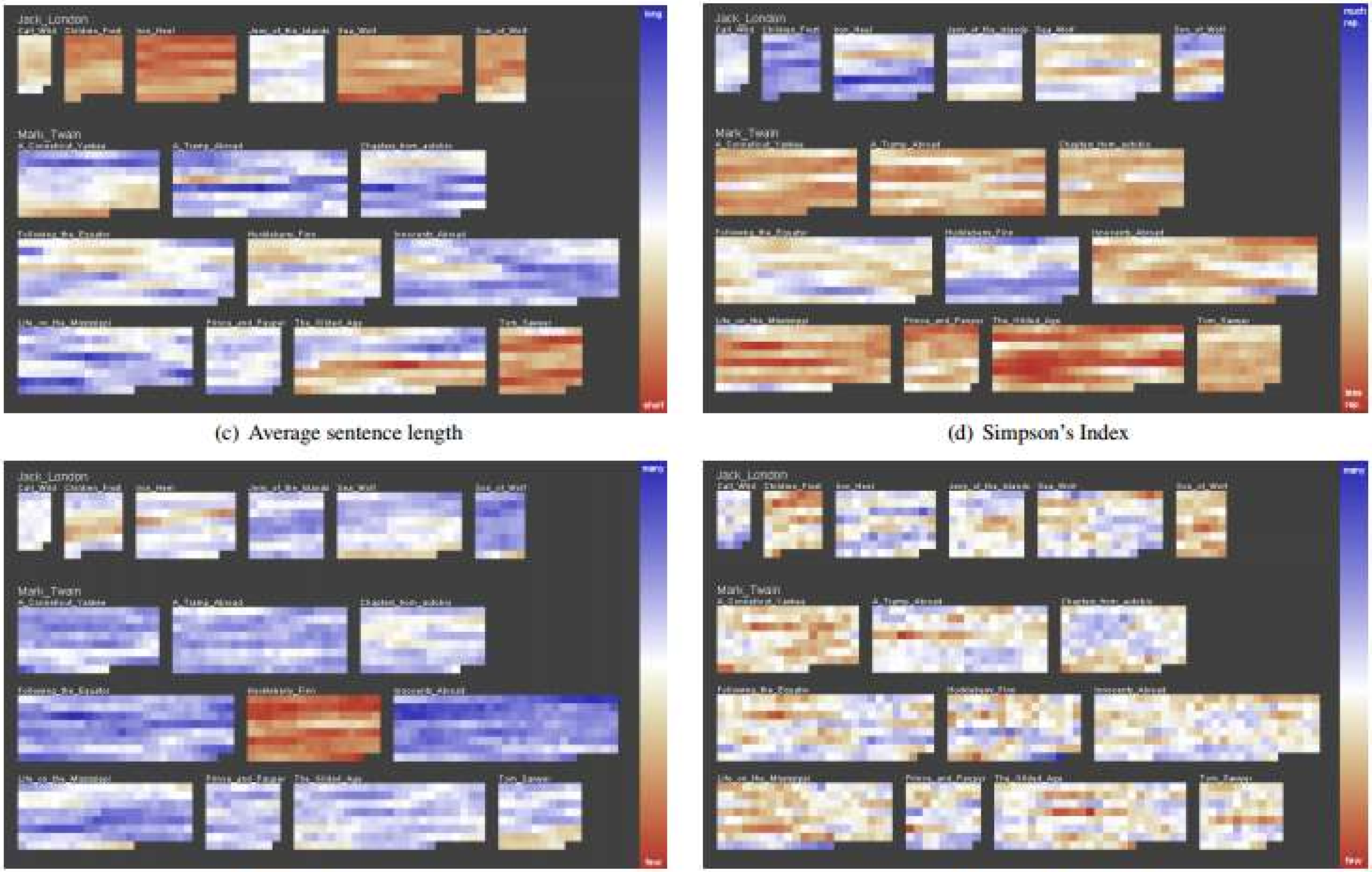
Hide common words ☑

Zoom

In Out Reset

Showing 84 of 759 terms

Literature Fingerprinting



(c) Average sentence length

(d) Simpson's Index

(e) Hapax Legomena

(f) Hapax Dislegomena

Keim, D. A., & Oelke, D. (2007). Literature Fingerprinting: A New Method for Visual Literary Analysis. In *2007 IEEE Symposium on Visual Analytics Science and Technology* (pp. 115–122).

Visual Readability Analysis

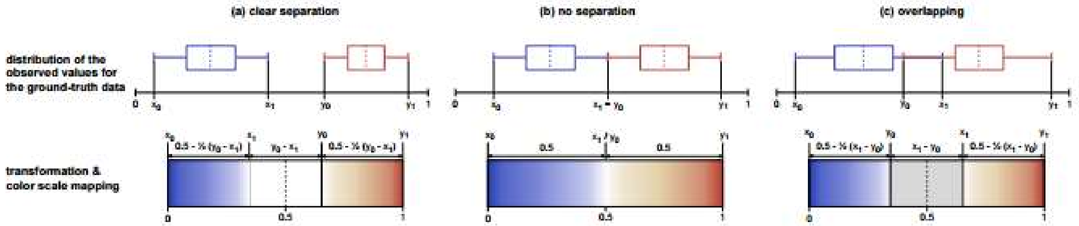


Figure 2: Normalization of the feature values is done relatively to the values that we observed for our ground-truth data set. The graphic shows the formulas and color scales for the 3 different cases that are possible.

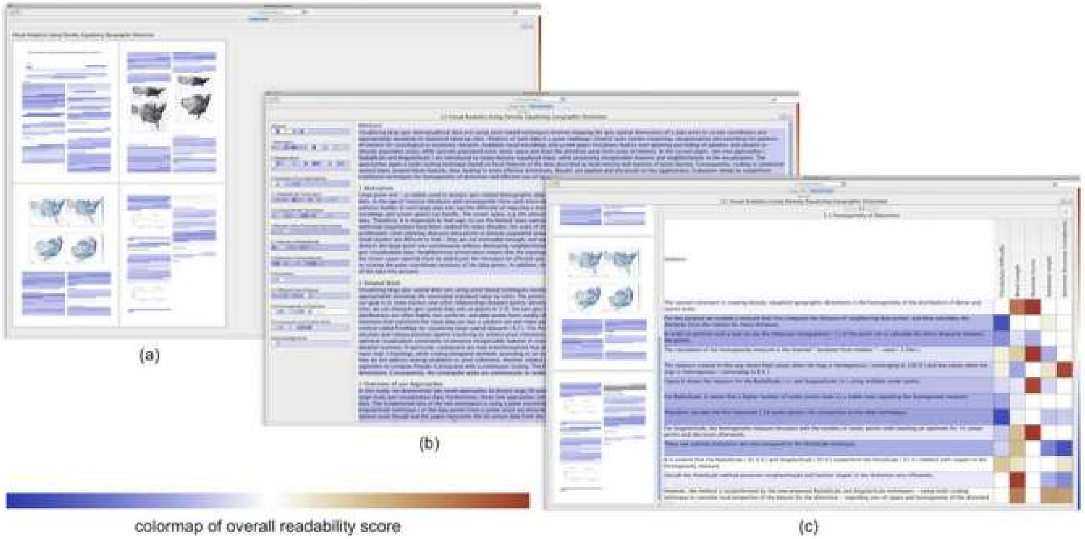
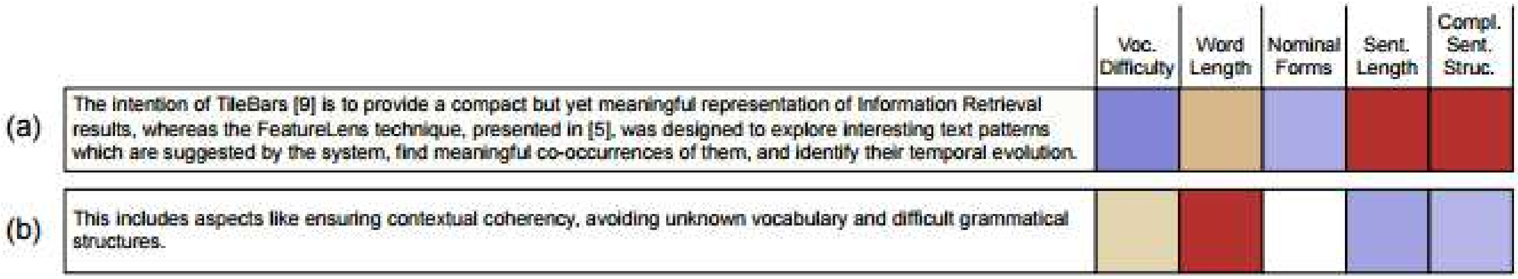


Figure 3: Screenshot of the VisRA tool on 3 different aggregation levels. (a) Corpus View (b) Block View (c) Detail View. To display single features, the colormap is generated as described in section 3.4 and figure 2.

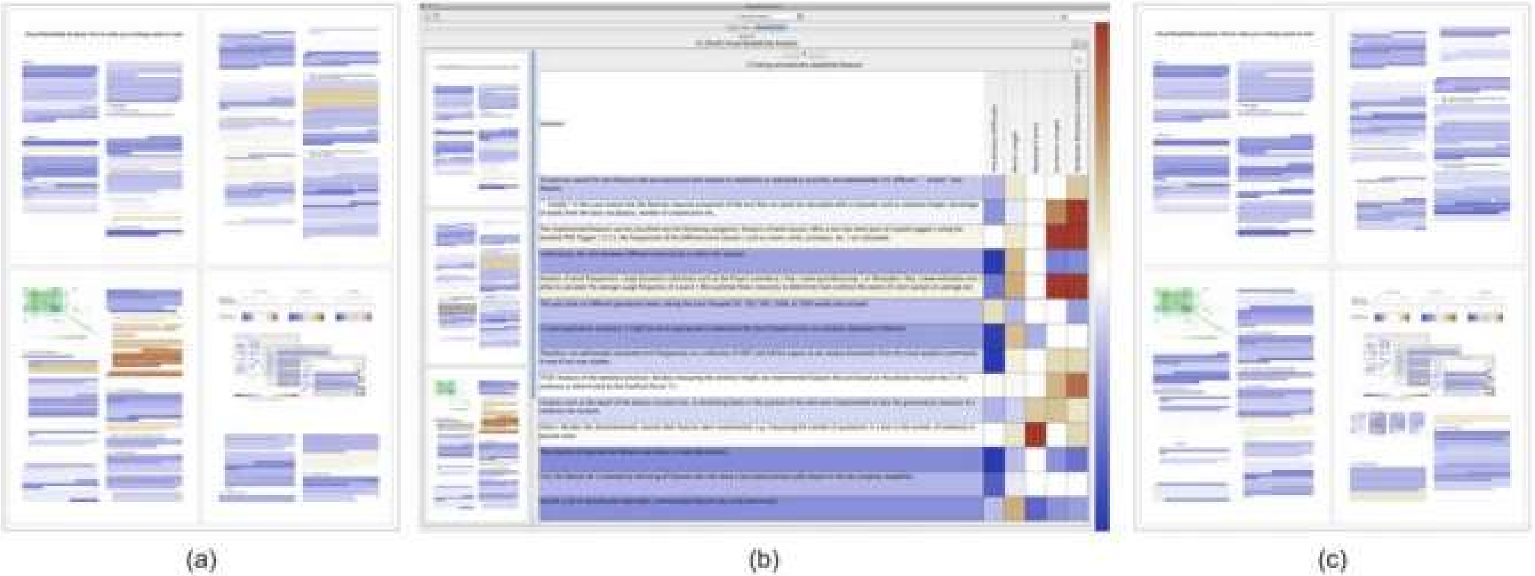
Oelke, D.; Spretke, D.; Stoffel, A.; Keim, D.A., "Visual Readability Analysis: How to Make Your Writings Easier to Read," *Visualization and Computer Graphics, IEEE Transactions on* , vol.18, no.5, pp.662,674, May 2012 doi: 10.1109/TVCG.2011.266

Visual Readability Analysis



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

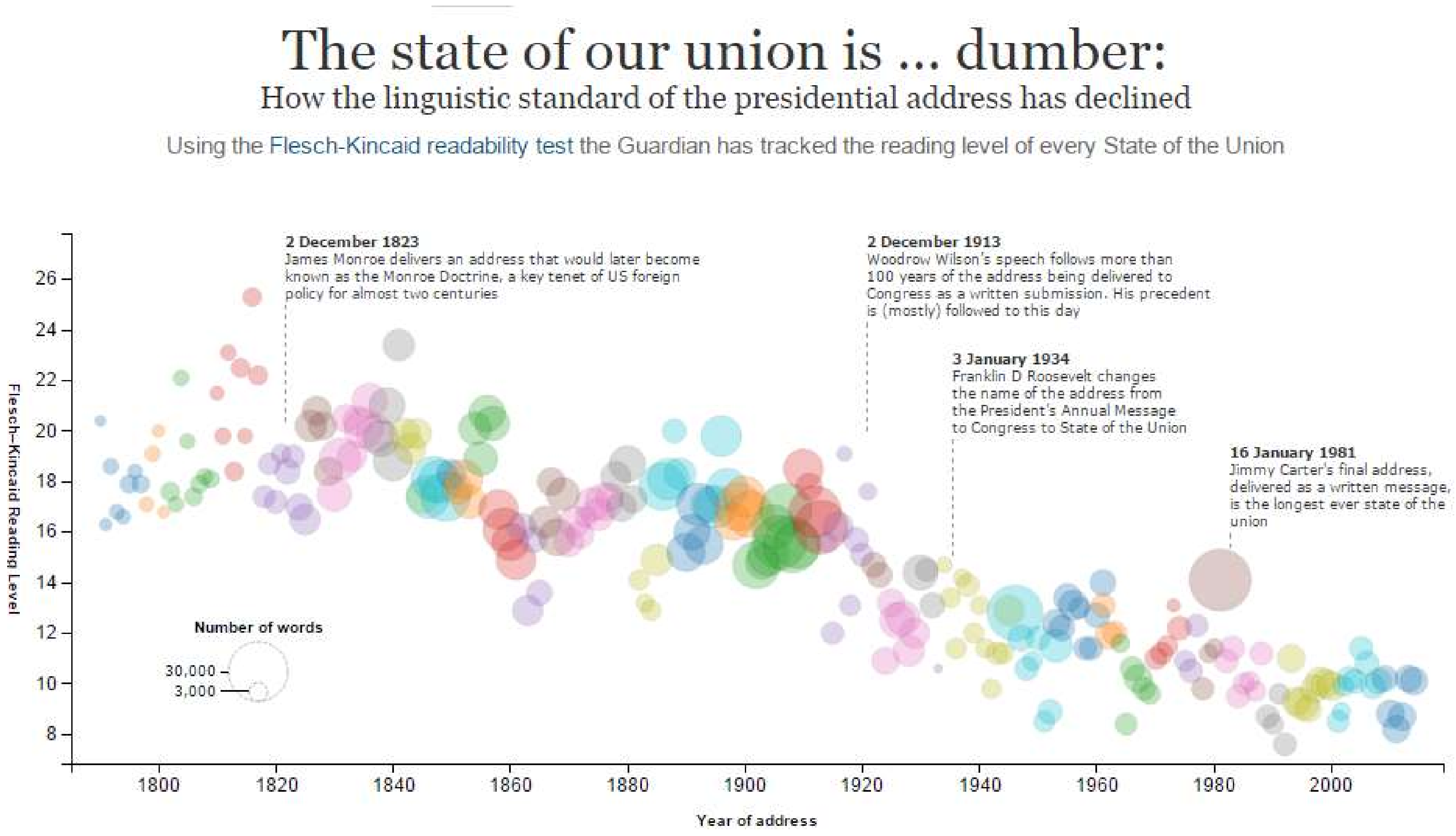
Figure 5: Two example sentences whose overall readability score is about the same. The detail view reveals the different reasons why the sentences are difficult to read.



(a) (b) (c)

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

Literary Analysis: Readability



The state of our union is ... dumber:

How the linguistic standard of the presidential address has declined

Using the Flesch-Kincaid readability test the Guardian has tracked the reading level of every State of the Union

2 December 1823
James Monroe delivers an address that would later become known as the Monroe Doctrine, a key tenet of US foreign policy for almost two centuries

2 December 1913
Woodrow Wilson's speech follows more than 100 years of the address being delivered to Congress as a written submission. His precedent is (mostly) followed to this day

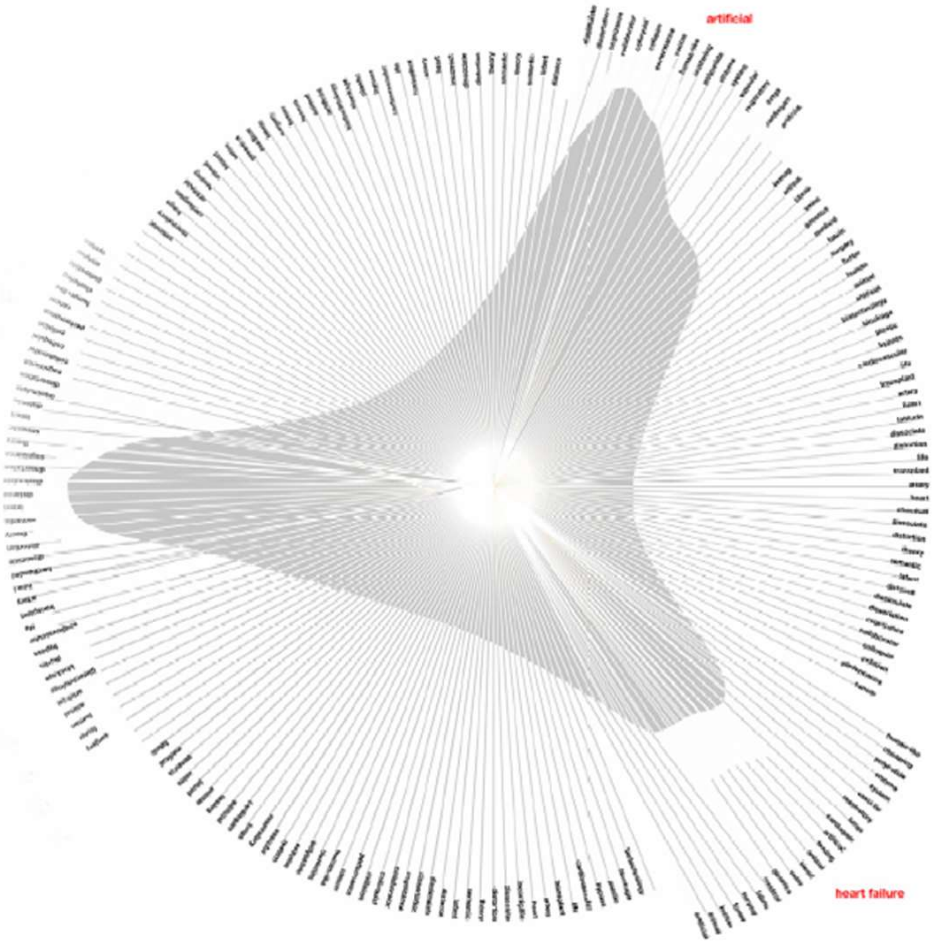
3 January 1934
Franklin D Roosevelt changes the name of the address from the President's Annual Message to Congress to State of the Union

16 January 1981
Jimmy Carter's final address, delivered as a written message, is the longest ever state of the union

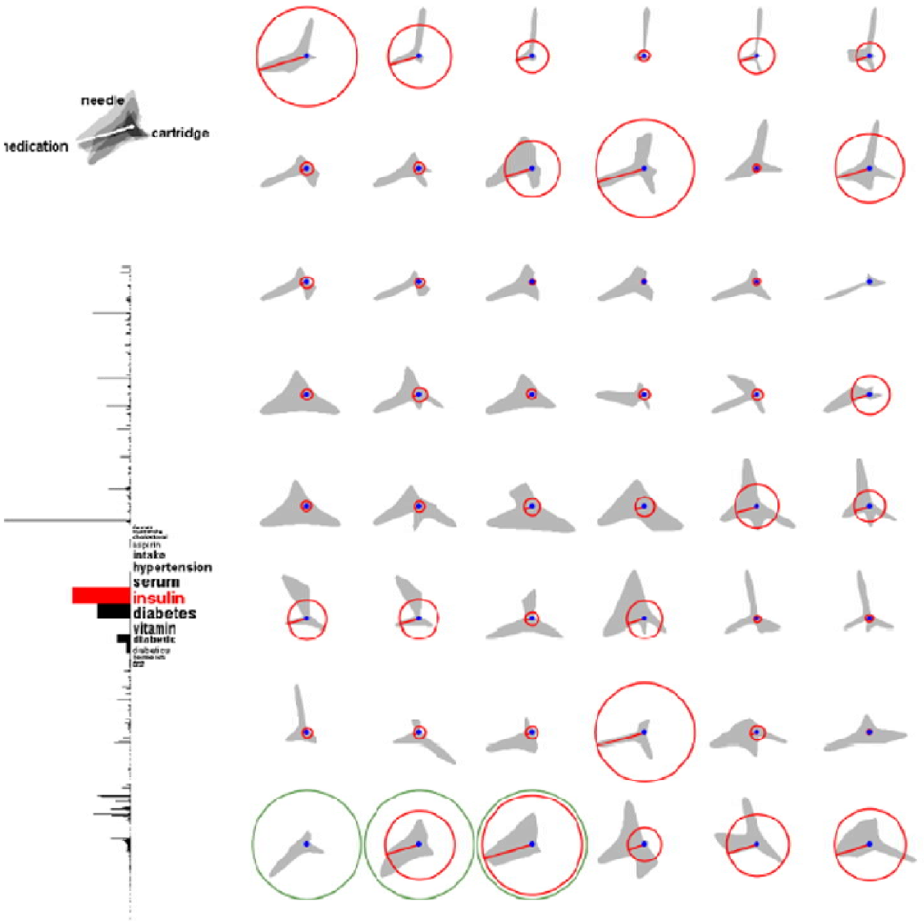
http://www.theguardian.com/world/interactive/2013/feb/12/state-of-the-union-reading-level

Gist Icons

- Word counts
- Automatic groupings
- Drill-down



Gist Icons



Chronic disease monitor

A system for monitoring a chronic disease is disclosed. The monitor includes a database for storing a plurality of patient data entries and sorts the patient data entries according to whether a test threshold is crossed. Each of the patient data entries includes personal information of a patient and a set of guidelines concerning the patient's care. The guideline represents a plurality of rules concerning a patient's care derived from accepted tests used to monitor the disease represented in an algorithm. A processor separates the patient entries designated by the user according to the test thresholds, such as for HbA1c, lipids, liver enzyme and microalbumin, for the disease of diabetes. If the test threshold value derived from the guideline is crossed, an alert sequence is activated, in which the patient is categorized as a high risk patient, the physician is notified, the patient is notified, the health care provider is notified, and the patient's treatment plan is altered to treat the high risk patient.

Diabetes management system and apparatus

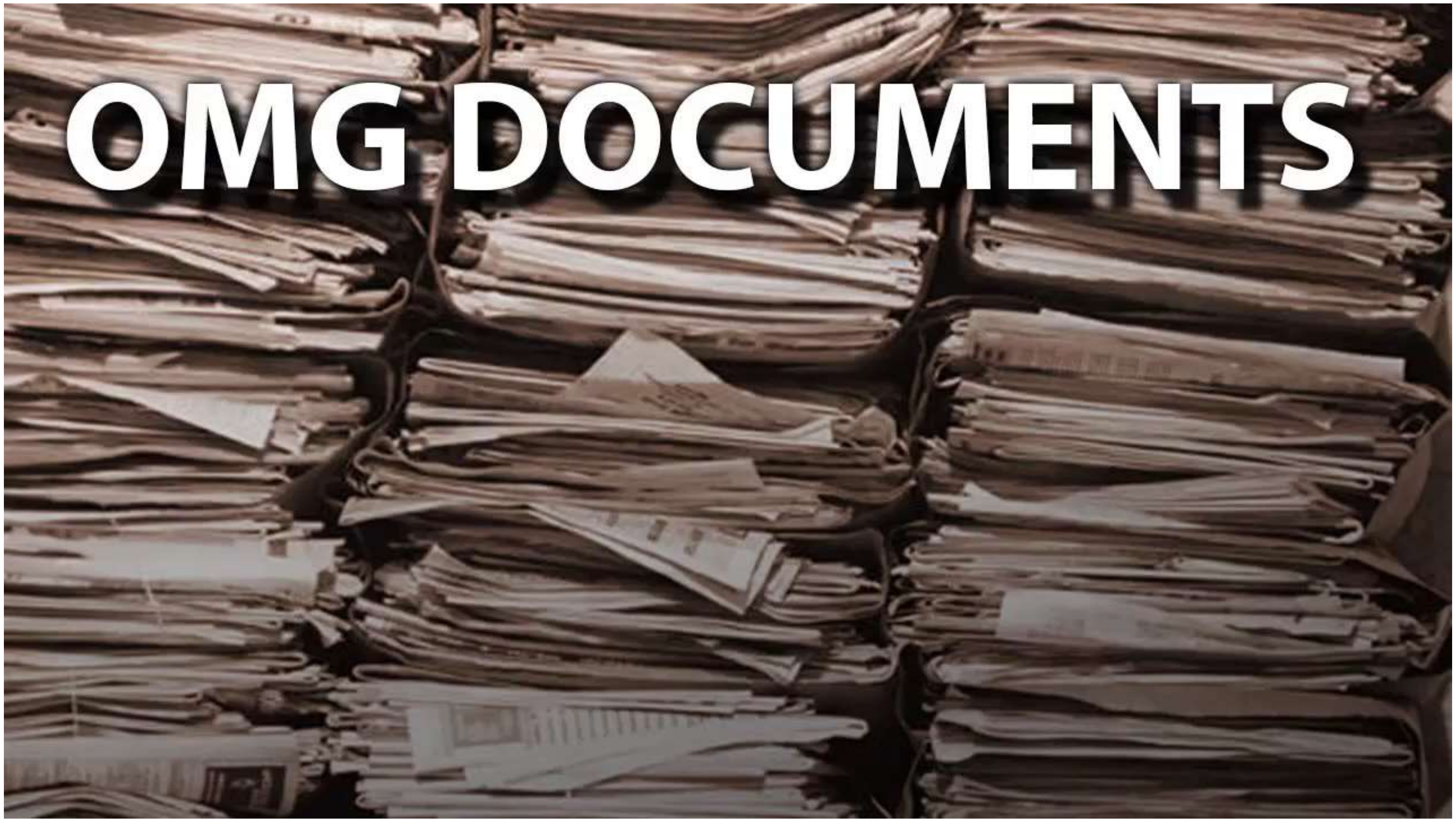
A system and apparatus for efficient medical control of a medical condition such as diabetes comprises a recorder, an interface and a master computer. The master computer develops a program of therapy which is downloaded into the recorder which then reminds the patient of any therapy due and records that the therapy has been effected. The record from the recorder is subsequently fed back to the master computer to improve or alter the therapy programme.

Method and system for the controlled release of biologicaly active

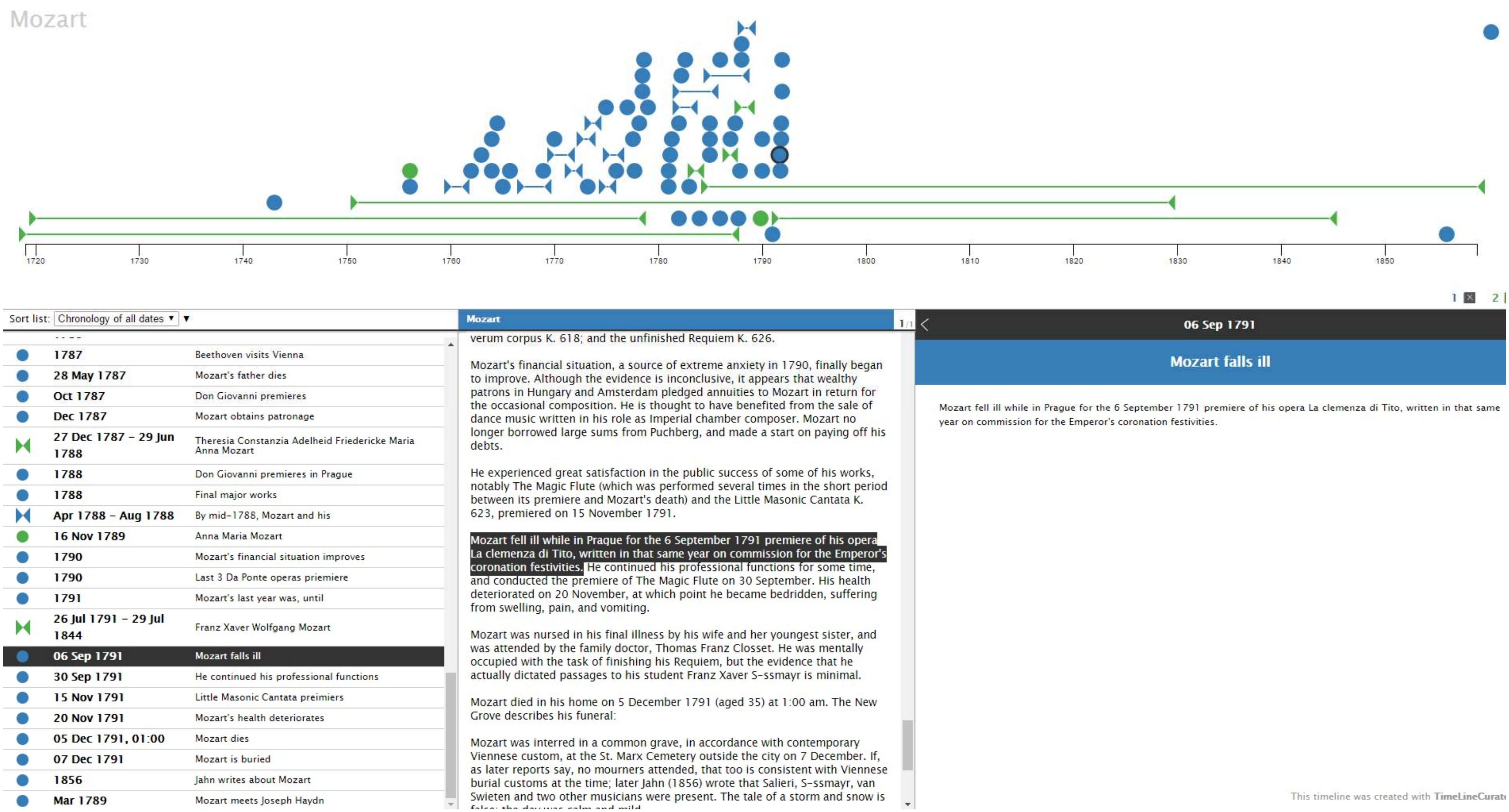
A process and system for controlled delivery of a biologically active substance to an animal body fluid which comprises contacting fluid with a reversible complex of a conjugate (1) and a binding macromolecule (2), wherein the conjugate (1) comprises a biologically active portion which is intended to be proportionately released into a body fluid stream in response to varying concentration levels of a component of the body fluid stream, and a complexing substrate portion which conjugates with the biologically active portion and which is characterized by affinity to the binding macromolecule (2), competitivey or non-competitively with the variable component of the body fluid; thereby causing the component present in the body fluid to complex to the binding macromolecule and thus releasing the conjugate (1) therefrom into the fluid.

Computational Journalism

- Overviewproject.org



Timeline Curator



Sort list: Chronology of all dates ▾

●	1787	Beethoven visits Vienna
●	28 May 1787	Mozart's father dies
●	Oct 1787	Don Giovanni premieres
●	Dec 1787	Mozart obtains patronage
⧓	27 Dec 1787 – 29 Jun 1788	Theresia Constanzia Adelheid Friedericke Maria Anna Mozart
●	1788	Don Giovanni premieres in Prague
●	1788	Final major works
⧓	Apr 1788 – Aug 1788	By mid-1788, Mozart and his
●	16 Nov 1789	Anna Maria Mozart
●	1790	Mozart's financial situation improves
●	1790	Last 3 Da Ponte operas premiere
●	1791	Mozart's last year was, until
⧓	26 Jul 1791 – 29 Jul 1844	Franz Xaver Wolfgang Mozart
●	06 Sep 1791	Mozart falls ill
●	30 Sep 1791	He continued his professional functions
●	15 Nov 1791	Little Masonic Cantata preimiers
●	20 Nov 1791	Mozart's health deteriorates
●	05 Dec 1791, 01:00	Mozart dies
●	07 Dec 1791	Mozart is buried
●	1856	Jahn writes about Mozart
●	Mar 1789	Mozart meets Joseph Haydn

Mozart 1/1

verum corpus K. 618; and the unfinished Requiem K. 626.

Mozart's financial situation, a source of extreme anxiety in 1790, finally began to improve. Although the evidence is inconclusive, it appears that wealthy patrons in Hungary and Amsterdam pledged annuities to Mozart in return for the occasional composition. He is thought to have benefited from the sale of dance music written in his role as Imperial chamber composer. Mozart no longer borrowed large sums from Puchberg, and made a start on paying off his debts.

He experienced great satisfaction in the public success of some of his works, notably The Magic Flute (which was performed several times in the short period between its premiere and Mozart's death) and the Little Masonic Cantata K. 623, premiered on 15 November 1791.

Mozart fell ill while in Prague for the 6 September 1791 premiere of his opera La clemenza di Tito, written in that same year on commission for the Emperor's coronation festivities. He continued his professional functions for some time, and conducted the premiere of The Magic Flute on 30 September. His health deteriorated on 20 November, at which point he became bedridden, suffering from swelling, pain, and vomiting.

Mozart was nursed in his final illness by his wife and her youngest sister, and was attended by the family doctor, Thomas Franz Closset. He was mentally occupied with the task of finishing his Requiem, but the evidence that he actually dictated passages to his student Franz Xaver S-ssmayr is minimal.

Mozart died in his home on 5 December 1791 (aged 35) at 1:00 am. The New Grove describes his funeral:

Mozart was interred in a common grave, in accordance with contemporary Viennese custom, at the St. Marx Cemetery outside the city on 7 December. If, as later reports say, no mourners attended, that too is consistent with Viennese burial customs at the time; later Jahn (1856) wrote that Salieri, S-ssmayr, van Swieten and two other musicians were present. The tale of a storm and snow is false; the day was calm and mild

06 Sep 1791

Mozart falls ill

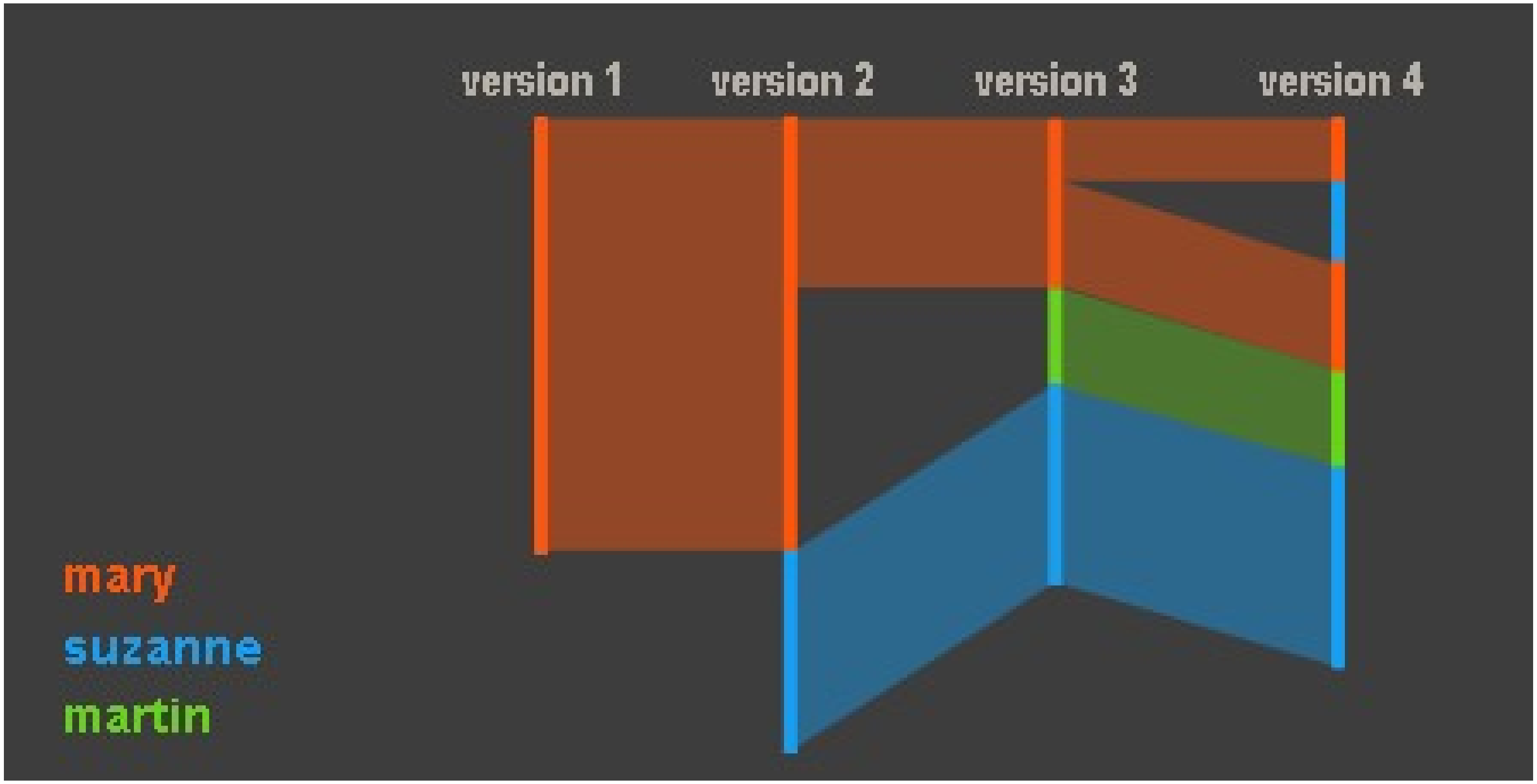
Mozart fell ill while in Prague for the 6 September 1791 premiere of his opera La clemenza di Tito, written in that same year on commission for the Emperor's coronation festivities.

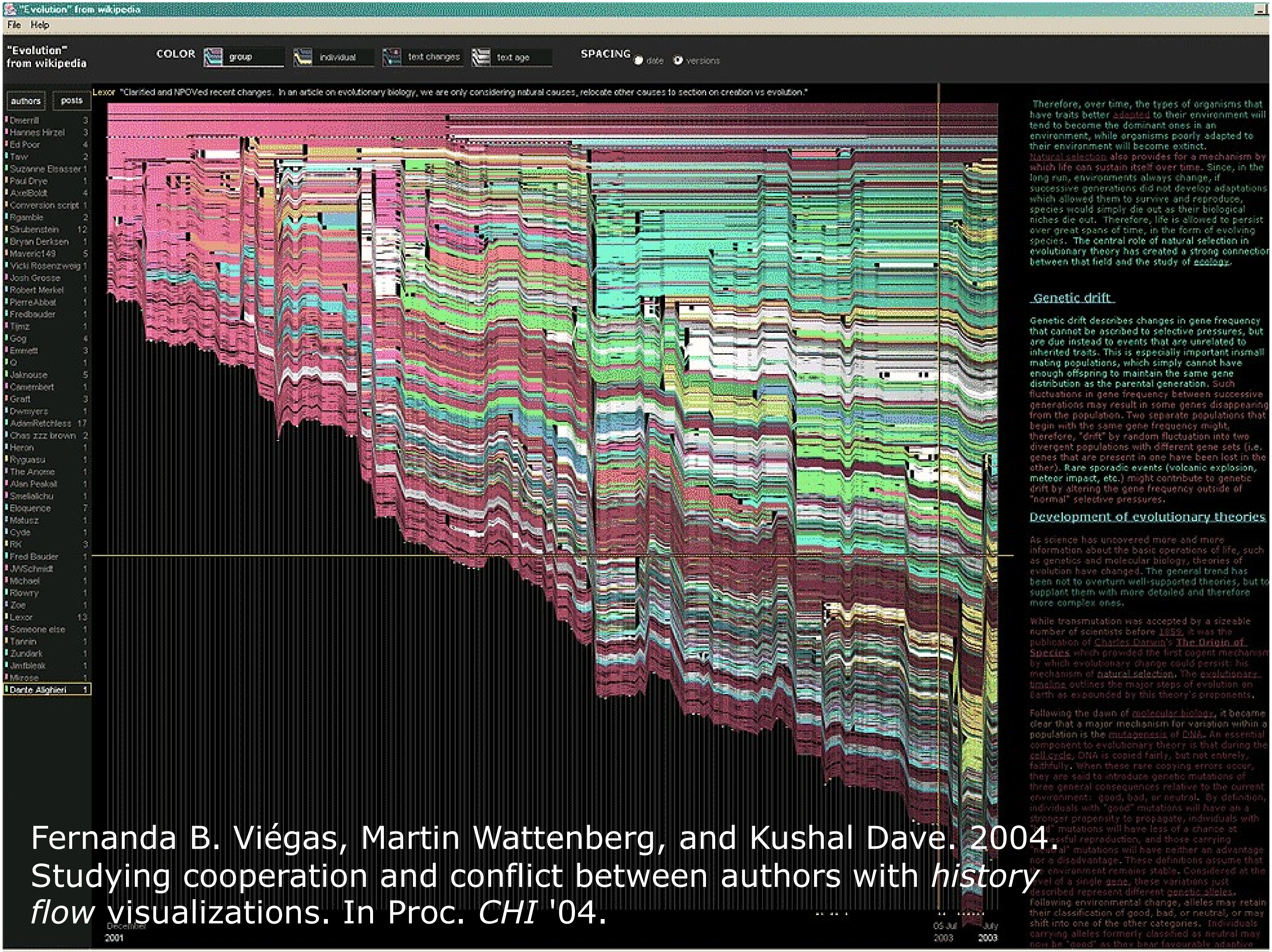
This timeline was created with TimeLineCurat

Fulda et al., http://www.cs.ubc.ca/group/infovis/software/TimeLineCurator/

Social Patterns from Text

- HistoryFlow – edits on Wikipedia





Lexor "Clarified and NPOVed recent changes. In an article on evolutionary biology, we are only considering natural causes, relocate other causes to section on creation vs evolution."

authors	posts
Dmerrill	3
Hannes Hirzel	3
Ed Poor	4
Taw	2
Suzanne Elsasser	1
Paul Drye	1
AxelBoldt	4
Conversion script	1
Rgamble	2
Slrubenstein	12
Bryan Derksen	1
Maveric149	5
Vicki Rosenzweig	1
Josh Grosse	1
Robert Merkel	1
PierreAbbat	1
Fredbauder	1
Tjmz	1
Gog	4
Emmett	3
Q	1
Jaknouse	5
Camembert	1
Graft	3
Dwmyers	1
AdamRetchless	17
Chas zzz brown	2
Heron	1
Ryguasu	1
The Anome	1
Alan Peakall	1
Smelialichu	1
Eloquence	7
Matusz	1
Cyde	1
RK	3
Fred Bauder	1
JWSchmidt	1
Michael	1
Rlowry	1
Zoe	1
Lexor	13
Someone else	1
Tannin	1
Zundark	1
Jimfbleak	1
Mirose	1
Dante Alighieri	1

Therefore, over time, the types of organisms that have traits better adapted to their environment will tend to become the dominant ones in an environment, while organisms poorly adapted to their environment will become extinct. Natural selection also provides for a mechanism by which life can sustain itself over time. Since, in the long run, environments always change, if successive generations did not develop adaptations which allowed them to survive and reproduce, species would simply die out as their biological niches die out. Therefore, life is allowed to persist over great spans of time, in the form of evolving species. The central role of natural selection in evolutionary theory has created a strong connection between that field and the study of ecology.

Genetic drift

Genetic drift describes changes in gene frequency that cannot be ascribed to selective pressures, but are due instead to events that are unrelated to inherited traits. This is especially important insmall mating populations, which simply cannot have enough offspring to maintain the same gene distribution as the parental generation. Such fluctuations in gene frequency between successive generations may result in some genes disappearing from the population. Two separate populations that begin with the same gene frequency might, therefore, "drift" by random fluctuation into two divergent populations with different gene sets (i.e. genes that are present in one have been lost in the other). Rare sporadic events (volcanic explosion, meteor impact, etc.) might contribute to genetic drift by altering the gene frequency outside of "normal" selective pressures.

Development of evolutionary theories

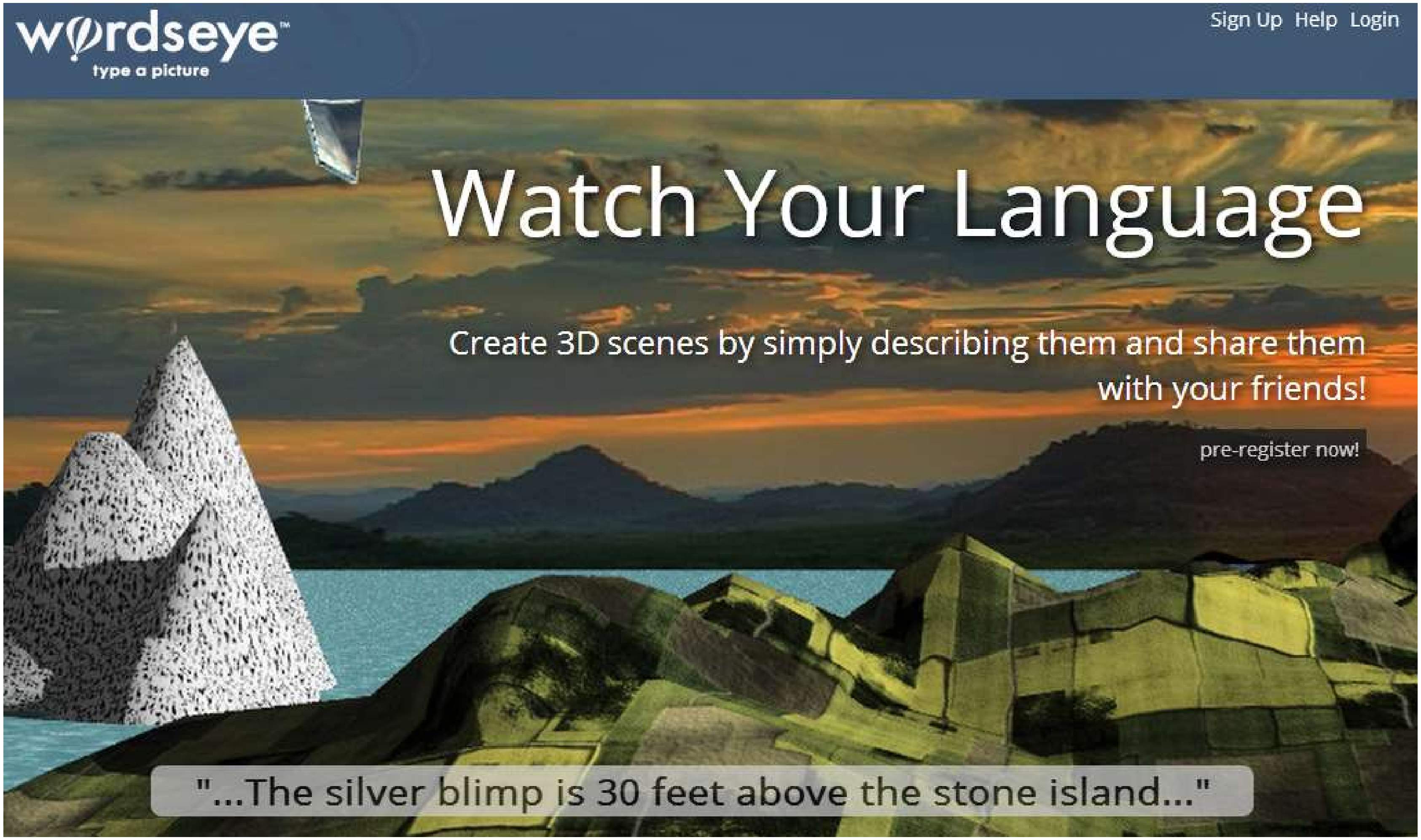
As science has uncovered more and more information about the basic operations of life, such as genetics and molecular biology, theories of evolution have changed. The general trend has been not to overturn well-supported theories, but to supplant them with more detailed and therefore more complex ones.

While transmutation was accepted by a sizeable number of scientists before 1859, it was the publication of Charles Darwin's The Origin of Species which provided the first cogent mechanism by which evolutionary change could persist: his mechanism of natural selection. The evolutionary timeline outlines the major steps of evolution on Earth as expounded by this theory's proponents.

Following the dawn of molecular biology, it became clear that a major mechanism for variation within a population is the mutagenesis of DNA. An essential component to evolutionary theory is that during the cell cycle, DNA is copied fairly, but not entirely, faithfully. When these rare copying errors occur, they are said to introduce genetic mutations of three general consequences relative to the current environment: good, bad, or neutral. By definition, individuals with "good" mutations will have an a stronger propensity to propagate, individuals with "bad" mutations will have less of a chance at successful reproduction, and those carrying "neutral" mutations will have neither an advantage nor a disadvantage. These definitions assume that the environment remains stable. Considered at the level of a single gene, these variations just described represent different genetic alleles. Following environmental change, alleles may retain their classification of good, bad, or neutral, or may shift into one of the other categories. Individuals carrying alleles formerly classified as neutral may now be "good" as they bear favourably adaptive

Fernanda B. Viégas, Martin Wattenberg, and Kushal Dave. 2004. Studying cooperation and conflict between authors with *history flow* visualizations. In Proc. *CHI* '04.

WordsEye.com



wordseye™ type a picture

Sign Up Help Login

Watch Your Language

Create 3D scenes by simply describing them and share them with your friends!

pre-register now!

"...The silver blimp is 30 feet above the stone island..."

SENTIMENT VISUALIZATION

Sentiment Analysis

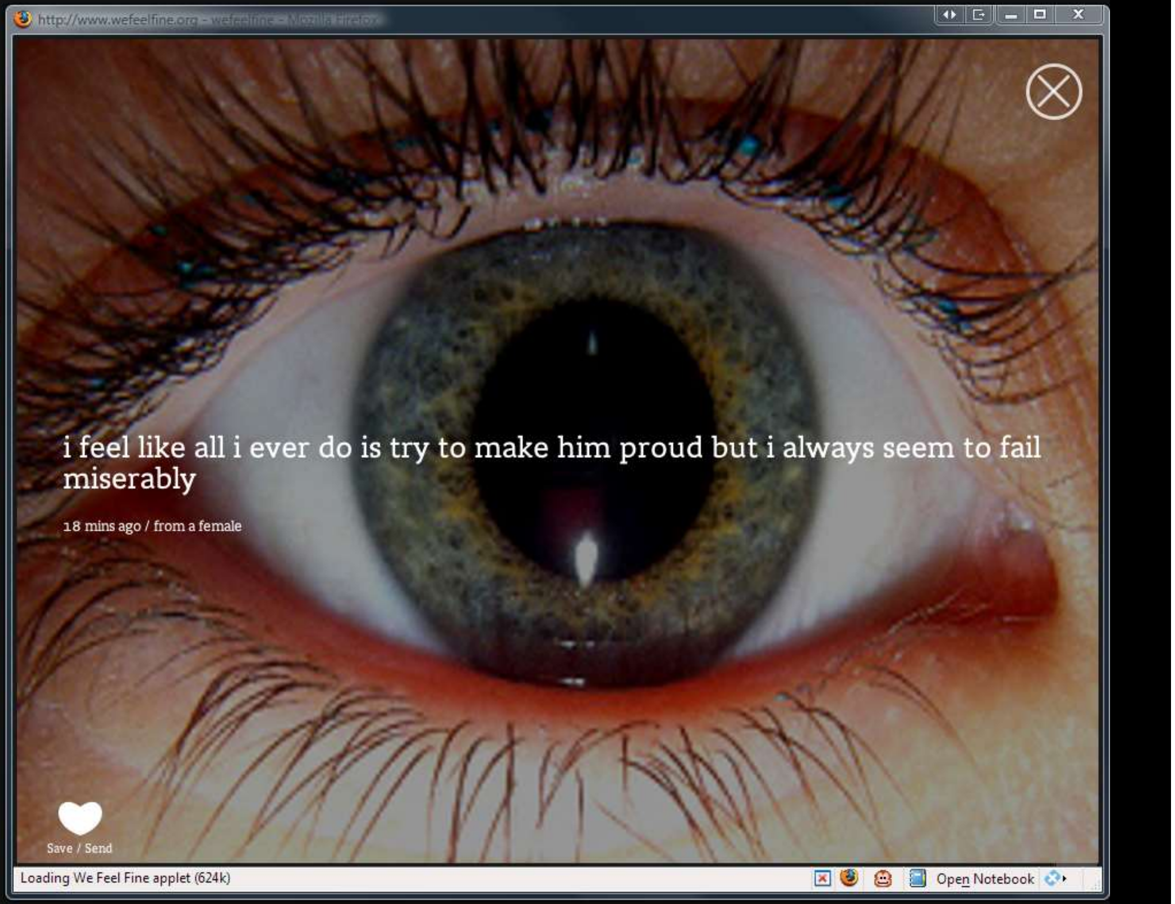
- Business intelligence:
  - Do people like my product/restaurant/movie/hotel?
  - Why or why not?
- Forensics and medicine:
  - State of mind analysis based on social media
- Emotional profiling / psycholinguistics
  - Understanding users -> individualization
  - Targeted advertising

Sentiment Analysis

- Language Processing:
  - Stemming
  - POS Tagging
  - Dependency Parsing
  - Named Entity Detection
- Granularity:
  - Positive/negative/uncertain
  - 8+ emotions
  - Word, sentence, paragraph, document, corpus level

Resources and Datasets

- NRC Word-Emotion Lexicon:
  - Saif Mohammad, 2013
    http://www.saifmohammad.com/WebPages/ResearchInterests.html
- LIWC:
  - James Pennybaker et al., 2007: http://www.liwc.net/
- Opinion Mining Dataset:
  - Bing Liu, 2004—current
    http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html


http://www.wefeelfine.org - wefeelfine - Mozilla Firefox

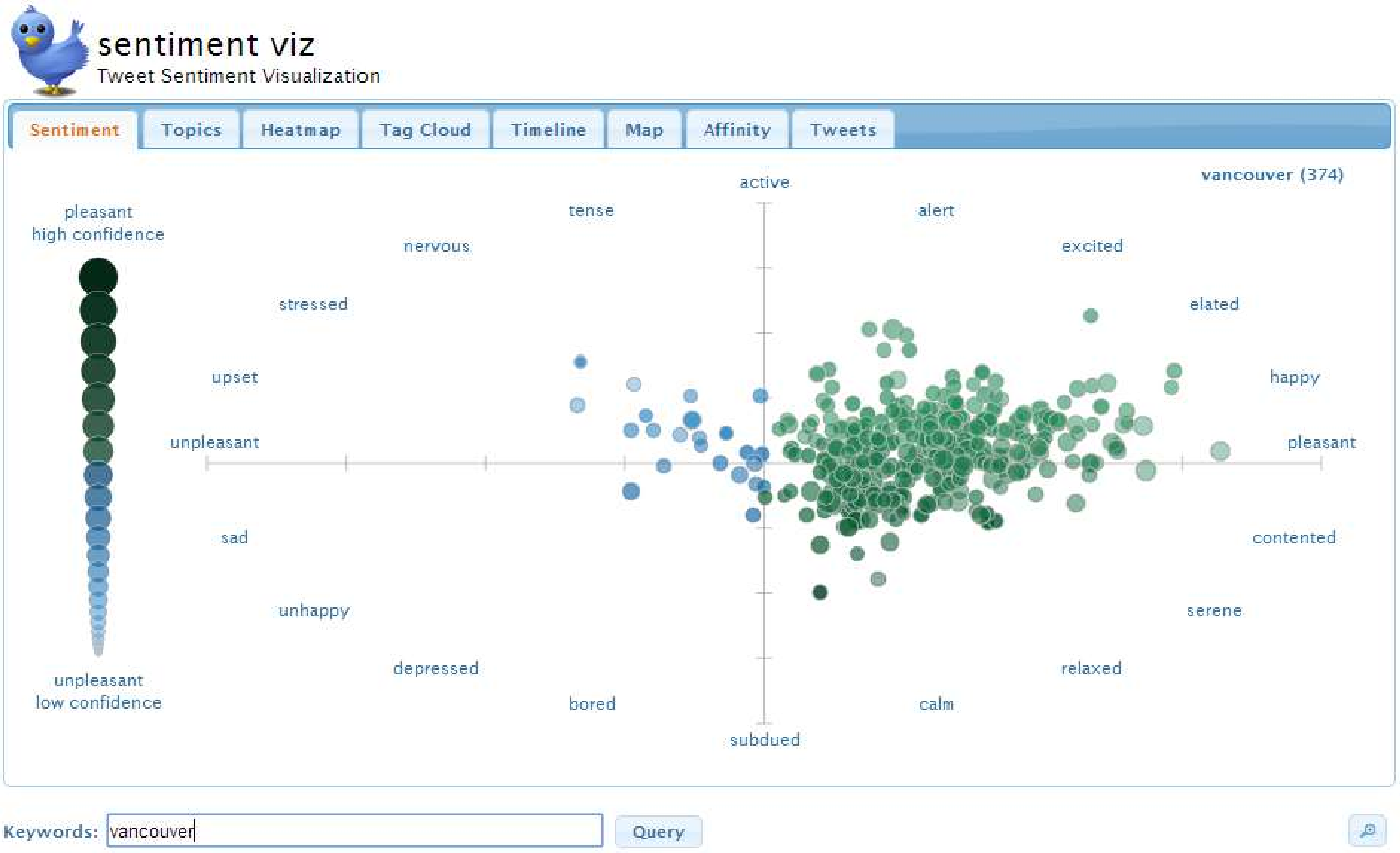
i feel like all i ever do is try to make him proud but i always seem to fail miserably

18 mins ago / from a female

Save / Send
Loading We Feel Fine applet (624k)
Open Notebook

(Harris, 2006)

Twitter Sentiment Viz



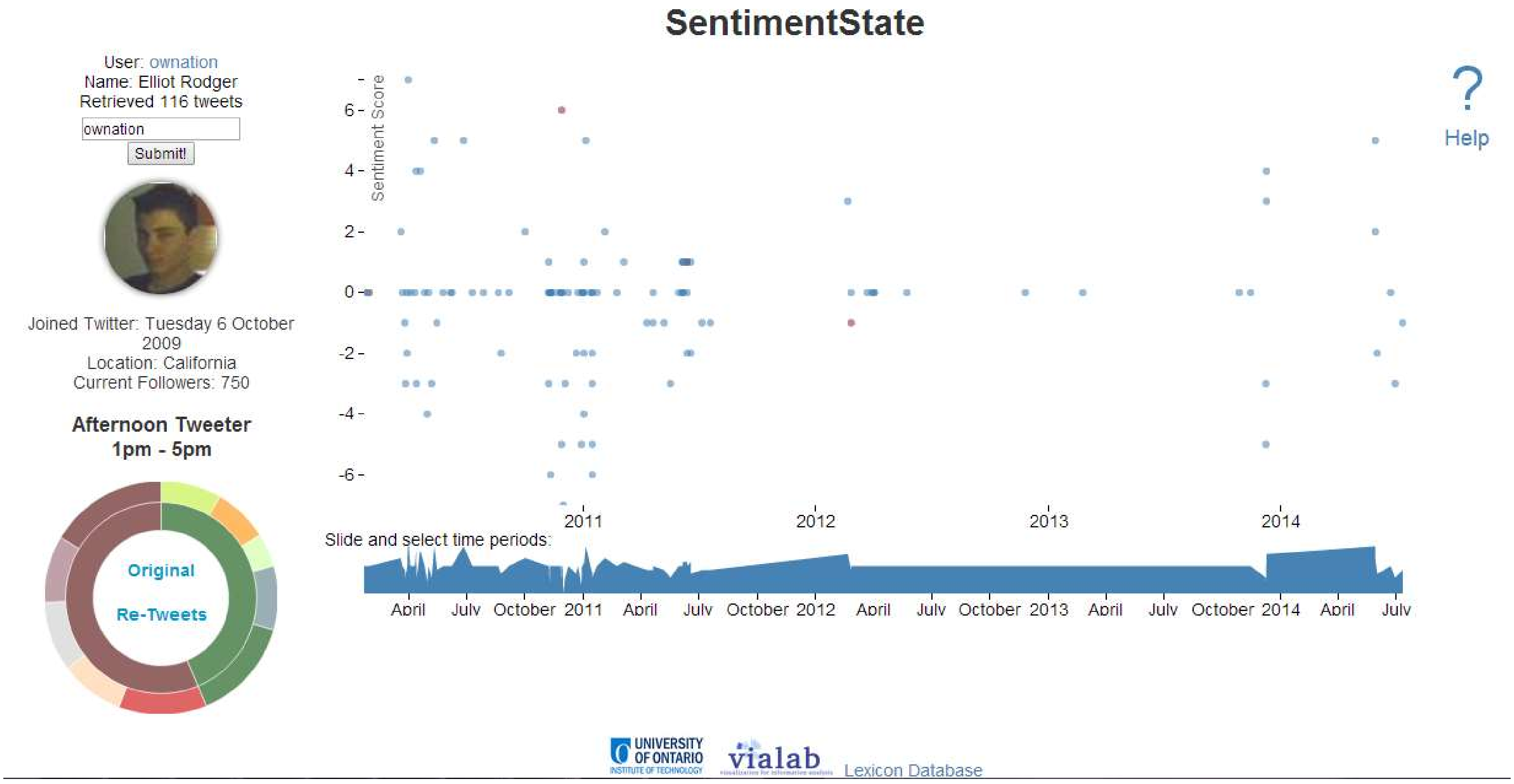
Healey and Ramaswamy, 2013. http://www.csc.ncsu.edu/faculty/healey/tweet_viz/

SentimentState

- Tweets over time, categorized using an emotion lexicon
- Examine Tweets in context, filter based on time and emotions

Scantlebury and Collins, 2014. http://vialab.science.uoit.ca/sentimentstate

SentimentState



SentimentState

User: ownation
Name: Elliot Rodger
Retrieved 116 tweets

ownation

Submit!

Joined Twitter: Tuesday 6 October 2009
Location: California
Current Followers: 750

Afternoon Tweeter
1pm - 5pm

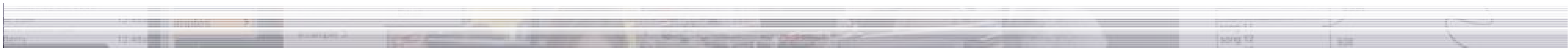
Slide and select time periods:

Lexicon Database

Scantlebury and Collins, 2014. http://vialab.science.uoit.ca/sentimentstate

This movie was actually neither that funny, nor super witty.

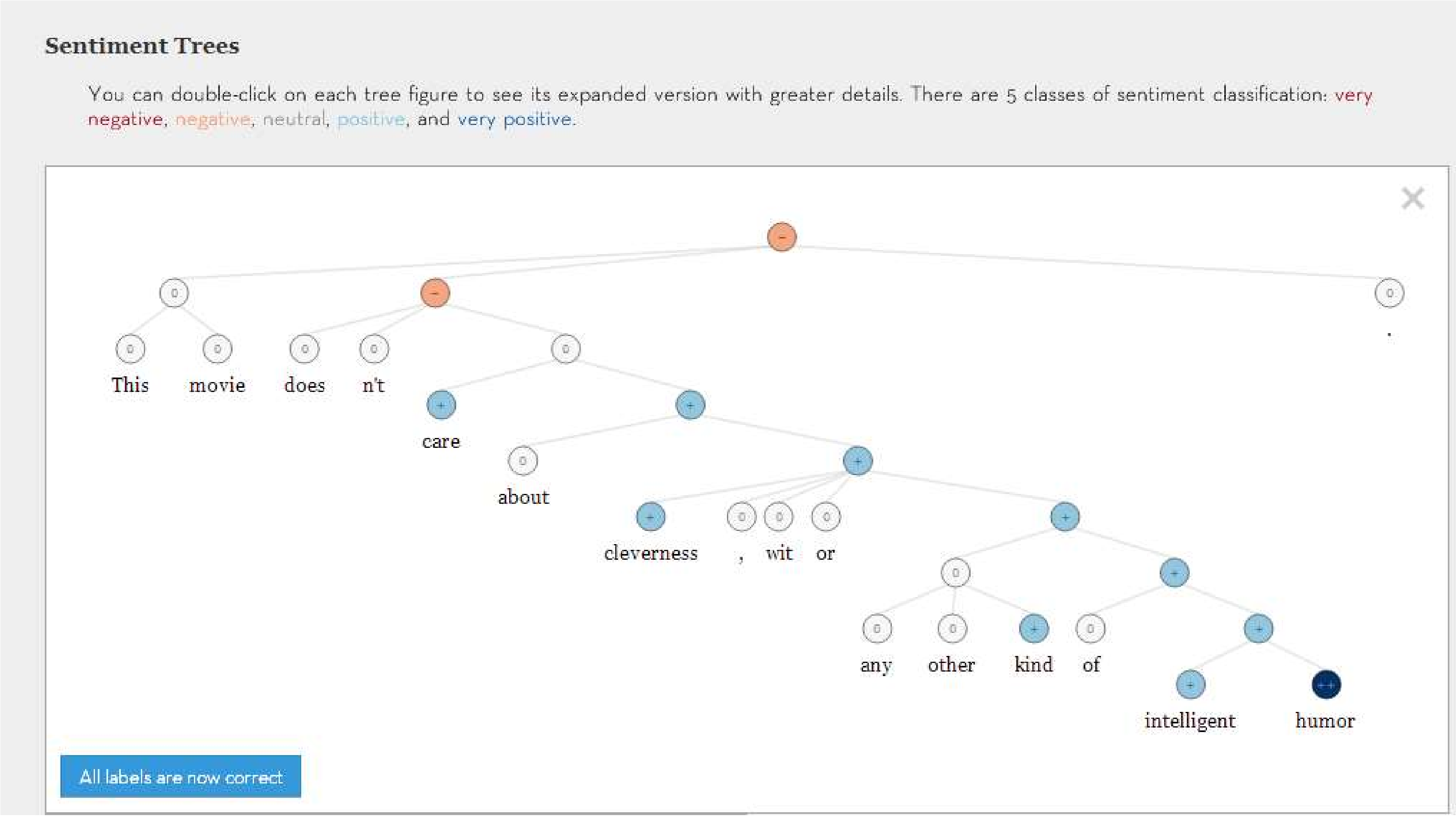
This movie was actually neither that funny, nor super witty.



Stanford Sentiment Parser

- Recursive neural network built on top of grammatical structures
- Trained on Stanford Sentiment Treebank
  - Parse trees labelled with sentiment scores
  - Crowed-sourced and editable

Socher et al. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. Conference on Empirical Methods in Natural Language Processing (EMNLP 2013).

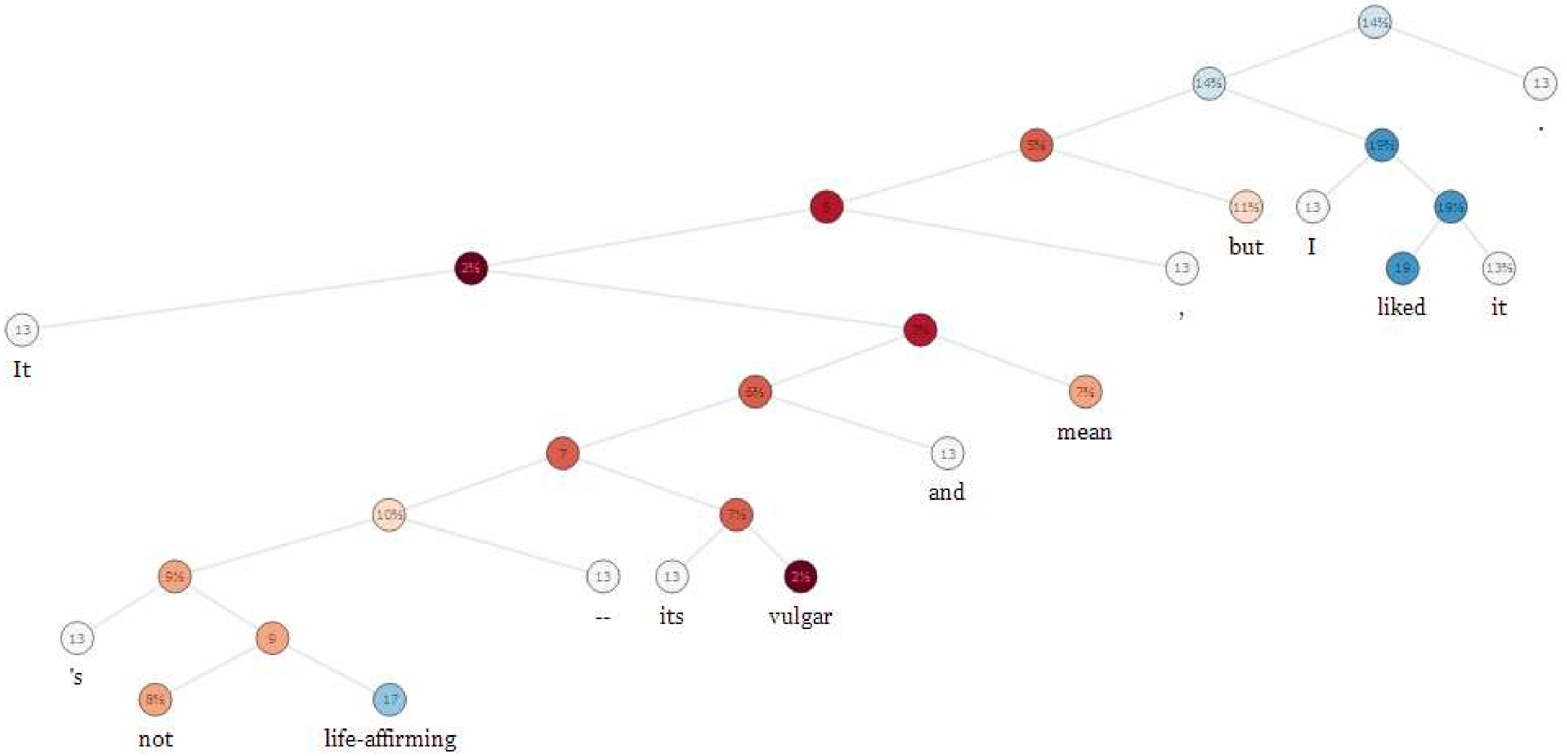


Sentiment Trees

You can double-click on each tree figure to see its expanded version with greater details. There are 5 classes of sentiment classification: very negative, negative, neutral, positive, and very positive.

Parsing is Needed!

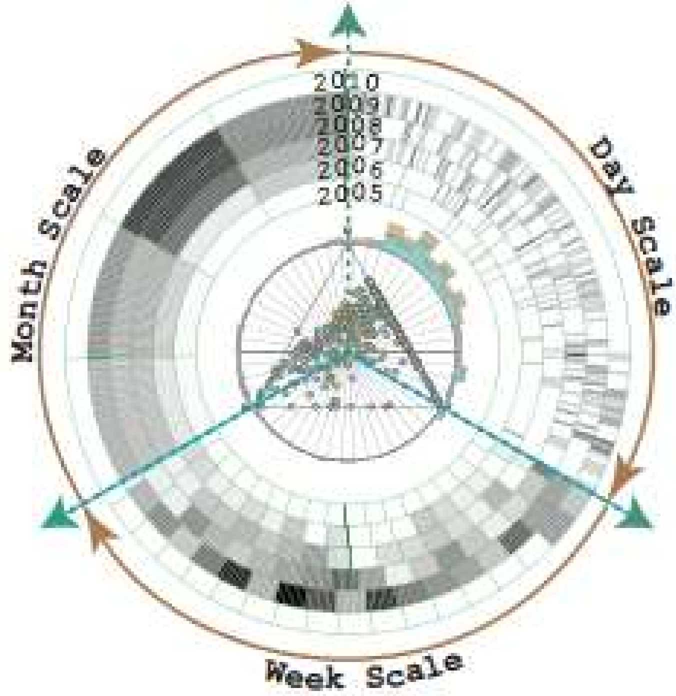
- Stanford Sentiment Treebank:
  - http://nlp.stanford.edu/sentiment/treebank.html



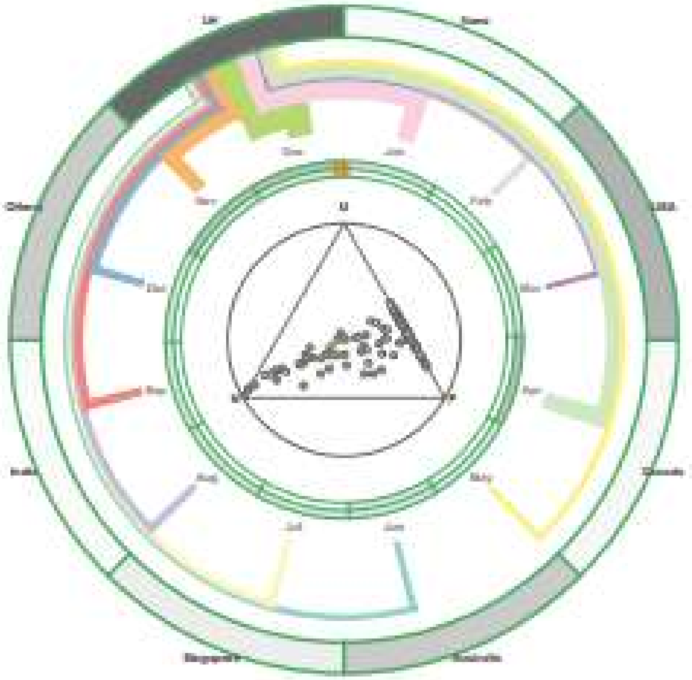
Challenges

- Word-counting techniques are fast, but inaccurate
  - Sarcasm, quotes, metaphorical language
- Accurate methods are slow/difficult to run over big data

Opinion Seer



(a)



(b)

Yingcai Wu et al. 2010. OpinionSeer: Interactive Visualization of Hotel Customer Feedback. *IEEE Transactions on Visualization and Computer Graphics* 16 (6), November 2010.

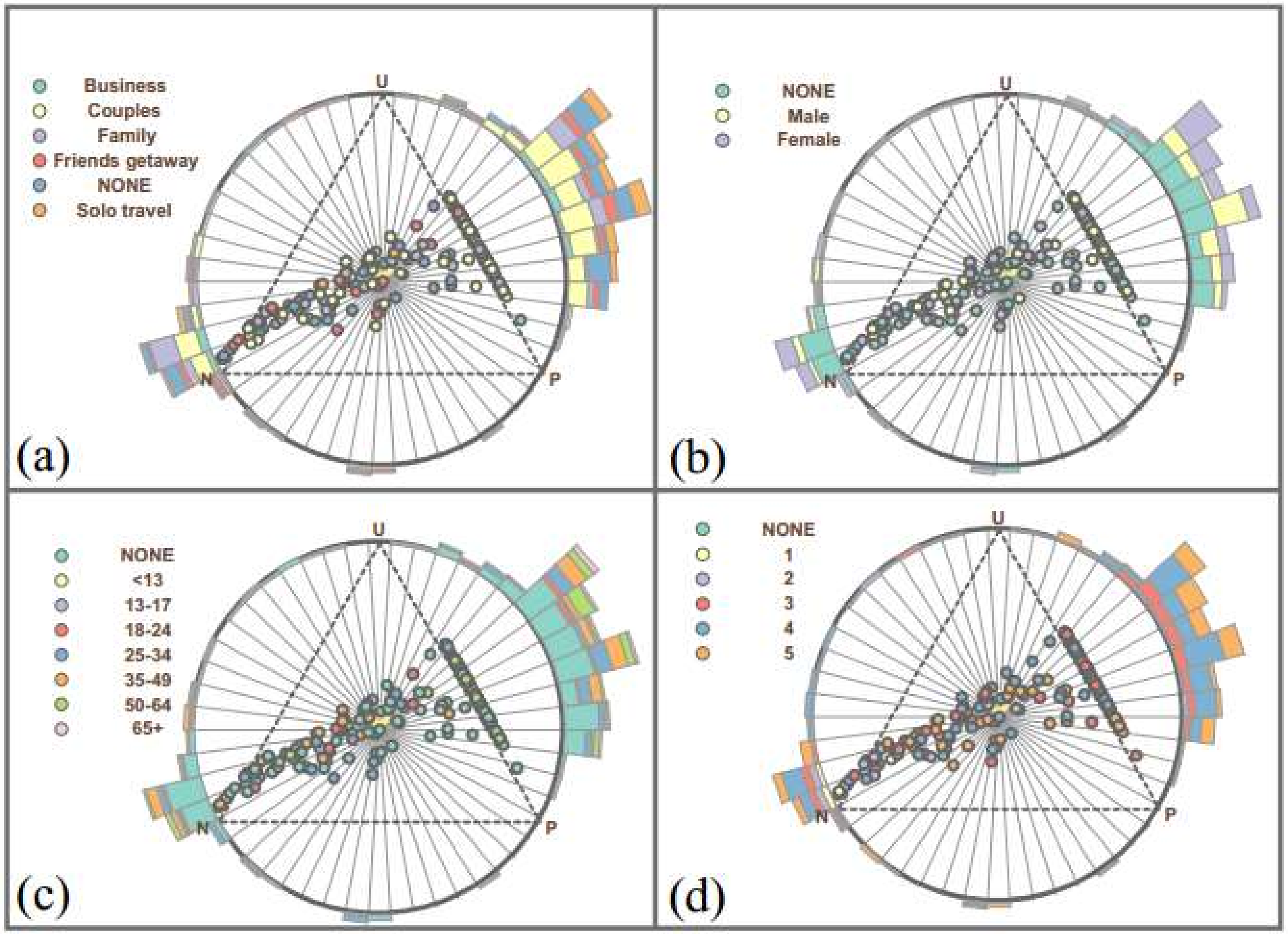
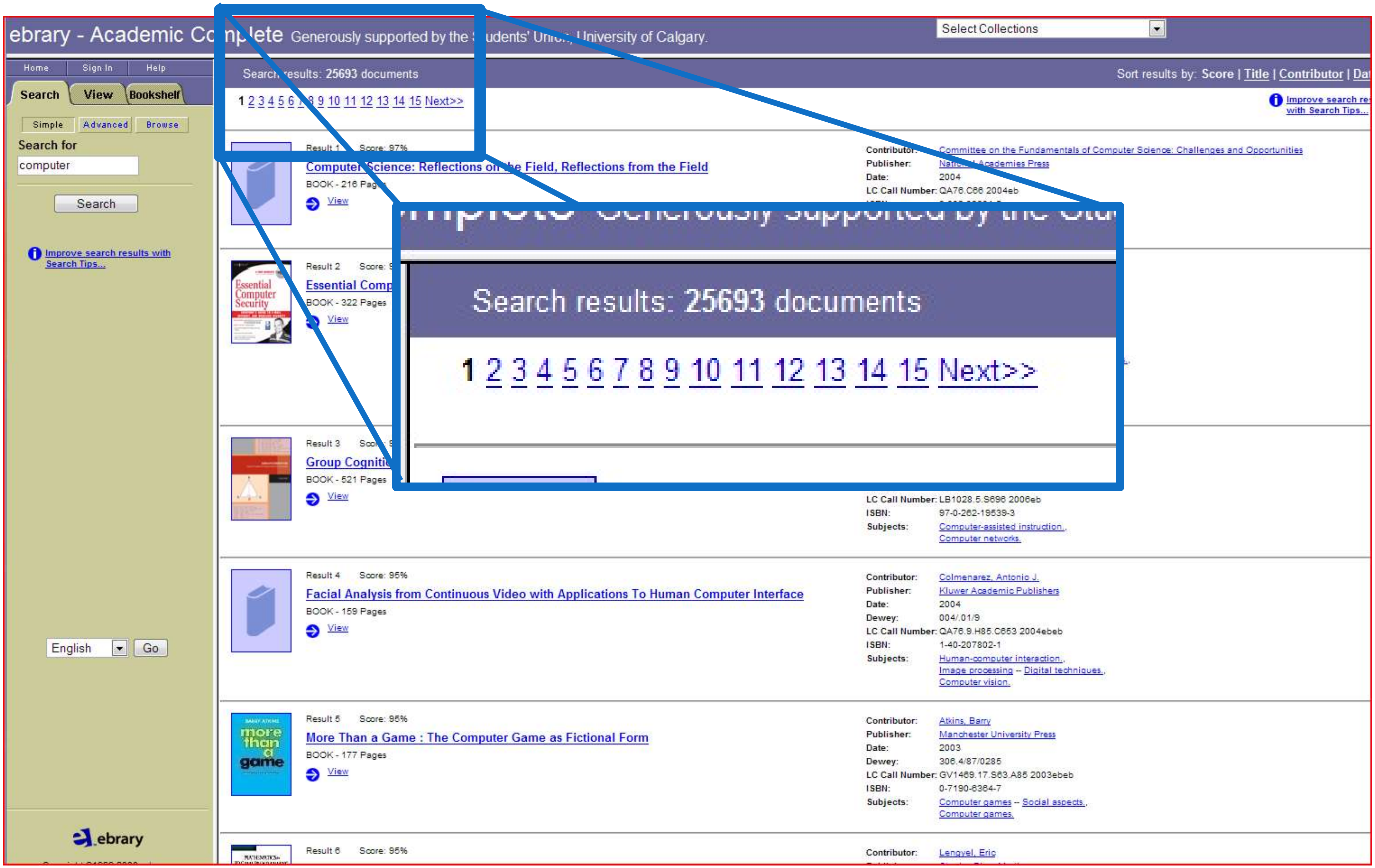


Fig. 8. OpinionSeer results showing how customer opinions are correlated with trip type, gender, age range, and ratings.

INFORMATION RETRIEVAL



ebrary - Academic Complete Generously supported by the Students' Union, University of Calgary.

Search for: computer

Search results: 25693 documents

Sort results by: Score | Title | Contributor | Dat

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 Next>>

Result 1 Score: 97%

Computer Science: Reflections on the Field, Reflections from the Field

BOOK - 216 Pages

View

Contributor: Committee on the Fundamentals of Computer Science: Challenges and Opportunities
Publisher: National Academies Press
Date: 2004
LC Call Number: QA76.C66 2004eb

Result 2 Score: 9

Essential Comp

BOOK - 322 Pages

View

Result 3 Score: 9

Group Cogniti

BOOK - 521 Pages

View

LC Call Number: LB1028.5.S696 2006eb
ISBN: 97-0-262-19539-3
Subjects: Computer-assisted instruction., Computer networks.

Result 4 Score: 95%

Facial Analysis from Continuous Video with Applications To Human Computer Interface

BOOK - 159 Pages

View

Contributor: Colmenarez, Antonio J.
Publisher: Kluwer Academic Publishers
Date: 2004
Dewey: 004/.01/9
LC Call Number: QA76.9.H85.C653 2004ebeb
ISBN: 1-40-207802-1
Subjects: Human-computer interaction., Image processing -- Digital techniques., Computer vision.

Result 5 Score: 95%

More Than a Game : The Computer Game as Fictional Form

BOOK - 177 Pages

View

Contributor: Atkins, Barry
Publisher: Manchester University Press
Date: 2003
Dewey: 306.4/87/0285
LC Call Number: GV1469.17.S63.A85 2003ebeb
ISBN: 0-7190-6364-7
Subjects: Computer games -- Social aspects., Computer games.

Result 6 Score: 95%

Contributor: Lengyel, Eric

Search results: 25693 documents

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 Next>>

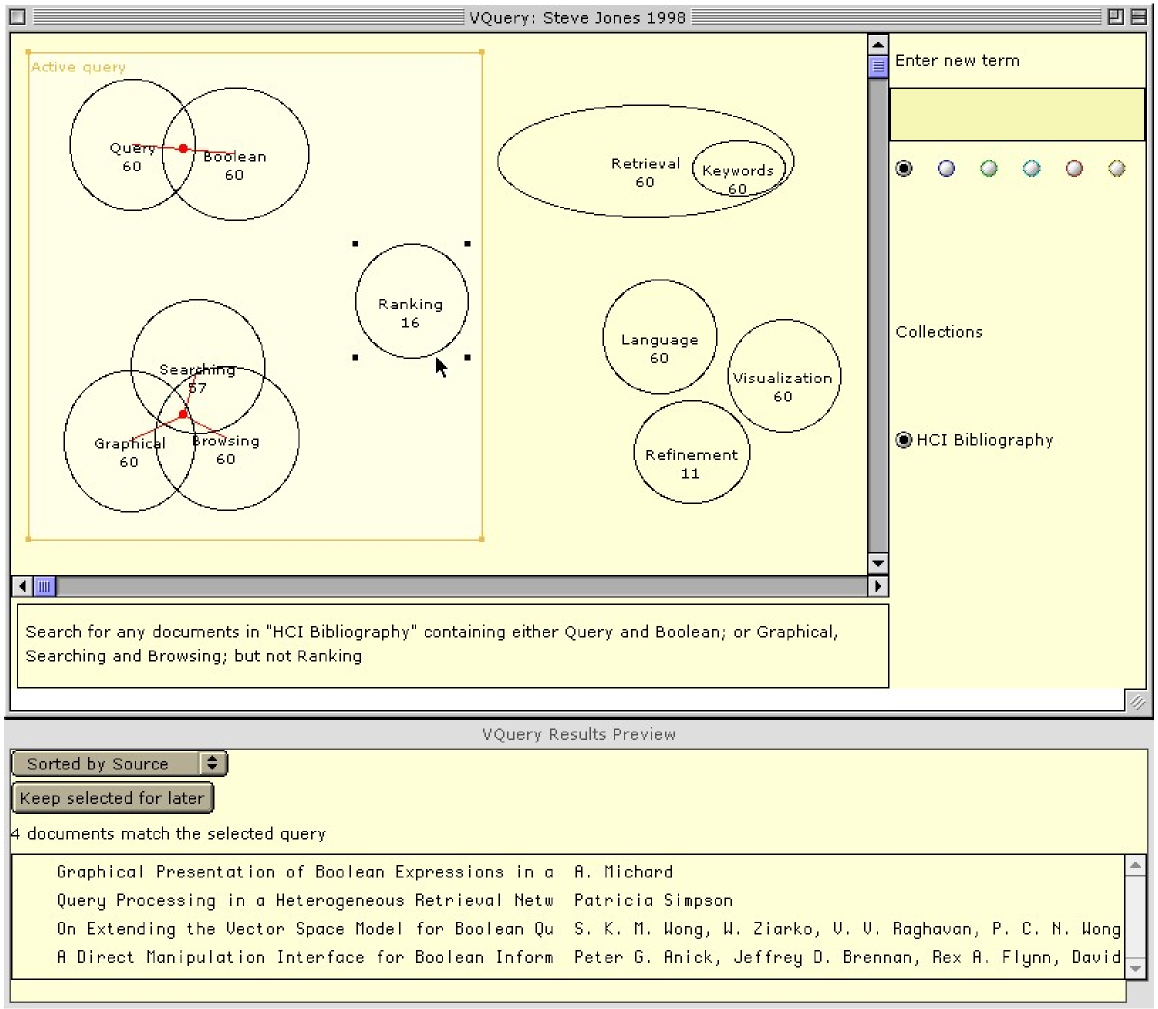
Information Retrieval

- Visual query formation
- Exploration of collections
- Single/comparative document content visualization

Visual Query Formation

- Rich specification of linguistic constraints

VQuery



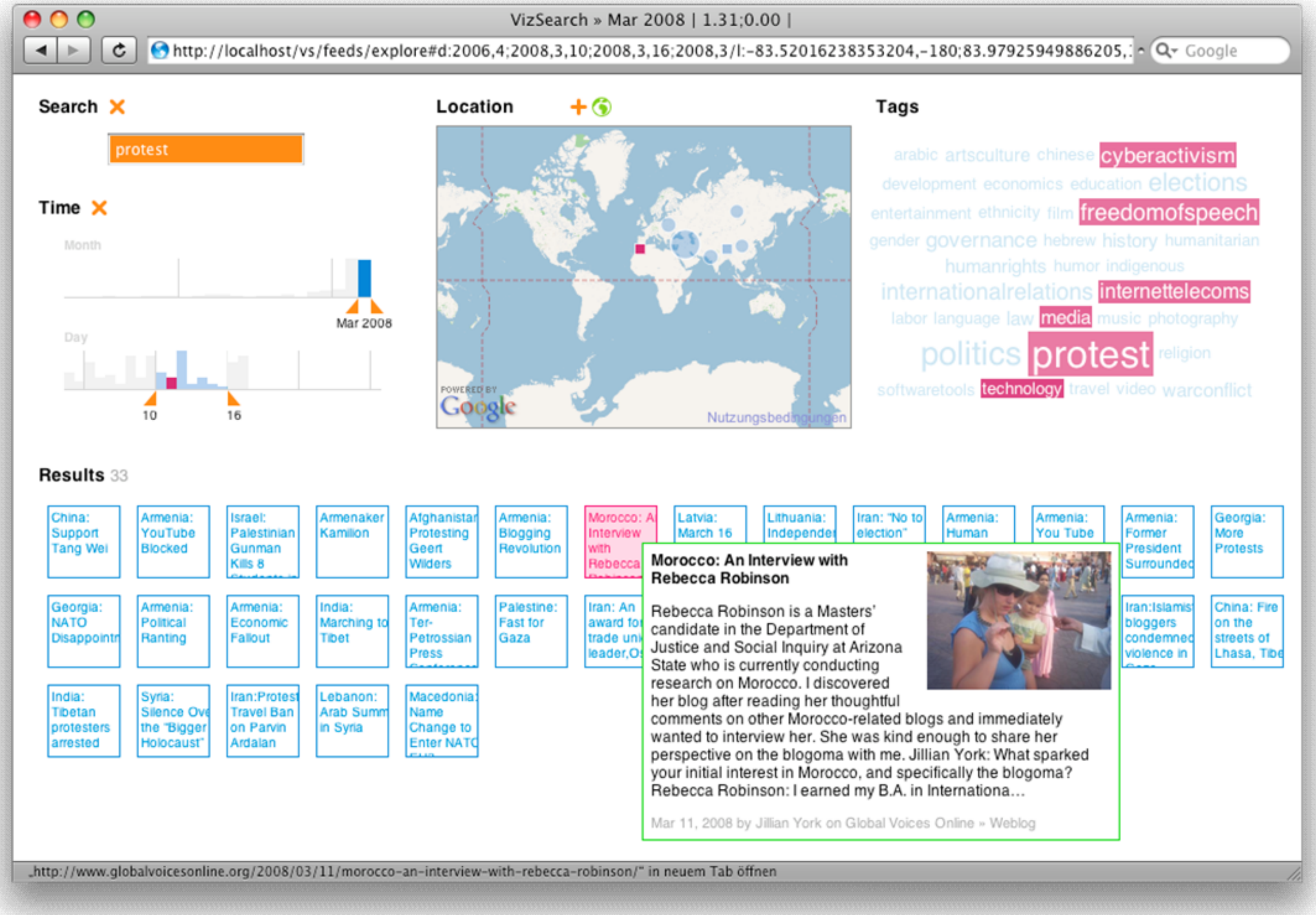
◉ HCI Bibliography

Search for any documents in "HCI Bibliography" containing either Query and Boolean; or Graphical, Searching and Browsing; but not Ranking

4 documents match the selected query

```
Graphical Presentation of Boolean Expressions in a  A. Michard
Query Processing in a Heterogeneous Retrieval Netw  Patricia Simpson
On Extending the Vector Space Model for Boolean Qu  S. K. M. Wong, W. Ziarko, V. V. Raghavan, P. C. N. Wong
A Direct Manipulation Interface for Boolean Inform  Peter G. Anick, Jeffrey D. Brennan, Rex A. Flynn, David
```

VisGets



VizSearch » Mar 2008 | 1.31;0.00 |

http://localhost/vs/feeds/explore#d:2006,4;2008,3,10;2008,3,16;2008,3/l:-83.52016238353204,-180;83.97925949886205,

Google
Search ✕
protest
Time ✕
Month
Mar 2008
Day
10
16
Location
Tags
arabic artsculture chinese cyberactivism
development economics education elections
entertainment ethnicity film freedomofspeech
gender governance hebrew history humanitarian
humanrights humor indigenous
internationalrelations internettelecoms
labor language law media music photography
politics protest religion
softwaretools technology travel video warconflict
POWERED BY Google
Nutzungsbedingungen
Results 33
China: Support Tang Wei
Armenia: YouTube Blocked
Israel: Palestinian Gunman Kills 8
Armenaker Kamilion
Afghanistan Protesting Geert Wilders
Armenia: Blogging Revolution
Morocco: An Interview with Rebecca
Latvia: March 16
Lithuania: Independen
Iran: "No to election"
Armenia: Human
Armenia: You Tube
Armenia: Former President Surrounded
Georgia: More Protests
Georgia: NATO Disappointm
Armenia: Political Ranting
Armenia: Economic Fallout
India: Marching to Tibet
Armenia: Ter-Petrossian Press
Palestine: Fast for Gaza
Iran: An award for trade uni leader,O
Iran:Islamist bloggers condemned violence in
China: Fire on the streets of Lhasa, Tibe
India: Tibetan protesters arrested
Syria: Silence Ove the "Bigger Holocaust"
Iran:Protest Travel Ban on Parvin Ardalan
Lebanon: Arab Summ in Syria
Macedonia: Name Change to Enter NATO

Morocco: An Interview with Rebecca Robinson

Rebecca Robinson is a Masters' candidate in the Department of Justice and Social Inquiry at Arizona State who is currently conducting research on Morocco. I discovered her blog after reading her thoughtful comments on other Morocco-related blogs and immediately wanted to interview her. She was kind enough to share her perspective on the blogoma with me. Jillian York: What sparked your initial interest in Morocco, and specifically the blogoma? Rebecca Robinson: I earned my B.A. in Internationa…

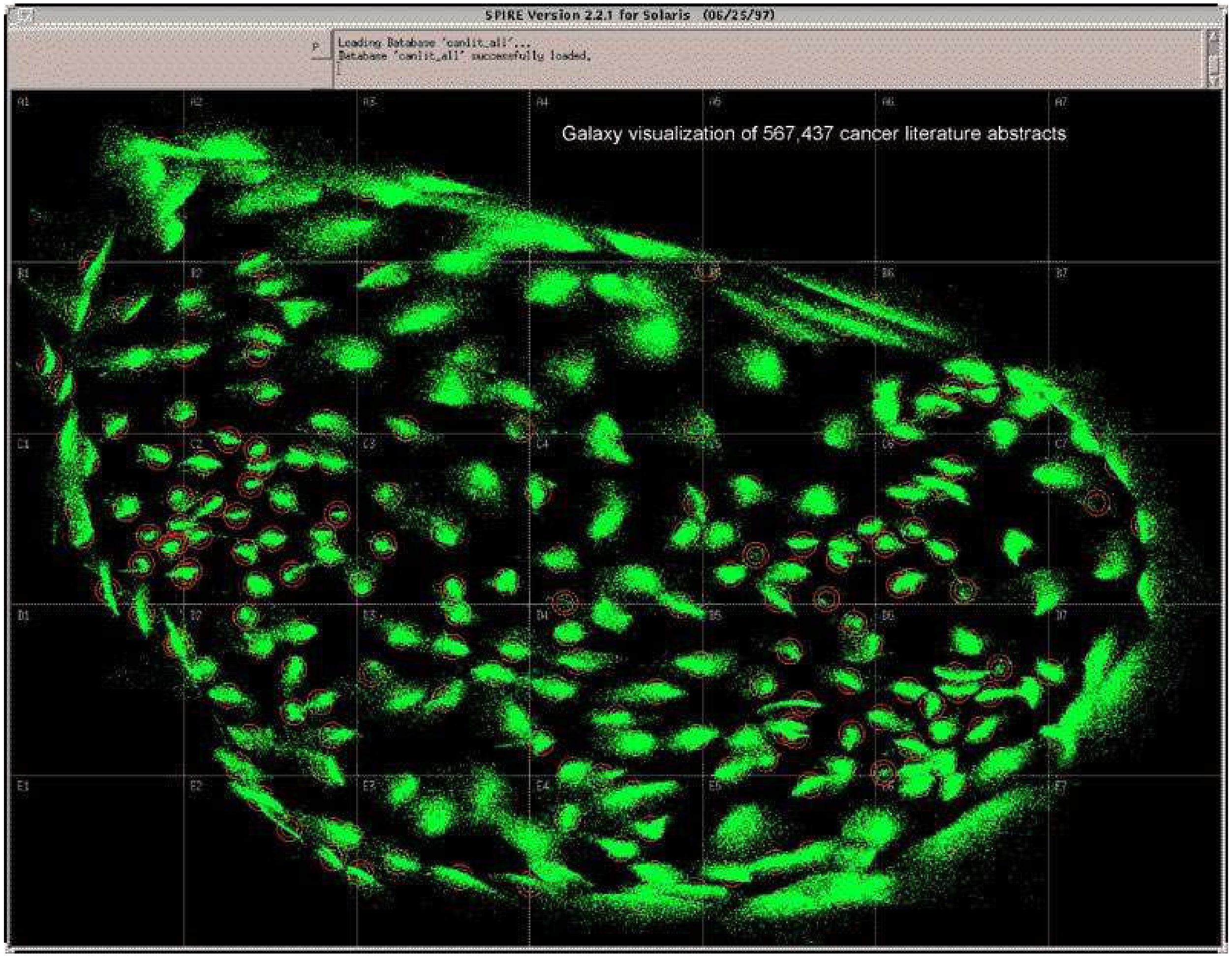
Mar 11, 2008 by Jillian York on Global Voices Online » Weblog

„http://www.globalvoicesonline.org/2008/03/11/morocco-an-interview-with-rebecca-robinson/" in neuem Tab öffnen

Exploration of Collections

- Provide overview of:
  - entire collection
  - subset matching a query
- Clustering and categorization

Galaxies

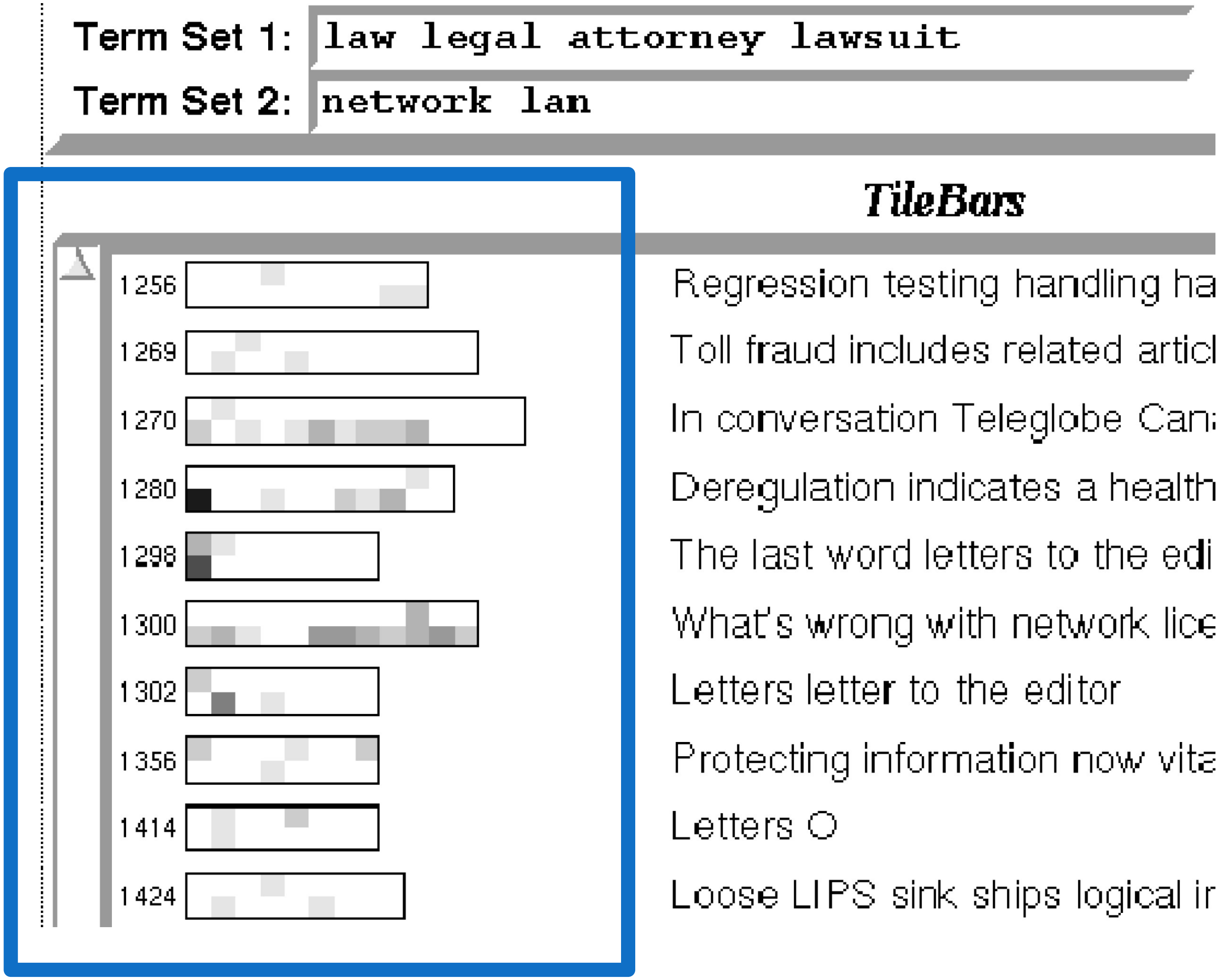


SPIRE Version 2.2.1 for Solaris (06/25/97)

```
Loading Database 'canlit_all'...
Database 'canlit_all' successfully loaded.
```

Galaxy visualization of 567,437 cancer literature abstracts

Tile Bars



Term Set 1: law legal attorney lawsuit

Term Set 2: network lan

TileBars

1256 Regression testing handling ha

1269 Toll fraud includes related articl

1270 In conversation Teleglobe Can

1280 Deregulation indicates a health

1298 The last word letters to the edi

1300 What's wrong with network lice

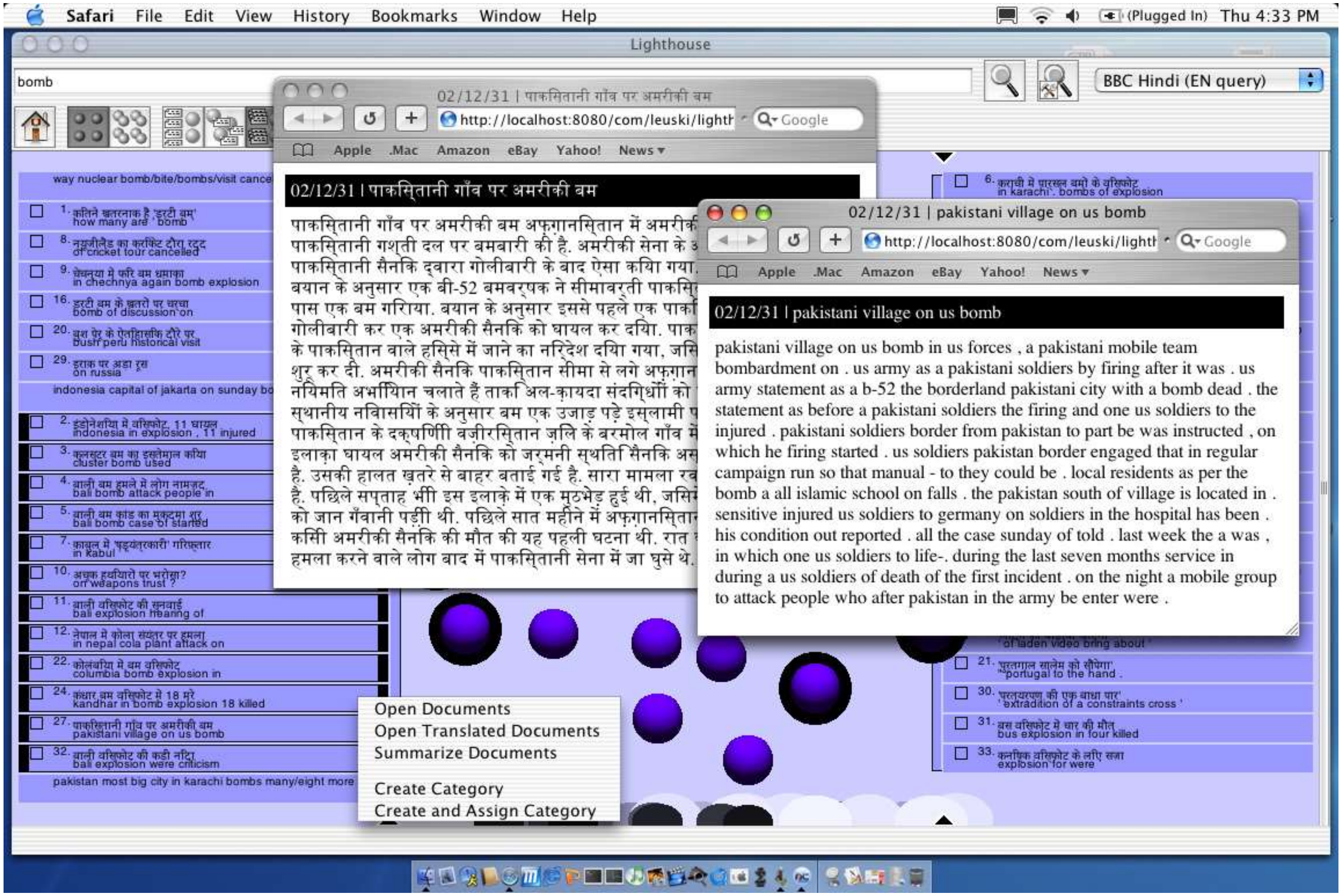
1302 Letters letter to the editor

1356 Protecting information now vita

1414 Letters O

1424 Loose LIPS sink ships logical ir

Lighthouse Cross-Lingual Search


Safari File Edit View History Bookmarks Window Help
(Plugged In) Thu 4:33 PM
Lighthouse
bomb
BBC Hindi (EN query)
Apple .Mac Amazon eBay Yahoo! News
Open Documents
Open Translated Documents
Summarize Documents
Create Category
Create and Assign Category

way nuclear bomb/bite/bombs/visit cance

□ 1. कतिने खतरनाक है 'डरटी बम'
how many are ' bomb

□ 8. न्यूज़ीलैंड का करकिेट दौरा रद्द
of cricket tour cancelled

□ 9. चेचन्या मे फरि बम धमाका
in chechnya again bomb explosion

□ 16. डरटी बम के खतरो पर चरचा
bomb of discussion on

□ 20. बुश पेरू के ऐतहािसकि दौरे पर
bush peru historical visit

□ 29. इराक पर अड़ा रूस
on russia

indonesia capital of jakarta on sunday bo

□ 2. इंडोनेशयिा मे वसि्फोट, 11 घायल
indonesia in explosion , 11 injured

□ 3. क्लस्टर बम का इस्तेमाल कयिा
cluster bomb used

□ 4. बाली बम हमले मे लोग नामज़द,
bali bomb attack people in

□ 5. बाली बम कांड का मुकदमा शुरू
bali bomb case of started

□ 7. काबुल मे 'षड्यंत्रकारी' गरिफ़्तार
in kabul

□ 10. अचूक हथयािरो पर भरोसा?
on weapons trust ?

□ 11. बाली वसि्फोट की सुनवाई
bali explosion hearing of

□ 12. नेपाल मे कोला संयंत्र पर हमला
in nepal cola plant attack on

□ 22. कोलंबयिा मे बम वसि्फोट
columbia bomb explosion in

□ 24. कंधार बम वसि्फोट मे 18 मरे
kandhar in bomb explosion 18 killed

□ 27. पाकसि्तानी गाँव पर अमरीकी बम
pakistani village on us bomb

□ 32. बाली वसि्फोट की कड़ी नदिा
bali explosion were criticism

pakistan most big city in karachi bombs many/eight more

02/12/31 | पाकसि्तानी गाँव पर अमरीकी बम

http://localhost:8080/com/leuski/lighth

02/12/31 | पाकसि्तानी गाँव पर अमरीकी बम

पाकसि्तानी गाँव पर अमरीकी बम अफ़ग़ानसि्तान में अमरीकि
पाकसि्तानी गश्ती दल पर बमबारी की है. अमरीकी सेना के अ
पाकसि्तानी सैनकि द्वारा गोलीबारी के बाद ऐसा कयिा गया.
बयान के अनुसार एक बी-52 बमवर्षक ने सीमावर्ती पाकसि्
पास एक बम गरिाया. बयान के अनुसार इससे पहले एक पाकि
गोलीबारी कर एक अमरीकी सैनकि को घायल कर दयिा. पाक
के पाकसि्तान वाले हसि्से में जाने का नरि्देश दयिा गया, जसि
शुरू कर दी. अमरीकी सैनकि पाकसि्तान सीमा से लगे अफ़ग़ान
नयिमति अभयिान चलाते हैं ताक‌ि अल-क़ायदा संदग‌्धिों को
स्थानीय नविासयिों के अनुसार बम एक उजाड़ पड़े इस्लामी प
पाकसि्तान के दक्षणिी वज़ीरसि्तान ज़लि‌े के बरमोल गाँव मे
इलाक़ा घायल अमरीकी सैनकि को जर्मनी स्थति सैनकि अस्
है. उसकी हालत ख़तरे से बाहर बताई गई है. सारा मामला रव
है. पछिले सप्ताह भी इस इलाक़े में एक मुठभेड़ हुई थी, जसिमे
को जान गँवानी पड़ी थी. पछिले सात महीने में अफ़ग़ानसि्तान
कसिी अमरीकी सैनकि की मौत की यह पहली घटना थी. रात
हमला करने वाले लोग बाद में पाकसि्तानी सेना में जा घुसे थे.

02/12/31 | pakistani village on us bomb

http://localhost:8080/com/leuski/lighth

02/12/31 | pakistani village on us bomb

pakistani village on us bomb in us forces , a pakistani mobile team bombardment on . us army as a pakistani soldiers by firing after it was . us army statement as a b-52 the borderland pakistani city with a bomb dead . the statement as before a pakistani soldiers the firing and one us soldiers to the injured . pakistani soldiers border from pakistan to part be was instructed , on which he firing started . us soldiers pakistan border engaged that in regular campaign run so that manual - to they could be . local residents as per the bomb a all islamic school on falls . the pakistan south of village is located in . sensitive injured us soldiers to germany on soldiers in the hospital has been . his condition out reported . all the case sunday of told . last week the a was , in which one us soldiers to life-. during the last seven months service in during a us soldiers of death of the first incident . on the night a mobile group to attack people who after pakistan in the army be enter were .

□ 6. कराची मे पारसल बमो के वसि्फोट
in karachi`. bombs of explosion

'of laden video bring about '

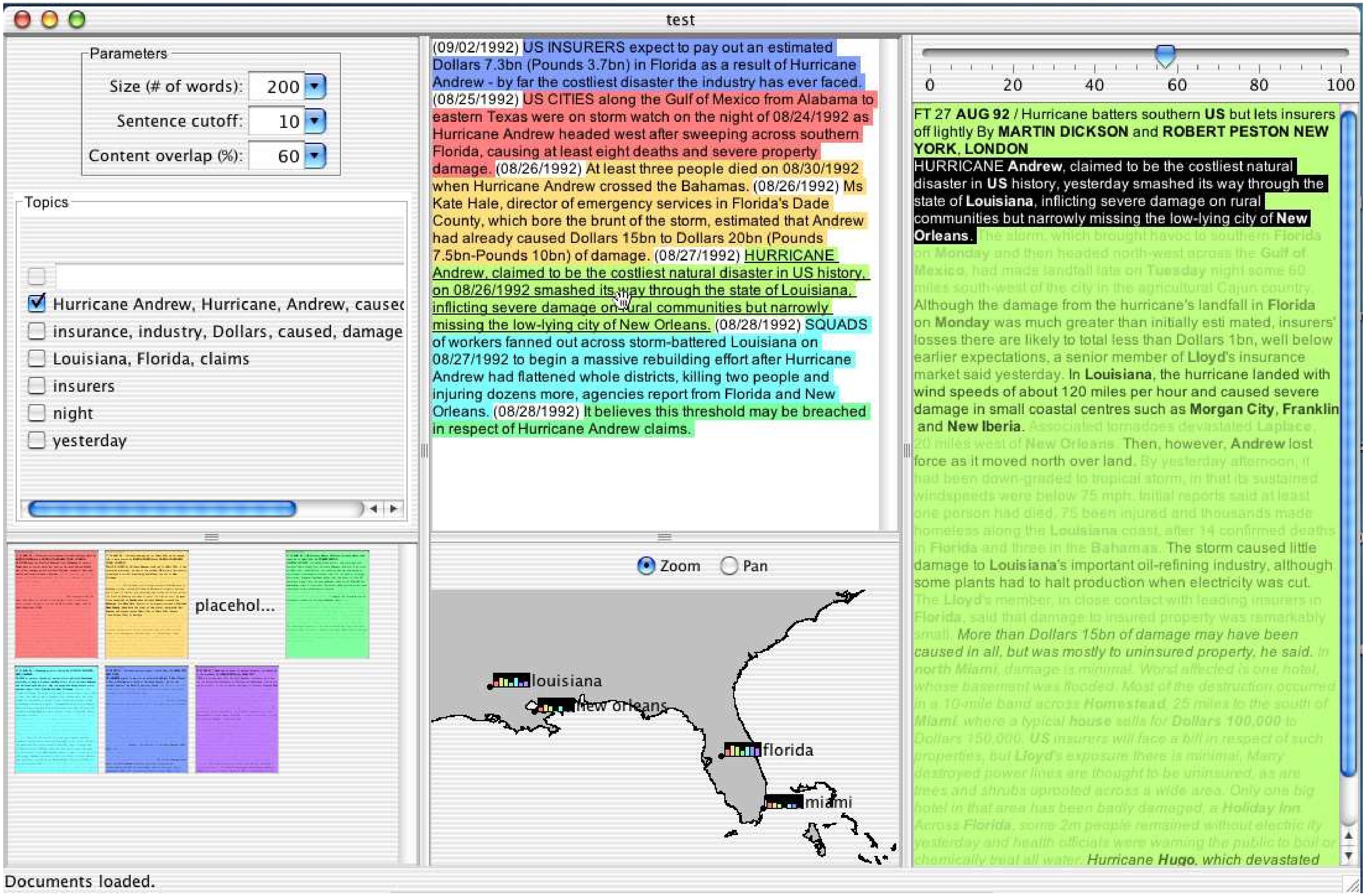
□ 21. पुरतगाल सालेम को सौंपेगा'
"portugal to the hand .

□ 30. प्रत्यरपण की एक बाधा पार'
' extradition of a constraints cross '

□ 31. बस वसि्फोट मे चार की मौत
bus explosion in four killed

□ 33. कनषि्क वसि्फोट के लएि सज़ा
explosion for were

iNeATS Summarizer



Parameters

Size (# of words): 200

Sentence cutoff: 10

Content overlap (%): 60

Topics

- ☑ Hurricane Andrew, Hurricane, Andrew, causec
- ☐ insurance, industry, Dollars, caused, damage
- ☐ Louisiana, Florida, claims
- ☐ insurers
- ☐ night
- ☐ yesterday

(09/02/1992) US INSURERS expect to pay out an estimated Dollars 7.3bn (Pounds 3.7bn) in Florida as a result of Hurricane Andrew - by far the costliest disaster the industry has ever faced. (08/25/1992) US CITIES along the Gulf of Mexico from Alabama to eastern Texas were on storm watch on the night of 08/24/1992 as Hurricane Andrew headed west after sweeping across southern Florida, causing at least eight deaths and severe property damage. (08/26/1992) At least three people died on 08/30/1992 when Hurricane Andrew crossed the Bahamas. (08/26/1992) Ms Kate Hale, director of emergency services in Florida's Dade County, which bore the brunt of the storm, estimated that Andrew had already caused Dollars 15bn to Dollars 20bn (Pounds 7.5bn-Pounds 10bn) of damage. (08/27/1992) HURRICANE Andrew, claimed to be the costliest natural disaster in US history, on 08/26/1992 smashed its way through the state of Louisiana, inflicting severe damage on rural communities but narrowly missing the low-lying city of New Orleans. (08/28/1992) SQUADS of workers fanned out across storm-battered Louisiana on 08/27/1992 to begin a massive rebuilding effort after Hurricane Andrew had flattened whole districts, killing two people and injuring dozens more, agencies report from Florida and New Orleans. (08/28/1992) It believes this threshold may be breached in respect of Hurricane Andrew claims.

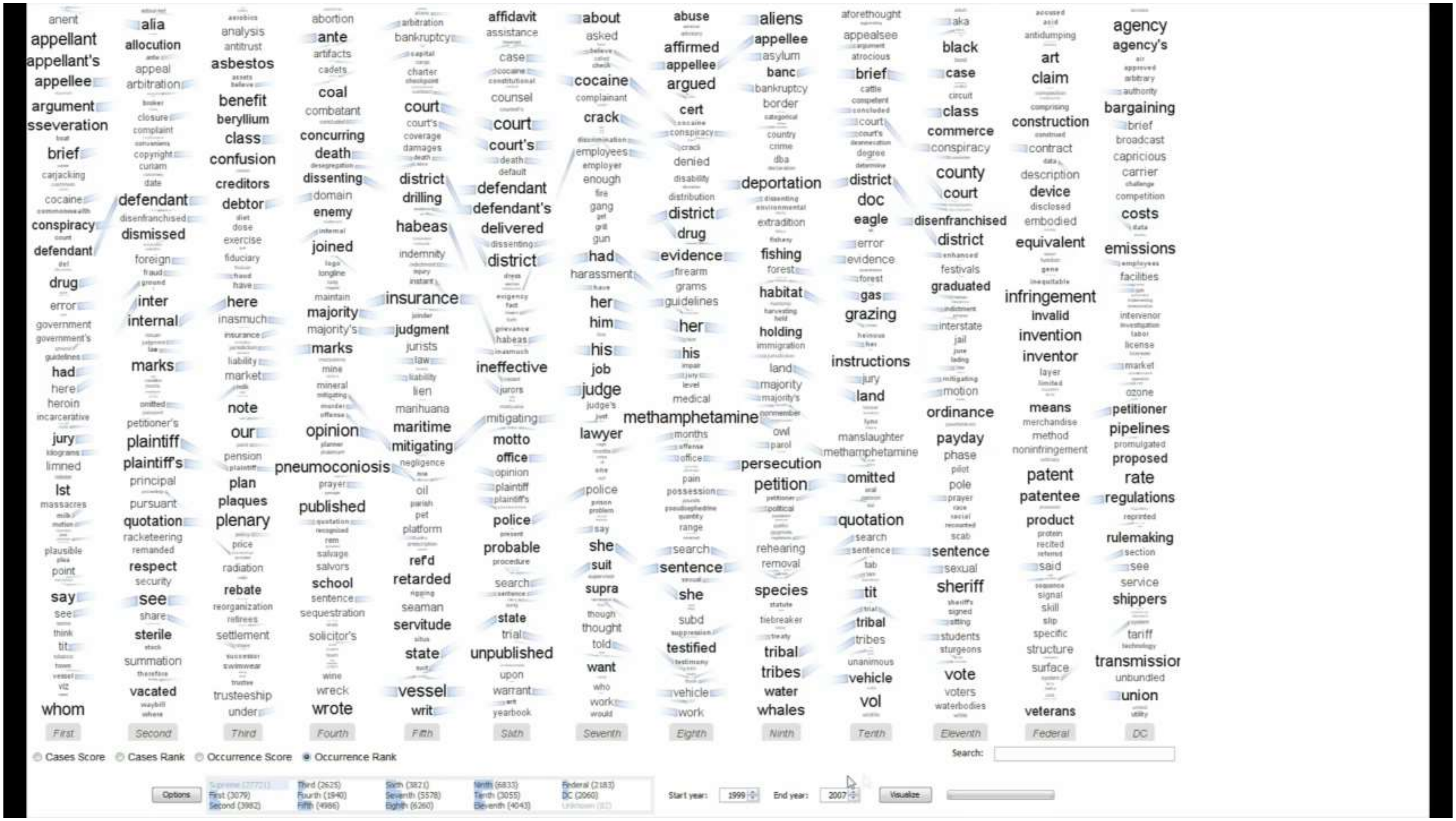
● Zoom ○ Pan

FT 27 AUG 92 / Hurricane batters southern **US** but lets insurers off lightly By **MARTIN DICKSON** and **ROBERT PESTON NEW YORK, LONDON**

HURRICANE **Andrew**, claimed to be the costliest natural disaster in **US** history, yesterday smashed its way through the state of **Louisiana**, inflicting severe damage on rural communities but narrowly missing the low-lying city of **New Orleans**. The storm, which brought havoc to southern **Florida** on **Monday** and then headed north-west across the **Gulf of Mexico**, had made landfall late on **Tuesday** night some 60 miles south-west of the city in the agricultural Cajun country. Although the damage from the hurricane's landfall in **Florida** on **Monday** was much greater than initially esti mated, insurers' losses there are likely to total less than Dollars 1bn, well below earlier expectations, a senior member of **Lloyd**'s insurance market said yesterday. In **Louisiana**, the hurricane landed with wind speeds of about 120 miles per hour and caused severe damage in small coastal centres such as **Morgan City**, **Franklin** and **New Iberia**. Associated tornadoes devastated **Laplace**, 20 miles west of **New Orleans**. Then, however, **Andrew** lost force as it moved north over land. By yesterday afternoon, it had been down-graded to tropical storm, in that its sustained windspeeds were below 75 mph. Initial reports said at least one person had died, 75 been injured and thousands made homeless along the **Louisiana** coast, after 14 confirmed deaths in **Florida** and three in the **Bahamas**. The storm caused little damage to **Louisiana**'s important oil-refining industry, although some plants had to halt production when electricity was cut. The **Lloyd's** member, in close contact with leading insurers in **Florida**, said that damage to insured property was remarkably small. *More than Dollars 15bn of damage may have been caused in all, but was mostly to uninsured property, he said. In **north Miami**, damage is minimal. Worst affected is one hotel, whose basement was flooded. Most of the destruction occurred in a 10-mile band across **Homestead**, 25 miles to the south of **Miami**, where a typical **house** sells for **Dollars 100,000** to Dollars 150,000. **US** insurers will face a bill in respect of such properties, but **Lloyd's** exposure there is minimal. Many destroyed power lines are thought to be uninsured, as are trees and shrubs uprooted across a wide area. Only one big hotel in that area has been badly damaged, a **Holiday Inn**. Across **Florida**, some 2m people remained without electric ity yesterday and health officials were warning the public to boil or chemically treat all water. Hurricane **Hugo**, which devastated*

Documents loaded.

Corpus Comparison

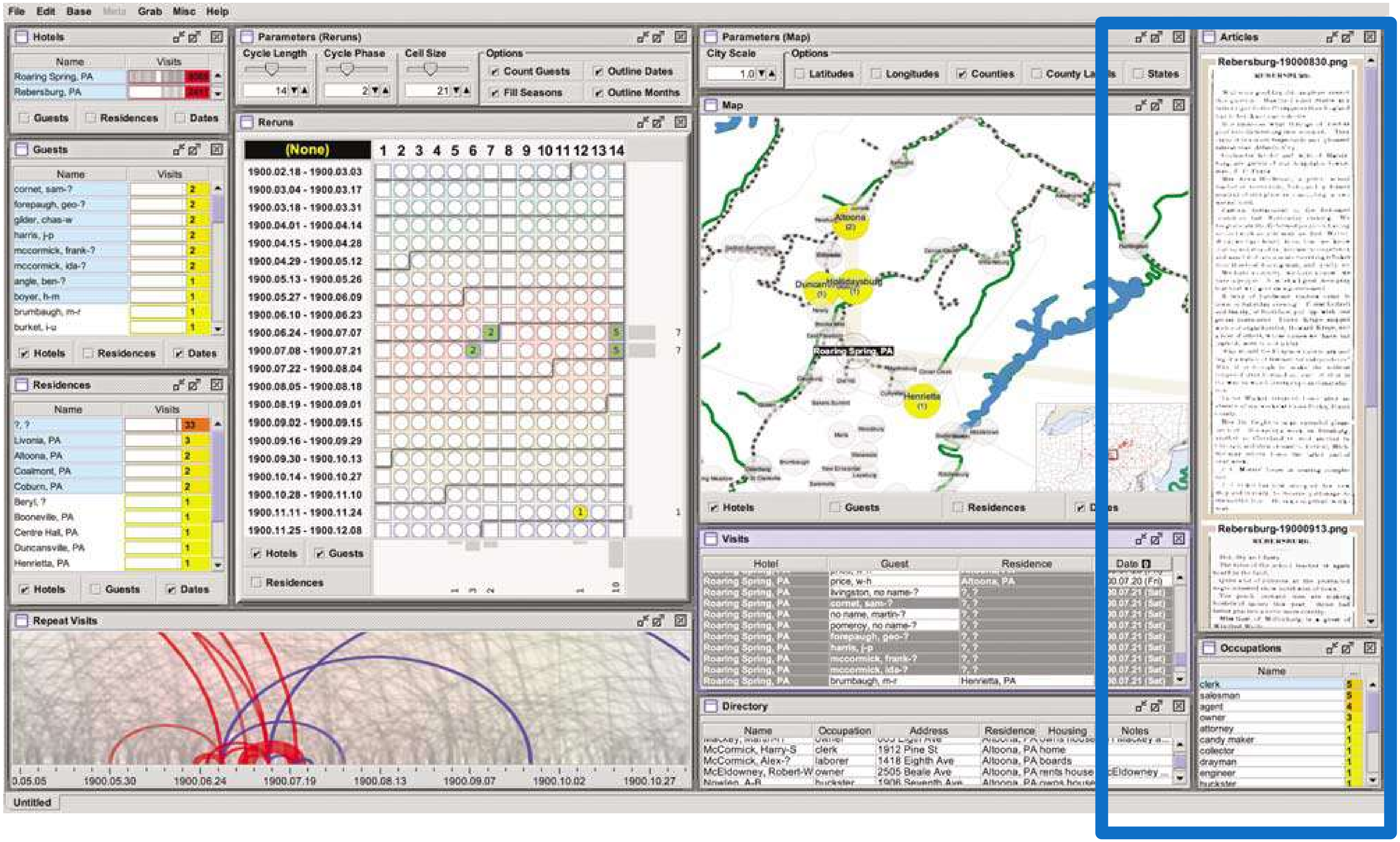


INTELLIGENCE ANALYSIS

Characteristics

- Multiple data streams
- Different data types: geolocation, phone calls, travel records, paper records, video surveillance, …
- Streaming data at different rates
- High cost of failure

Hotel Visits



Hotels

Name	Visits
Roaring Spring, PA	[illegible]
Rebersburg, PA	[illegible]

□ Guests □ Residences □ Dates

Guests

Name	Visits
cornet, sam-?	2
forepaugh, geo-?	2
gilder, chas-w	2
harris, j-p	2
mccormick, frank-?	2
mccormick, ida-?	2
angle, ben-?	1
boyer, h-m	1
brumbaugh, m-r	1
burkel, l-u	1

☑ Hotels □ Residences ☑ Dates

Residences

Name	Visits
?, ?	33
Livonia, PA	3
Altoona, PA	2
Coalmont, PA	2
Coburn, PA	2
Beryl, ?	1
Booneville, PA	1
Centre Hall, PA	1
Duncansville, PA	1
Henrietta, PA	1

☑ Hotels □ Guests ☑ Dates

Parameters (Reruns): Cycle Length 14; Cycle Phase 2; Cell Size 21; Options: ☑ Count Guests ☑ Outline Dates ☑ Fill Seasons ☑ Outline Months

Reruns: (None) 1 2 3 4 5 6 7 8 9 10 11 12 13 14

1900.02.18 - 1900.03.03
1900.03.04 - 1900.03.17
1900.03.18 - 1900.03.31
1900.04.01 - 1900.04.14
1900.04.15 - 1900.04.28
1900.04.29 - 1900.05.12
1900.05.13 - 1900.05.26
1900.05.27 - 1900.06.09
1900.06.10 - 1900.06.23
1900.06.24 - 1900.07.07
1900.07.08 - 1900.07.21
1900.07.22 - 1900.08.04
1900.08.05 - 1900.08.18
1900.08.19 - 1900.09.01
1900.09.02 - 1900.09.15
1900.09.16 - 1900.09.29
1900.09.30 - 1900.10.13
1900.10.14 - 1900.10.27
1900.10.28 - 1900.11.10
1900.11.11 - 1900.11.24
1900.11.25 - 1900.12.08

☑ Hotels ☑ Guests □ Residences

Parameters (Map): City Scale 1.0; Options: □ Latitudes □ Longitudes ☑ Counties □ County Labels □ States

Map: ☑ Hotels □ Guests □ Residences ☑ Dates

Visits

Hotel	Guest	Residence	Date
Roaring Spring, PA	price, w-h	Altoona, PA	00.07.20 (Fri)
Roaring Spring, PA	livingston, no name-?	?, ?	00.07.21 (Sat)
Roaring Spring, PA	cornet, sam-?	?, ?	00.07.21 (Sat)
Roaring Spring, PA	no name, martin-?	?, ?	00.07.21 (Sat)
Roaring Spring, PA	pomeroy, no name-?	?, ?	00.07.21 (Sat)
Roaring Spring, PA	forepaugh, geo-?	?, ?	00.07.21 (Sat)
Roaring Spring, PA	harris, j-p	?, ?	00.07.21 (Sat)
Roaring Spring, PA	mccormick, frank-?	?, ?	00.07.21 (Sat)
Roaring Spring, PA	mccormick, ida-?	?, ?	00.07.21 (Sat)
Roaring Spring, PA	brumbaugh, m-r	Henrietta, PA	00.07.21 (Sat)

Directory

Name	Occupation	Address	Residence	Housing	Notes
McCormick, Harry-S	clerk	1912 Pine St	Altoona, PA	home	
McCormick, Alex-?	laborer	1418 Eighth Ave	Altoona, PA	boards	
McEldowney, Robert-W	owner	2505 Beale Ave	Altoona, PA	rents house	cEldowney ...
Noslen, A-R	huckster	1906 Seventh Ave	Altoona, PA	owns house	

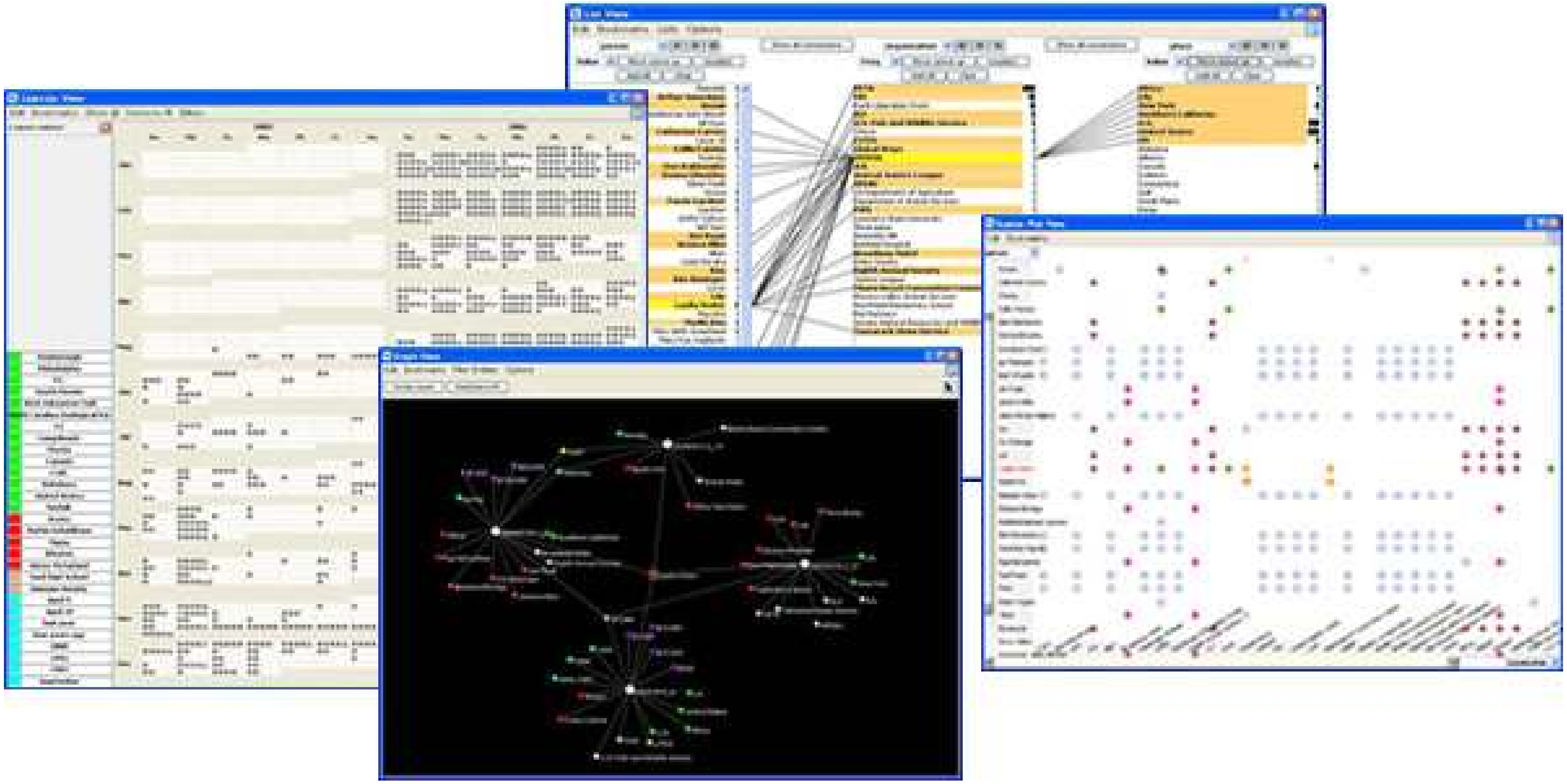
Articles: Rebersburg-19000830.png; Rebersburg-19000913.png

Occupations

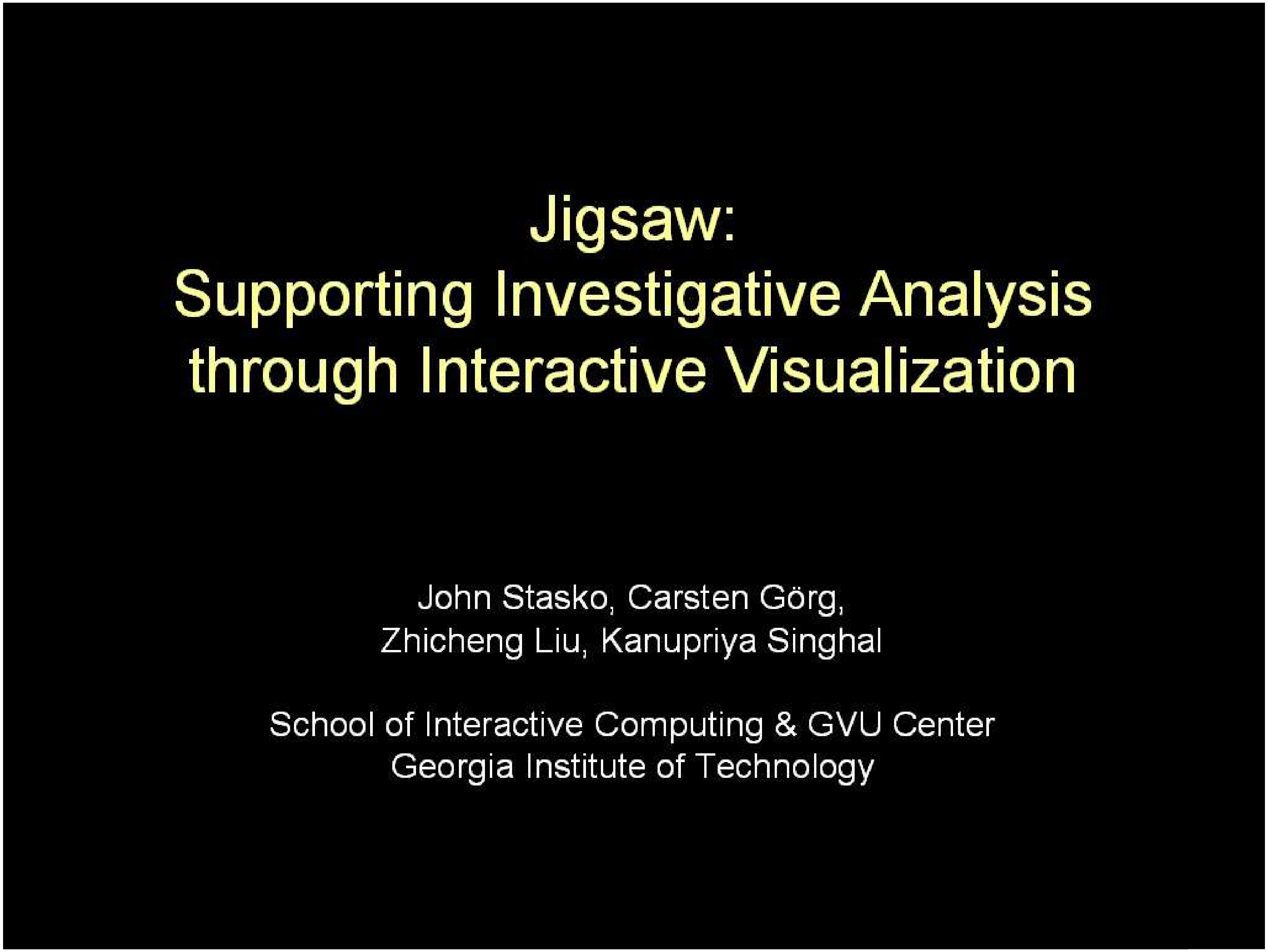
Name	
clerk	5
salesman	5
agent	4
owner	3
attorney	1
candy maker	1
collector	1
drayman	1
engineer	1
huckster	1

Weaver, C.; Fyfe, D.; Robinson, A.; Holdsworth, D.; Peuquet, D.; MacEachren, A.M., "Visual Analysis of Historic Hotel Visitation Patterns," IEEE VAST, 2006

Jigsaw



John Stasko and colleagues, numerous papers: http://www.cc.gatech.edu/gvu/ii/jigsaw/



Jigsaw: Supporting Investigative Analysis through Interactive Visualization

John Stasko, Carsten Görg,
Zhicheng Liu, Kanupriya Singhal

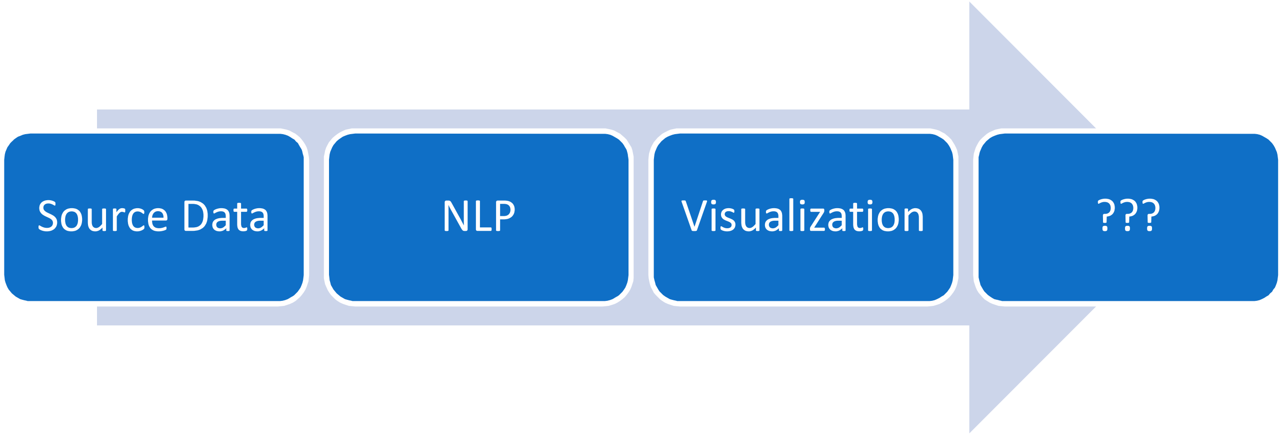
School of Interactive Computing & GVU Center
Georgia Institute of Technology

Text Visualization

OPEN RESEARCH AREAS

Trust

Unquantifiable uncertainty



Source Data

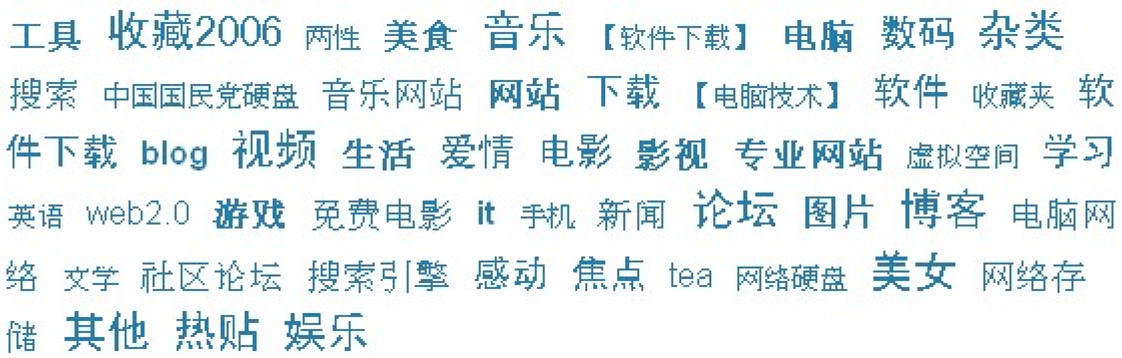
NLP

Visualization

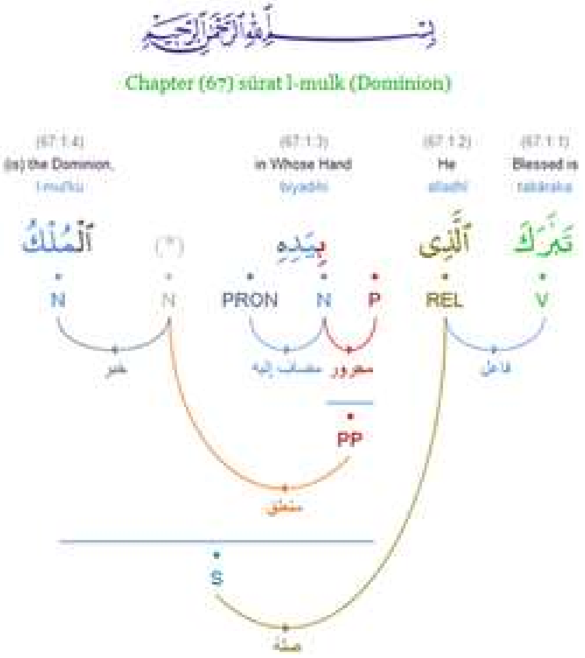
???

Multilanguage

- Do the same techniques work for non-Western languages?



工具 收藏2006 两性 美食 音乐 【软件下载】 电脑 数码 杂类 搜索 中国国民党硬盘 音乐网站 网站 下载 【电脑技术】 软件 收藏夹 软件下载 blog 视频 生活 爱情 电影 影视 专业网站 虚拟空间 学习 英语 web2.0 游戏 免费电影 it 手机 新闻 论坛 图片 博客 电脑网络 文学 社区论坛 搜索引擎 感动 焦点 tea 网络硬盘 美女 网络存储 其他 热贴 娱乐



Term / Concept Ambiguity

- Most meaning comes from our minds and common understanding.

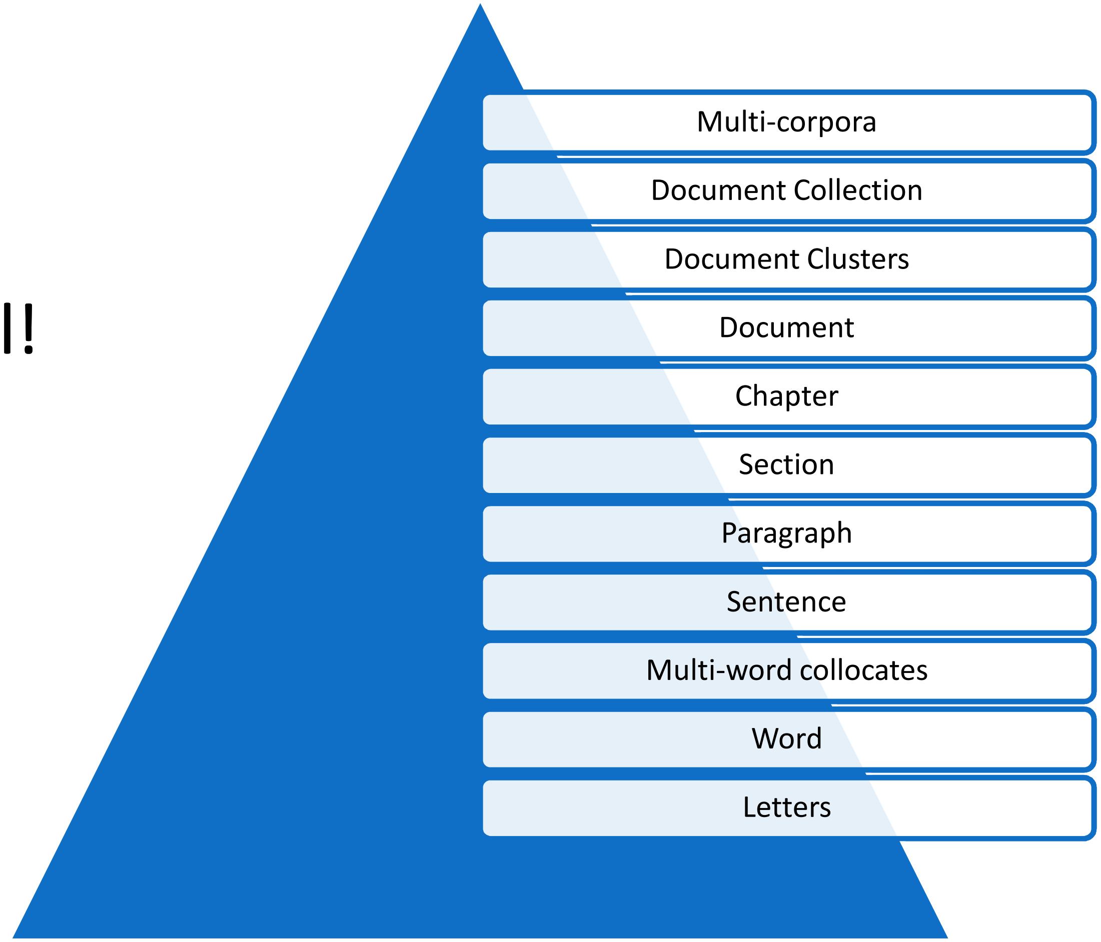
- "How much is that doggy in the window?"
  - how much: social system of barter and trade (not the size of the dog)
  - "doggy" implies childlike, plaintive, probably cannot do the purchasing on their own
  - "in the window" implies behind a store window, not really inside a window, requires notion of window shopping

(Hearst, 2006)

Finding 'Sweet Spots' in the Hierarchy

+ Meta data at each level!

Different levels for each genre?



Multi-corpora

Document Collection

Document Clusters

Document

Chapter

Section

Paragraph

Sentence

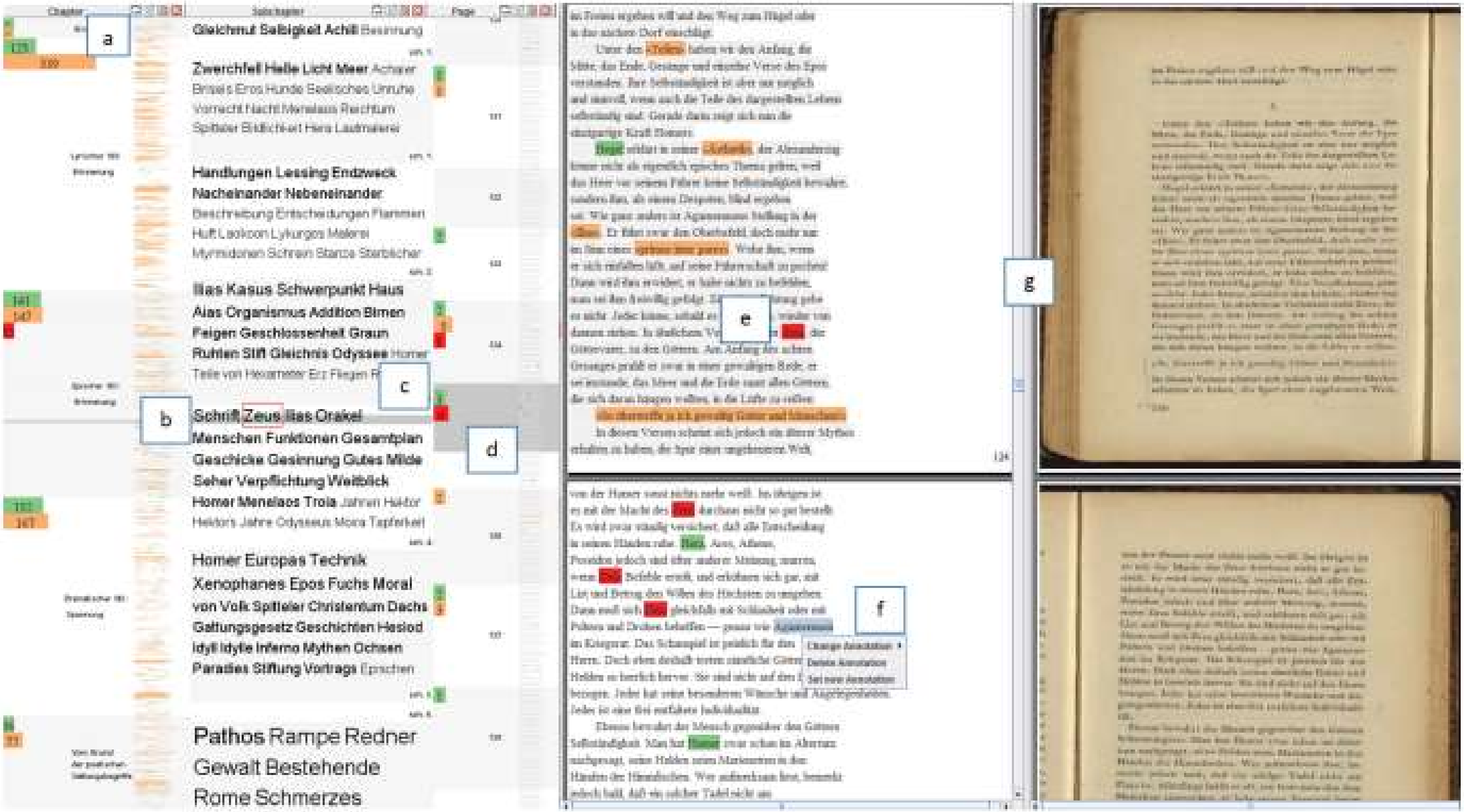
Multi-word collocates

Word

Letters

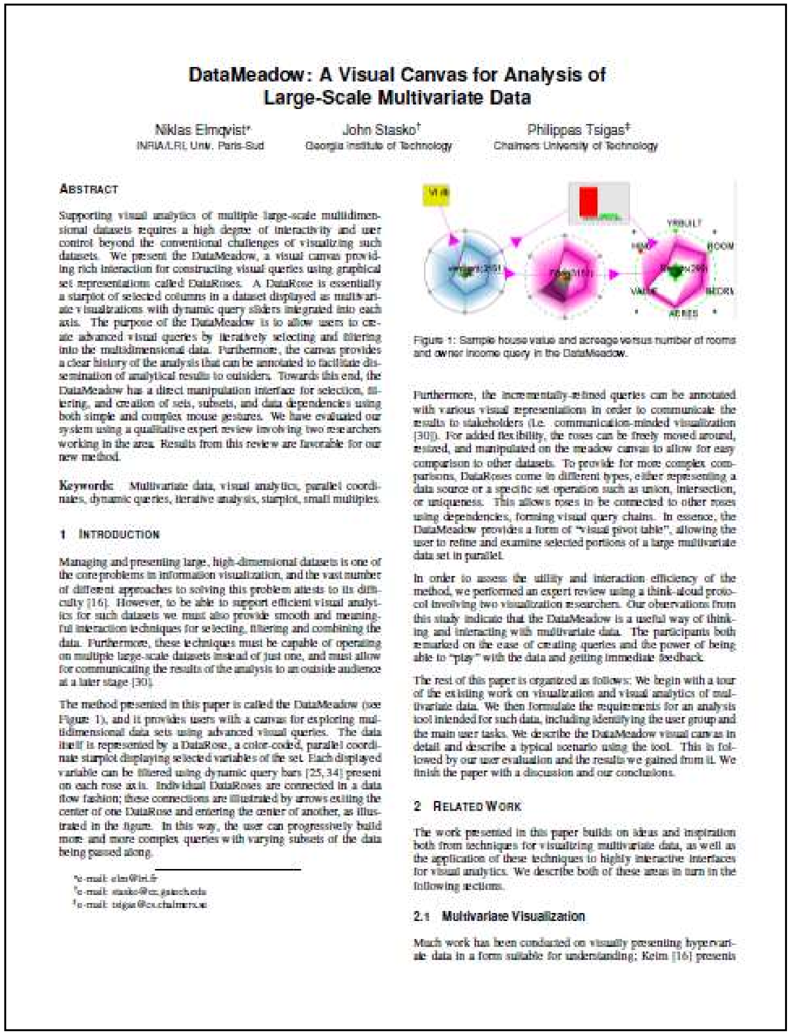
?

VariFocalReader


Chapter
Subchapter
Page
a
b
c
d
e
f
g
Gleichmut Selbigkeit Achill Besinnung
Zwerchfell Helle Licht Meer Achaier
Briseis Eros Hunde Seelisches Unruhe
Vorrecht Nacht Menelaos Reichtum
Spitteler Bildlichkeit Hera Lautmalerei
Handlungen Lessing Endzweck
Nacheinander Nebeneinander
Beschreibung Entscheidungen Flammen
Huft Laokoon Lykurgos Malerei
Myrmidonen Schranken Stanze Sterblicher
Ilias Kasus Schwerpunkt Haus
Aias Organismus Addition Birnen
Feigen Geschlossenheit Graun
Ruhten Stift Gleichnis Odyssee Homer
Teile von Hexameter Erz Fliegen
Schrift Zeus Ilias Orakel
Menschen Funktionen Gesamtplan
Geschicke Gesinnung Gutes Milde
Seher Verpflichtung Weitblick
Homer Menelaos Troia Jahren Hektor
Hektors Jahre Odysseus Moira Tapferkeit
Homer Europas Technik
Xenophanes Epos Fuchs Moral
von Volk Spitteler Christentum Dachs
Gattungsgesetz Geschichten Hesiod
Idyll Idylle Inferno Mythen Ochsen
Paradies Stiftung Vortrags Epischen
Pathos Rampe Redner
Gewalt Bestehende
Rome Schmerzes
Change Annotation
Delete Annotation
Set new Annotation

S. Koch, M. John, M. Wörner, A. Müller and T. Ertl, "VarifocalReader — In-Depth Visual Analysis of Large Text Documents," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1723-1732, Dec. 31 2014.

Multi-Media Documents



DataMeadow: A Visual Canvas for Analysis of Large-Scale Multivariate Data

Niklas Elmqvist*
INRIA/LRI, Univ. Paris-Sud

John Stasko†
Georgia Institute of Technology

Philippas Tsigas‡
Chalmers University of Technology

Abstract

Supporting visual analytics of multiple large-scale multidimensional datasets requires a high degree of interactivity and user control beyond the conventional challenges of visualizing such datasets. We present the DataMeadow, a visual canvas providing rich interaction for constructing visual queries using graphical set representations called DataRoses. A DataRose is essentially a starplot of selected columns in a dataset displayed as multivariate visualizations with dynamic query sliders integrated into each axis. The purpose of the DataMeadow is to allow users to create advanced visual queries by iteratively selecting and filtering into the multidimensional data. Furthermore, the canvas provides a clear history of the analysis that can be annotated to facilitate dissemination of analytical results to outsiders. Towards this end, the DataMeadow has a direct manipulation interface for selection, filtering, and creation of sets, subsets, and data dependencies using both simple and complex mouse gestures. We have evaluated our system using a qualitative expert review involving two researchers working in the area. Results from this review are favorable for our new method.

Keywords: Multivariate data, visual analytics, parallel coordinates, dynamic queries, iterative analysis, starplot, small multiples.

1 Introduction

Managing and presenting large, high-dimensional datasets is one of the core problems in information visualization, and the vast number of different approaches to solving this problem attests to its difficulty [16]. However, to be able to support efficient visual analytics for such datasets we must also provide smooth and meaningful interaction techniques for selecting, filtering and combining the data. Furthermore, these techniques must be capable of operating on multiple large-scale datasets instead of just one, and must allow for communicating the results of the analysis to an outside audience at a later stage [30].

The method presented in this paper is called the DataMeadow (see Figure 1), and it provides users with a canvas for exploring multidimensional data sets using advanced visual queries. The data itself is represented by a DataRose, a color-coded, parallel coordinate starplot displaying selected variables of the set. Each displayed variable can be filtered using dynamic query bars [25, 34] present on each rose axis. Individual DataRoses are connected in a data flow fashion; these connections are illustrated by arrows exiting the center of one DataRose and entering the center of another, as illustrated in the figure. In this way, the user can progressively build more and more complex queries with varying subsets of the data being passed along.

Figure 1: Sample house value and acreage versus number of rooms and owner income query in the DataMeadow.

Furthermore, the incrementally-refined queries can be annotated with various visual representations in order to communicate the results to stakeholders (i.e. communication-minded visualization [30]). For added flexibility, the roses can be freely moved around, resized, and manipulated on the meadow canvas to allow for easy comparison to other datasets. To provide for more complex comparisons, DataRoses come in different types, either representing a data source or a specific set operation such as union, intersection, or uniqueness. This allows roses to be connected to other roses using dependencies, forming visual query chains. In essence, the DataMeadow provides a form of "visual pivot table", allowing the user to refine and examine selected portions of a large multivariate data set in parallel.

In order to assess the utility and interaction efficiency of the method, we performed an expert review using a think-aloud protocol involving two visualization researchers. Our observations from this study indicate that the DataMeadow is a useful way of thinking and interacting with multivariate data. The participants both remarked on the ease of creating queries and the power of being able to "play" with the data and getting immediate feedback.

The rest of this paper is organized as follows: We begin with a tour of the existing work on visualization and visual analytics of multivariate data. We then formulate the requirements for an analysis tool intended for such data, including identifying the user group and the main user tasks. We describe the DataMeadow visual canvas in detail and describe a typical scenario using the tool. This is followed by our user evaluation and the results we gained from it. We finish the paper with a discussion and our conclusions.

2 Related Work

The work presented in this paper builds on ideas and inspiration both from techniques for visualizing multivariate data, as well as the application of these techniques to highly interactive interfaces for visual analytics. We describe both of these areas in turn in the following sections.

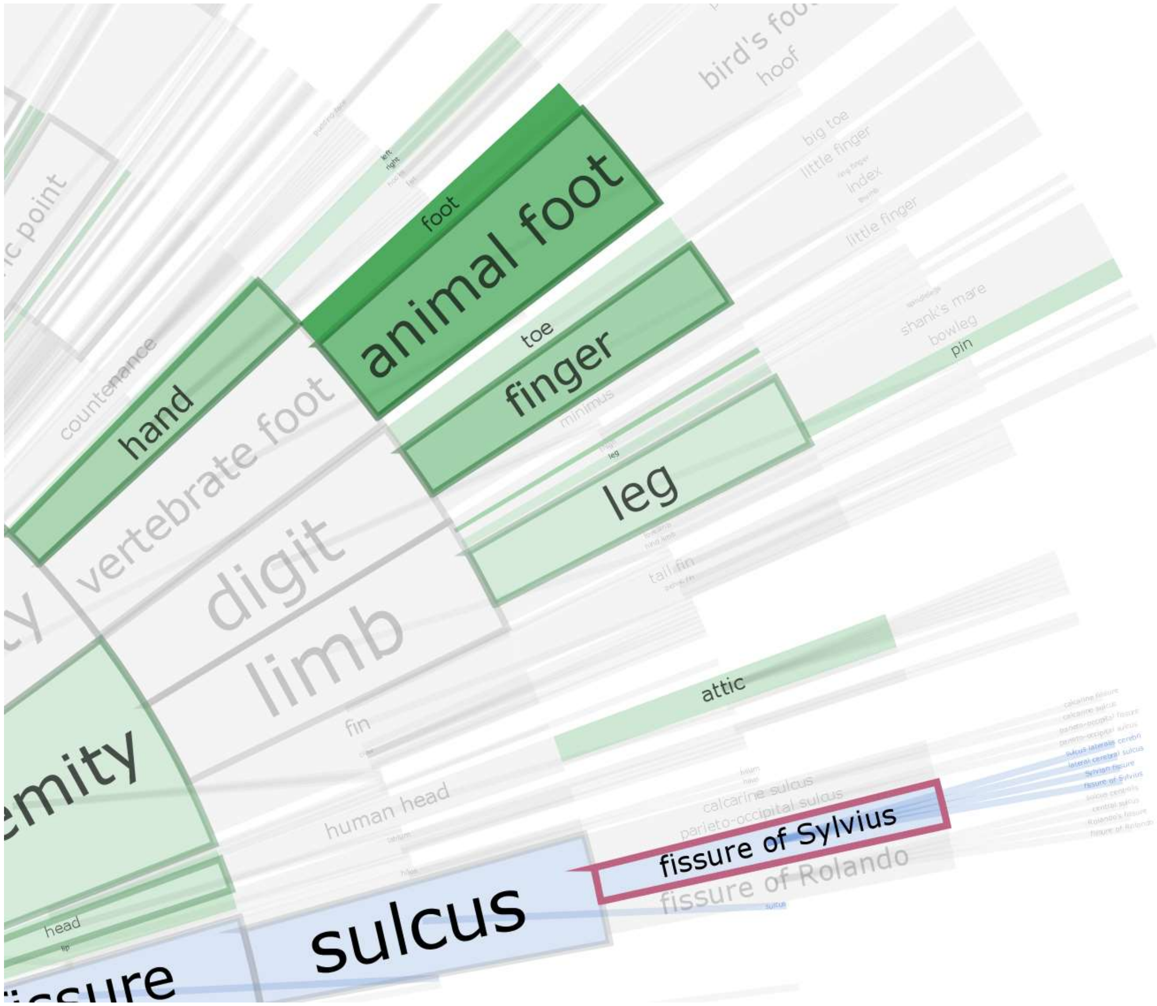
2.1 Multivariate Visualization

Much work has been conducted on visually presenting hypervariate data in a form suitable for understanding; Keim [16] presents

*e-mail: elm@lri.fr
†e-mail: stasko@cc.gatech.edu
‡e-mail: tsigas@cs.chalmers.se

Audio + Text Analysis (e.g. court proceedings)

Appropriate Abstraction

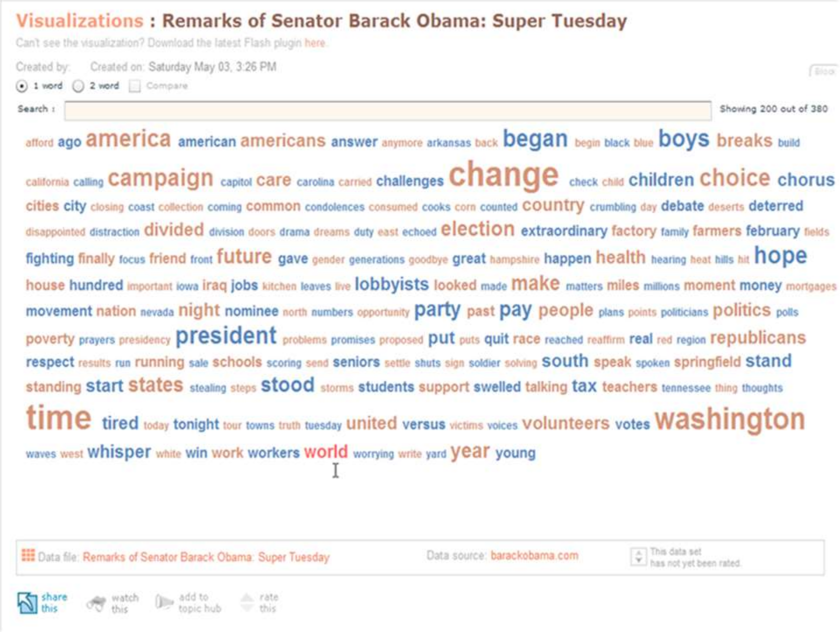


Proper Nouns

- Do not appear in ontologies like WordNet
- Challenging to translate

- Holy Grail: NNPs + sentiment + visualization

Interactivity: Linking to Text



Visualizations : Remarks of Senator Barack Obama: Super Tuesday

Can't see the visualization? Download the latest Flash plugin here.

Created by: Created on: Saturday May 03, 3:26 PM

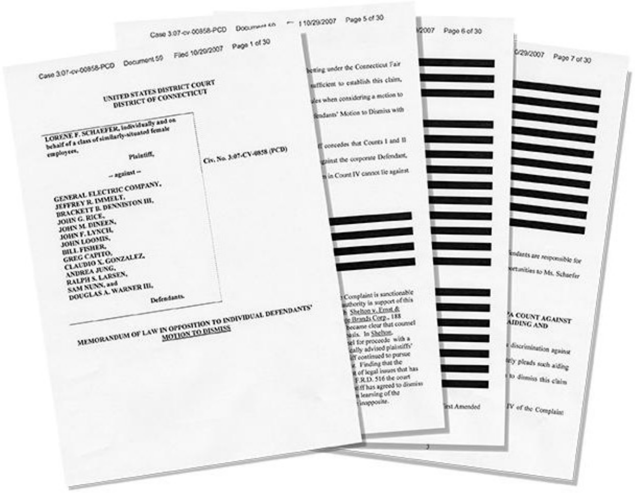
● 1 word ○ 2 word □ Compare

Search : Showing 200 out of 380

afford ago america american americans answer anymore arkansas back began begin black blue boys breaks build california calling campaign capitol care carolina carried challenges change check child children choice chorus cities city closing coast collection coming common condolences consumed cooks corn counted country crumbling day debate deserts deterred disappointed distraction divided division doors drama dreams duty east echoed election extraordinary factory family farmers february fields fighting finally focus friend front future gave gender generations goodbye great hampshire happen health hearing heat hills hit hope house hundred important iowa iraq jobs kitchen leaves live lobbyists looked made make matters miles millions moment money mortgages movement nation nevada night nominee north numbers opportunity party past pay people plans points politicians politics polls poverty prayers presidency president problems promises proposed put puts quit race reached reaffirm real red region republicans respect results run running sale schools scoring send seniors settle shuts sign soldier solving south speak spoken springfield stand standing start states stealing steps stood storms students support swelled talking tax teachers tennessee thing thoughts time tired today tonight tour towns truth tuesday united versus victims voices volunteers votes washington waves west whisper white win work workers world worrying write yard year young

Data file: Remarks of Senator Barack Obama: Super Tuesday Data source: barackobama.com This data set has not yet been rated.

share this
watch this
add to topic hub
rate this



Case 3:07-cv-00858-PCD Document 50 Filed 10/20/2007 Page 1 of 30

UNITED STATES DISTRICT COURT
DISTRICT OF CONNECTICUT

LORENE F. SCHAEFER, individually and on behalf of a class of similarly-situated female employees,

Plaintiff,

-- against --

GENERAL ELECTRIC COMPANY,
JEFFREY R. IMMELT,
BRACKETT B. DENNISTON III,
JOHN G. RICE,
JOHN M. DINEEN,
JOHN F. LYNCH,
JOHN LOOMIS,
BILL FISHER,
GREG CAPITO,
CLAUDIO X. GONZALEZ,
ANDREA JUNG,
RALPH S. LARSEN,
SAM NUNN, and
DOUGLAS A. WARNER III,

Defendants.

Civ. No. 3:07-CV-0858 (PCD)

MEMORANDUM OF LAW IN OPPOSITION TO INDIVIDUAL DEFENDANTS' MOTION TO DISMISS

Case 3:07-cv-00858-PCD Document 50 1 10/29/2007 Page 5 of 30

/2007 Page 6 of 30

/29/2007 Page 7 of 30

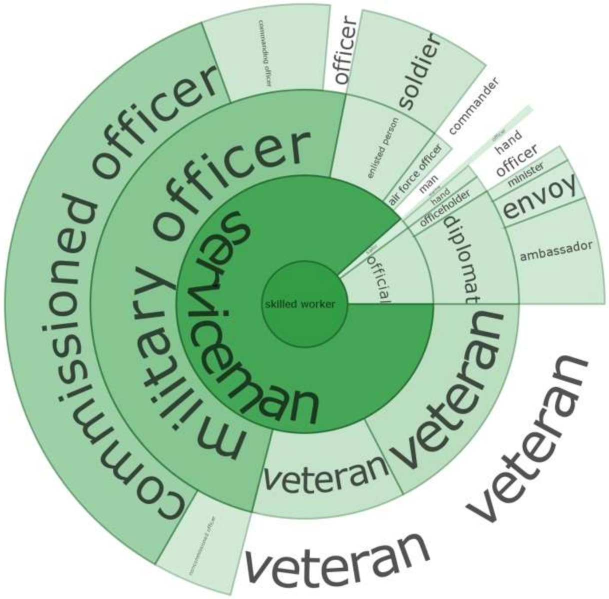
[redacted]

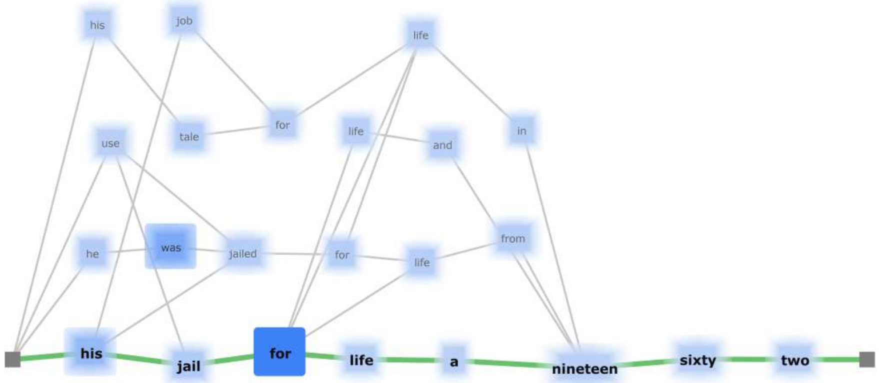
[redacted]

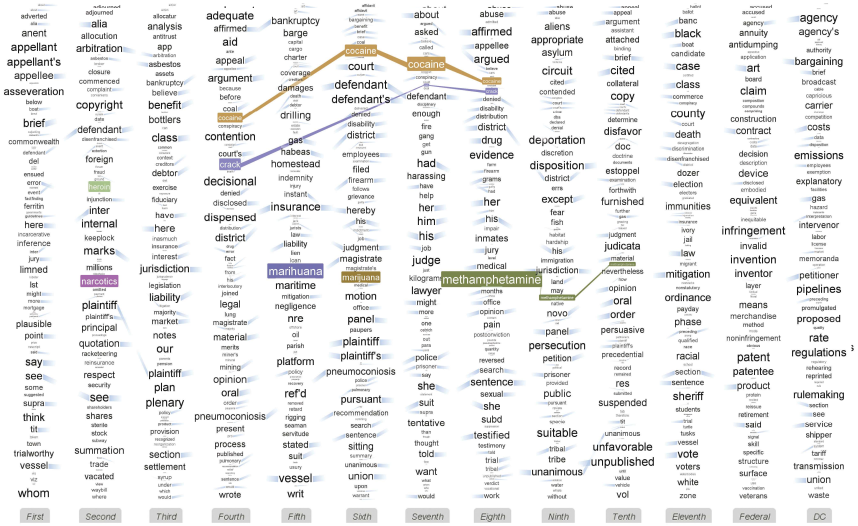
[redacted]

Legibility

- Screen Real Estate

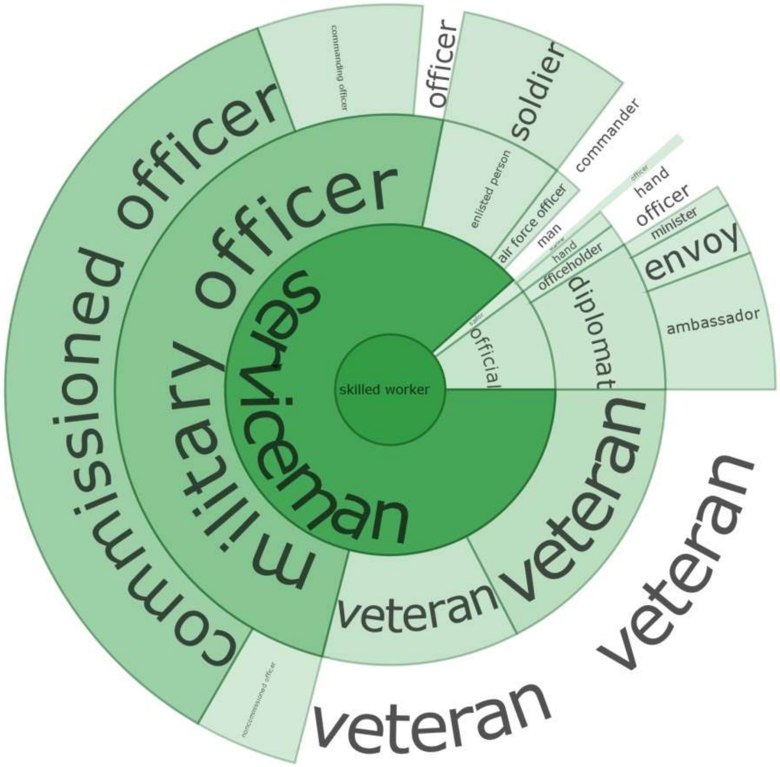


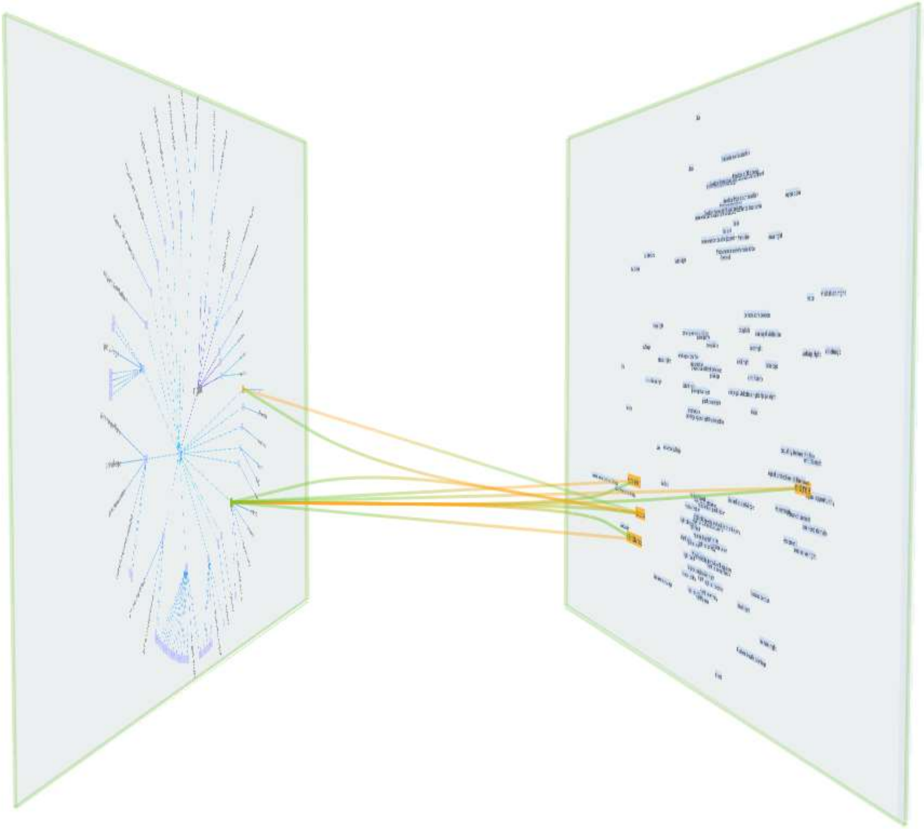




Legibility

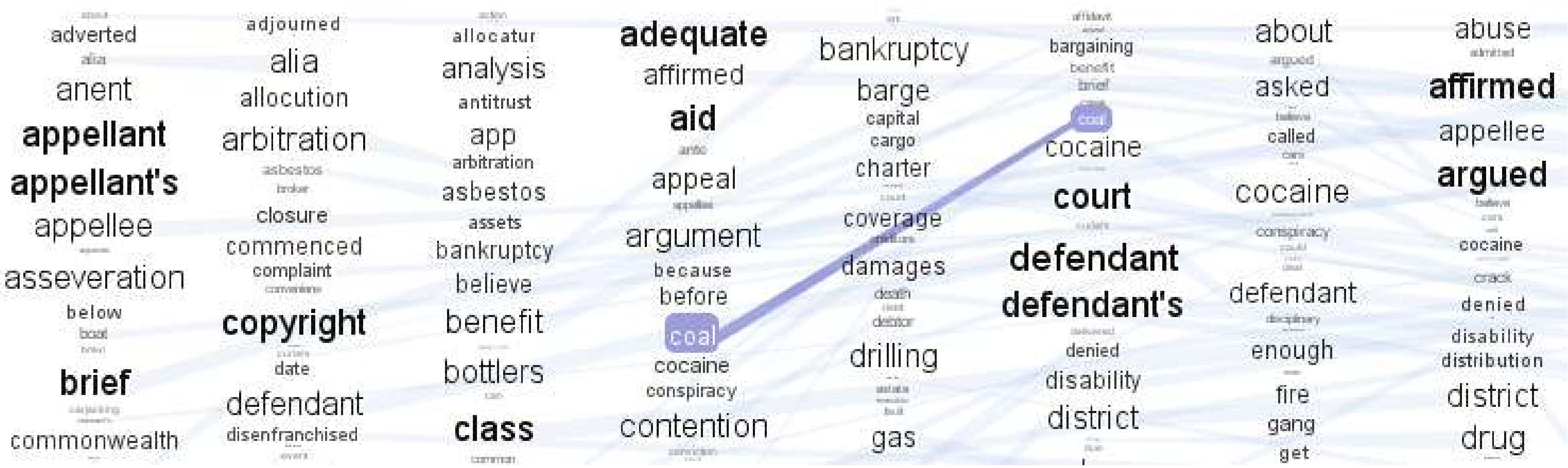
- Orientation

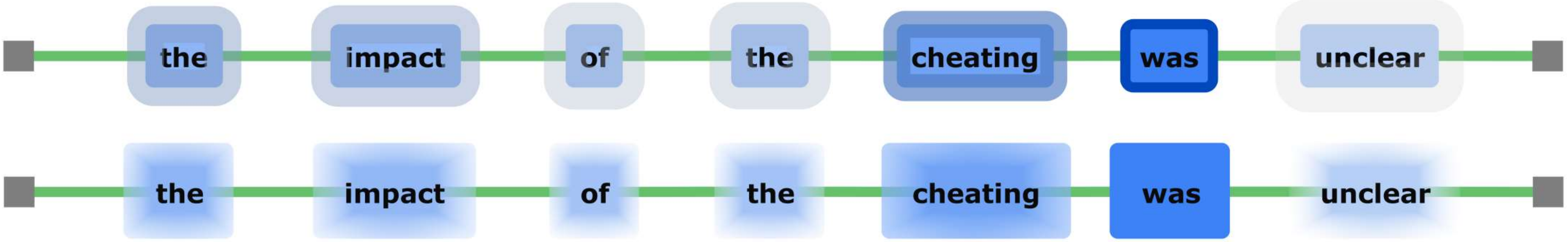




Legibility

- Overlay or Background Interference



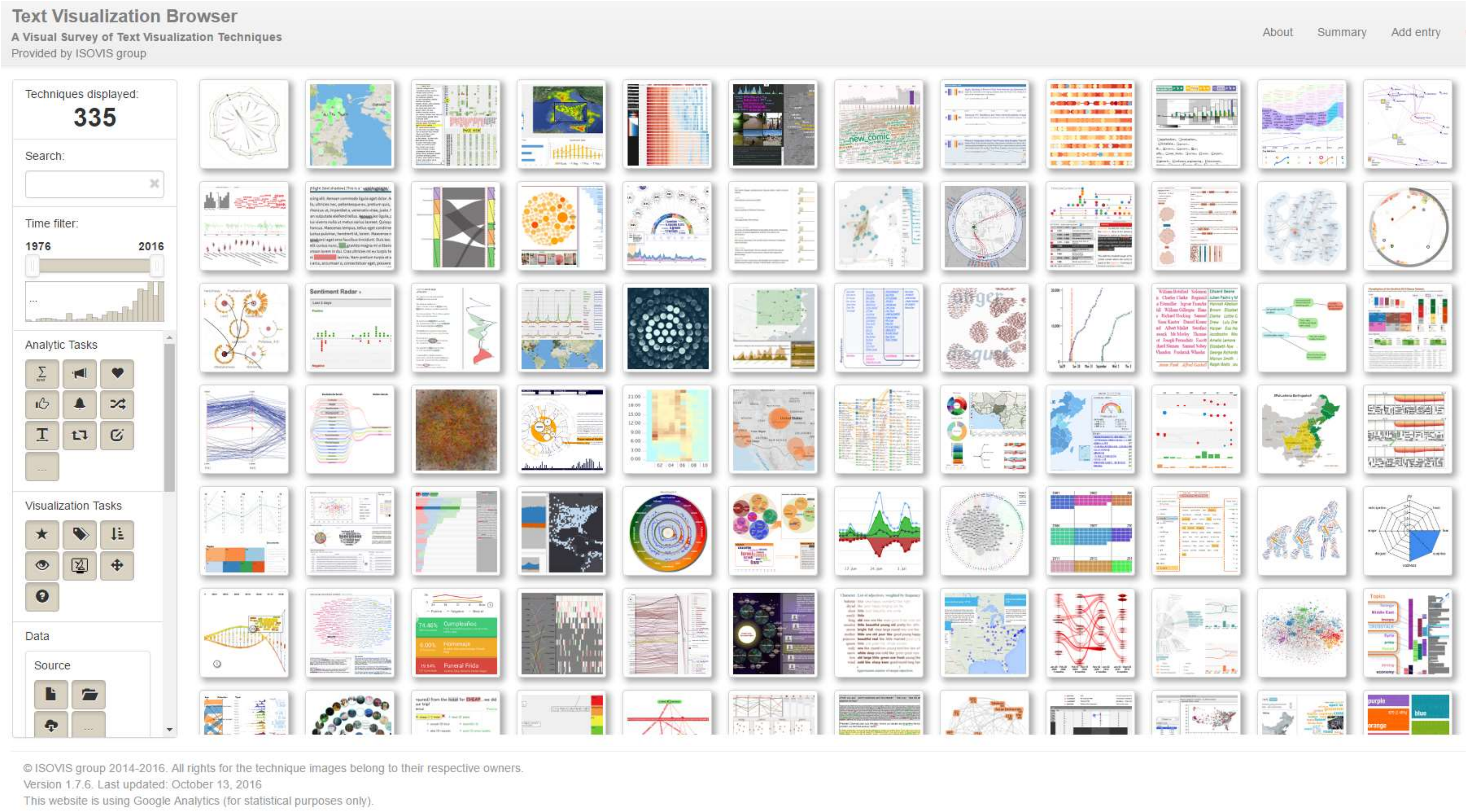


Domains of Application

- Medicine: electronic medical records
- Business: social media analytics, corporate document collection management
- Crime Prevention and Intelligence Analysis: find threats in communications and blogs
- Legal: sift through evidence, e.g. millions of emails, to investigate fraud
- Literary and History Scholarship

CONCLUSION

TextVis Survey



Text Visualization Browser

A Visual Survey of Text Visualization Techniques

Provided by ISOVIS group

Techniques displayed:
335

Search:

Time filter:
1976 2016

© ISOVIS group 2014-2016. All rights for the technique images belong to their respective owners.
Version 1.7.6. Last updated: October 13, 2016
This website is using Google Analytics (for statistical purposes only).

http://textvis.lnu.se/

Summary

- Text visualization is an exciting area of ongoing research – check out recent workshop papers at textvis.org and vis4dh.org
- Text visualization bridges visualization design, interaction design, and natural language processing