

TEXT VISUALIZATION

KIRELL BENZI, PH.D.

Why visualize text?

Faster understanding: get quick insights on what I am reading

Comparison: compare document collections, or inspect evolution of collection over time

Clustering (grouping): classification or topic extractions

Correlation: compare patterns in the current text to other datasets

Characteristics of a textual representation

Abstract representation of the data

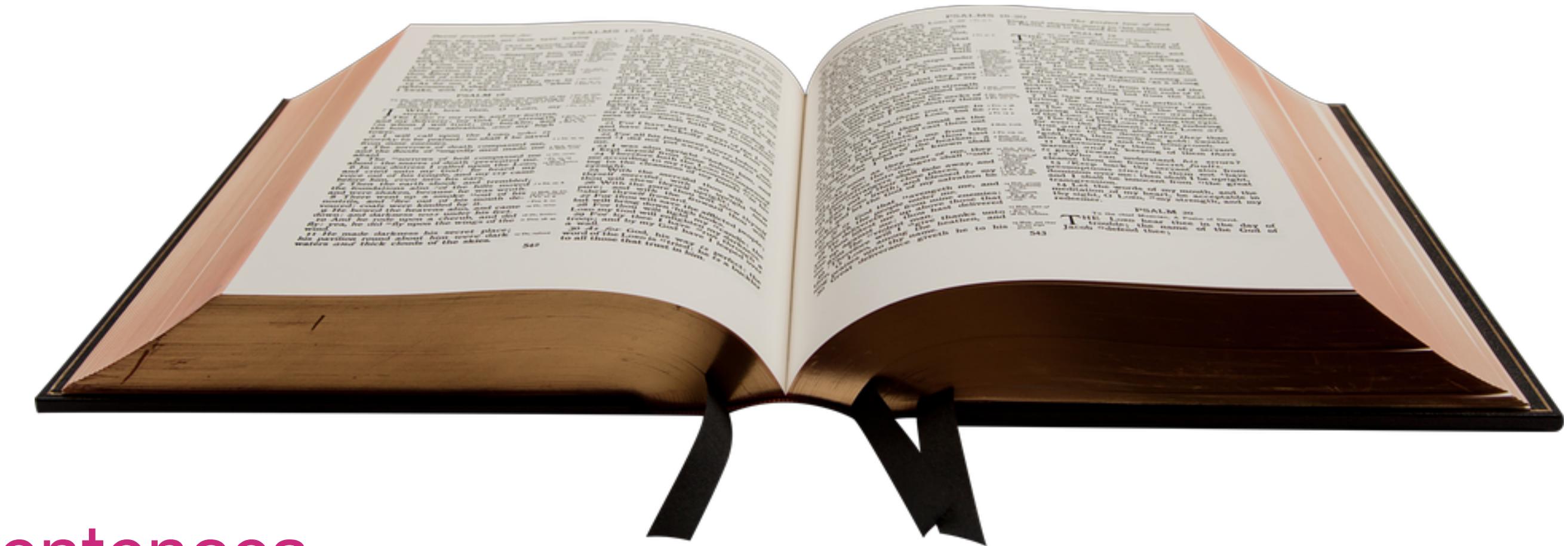
Powerful representation to describe

Linear perception

Semi-structured

grammar, punctuation, words, sentences

paragraphs, typography, calligraphy



Different across population groups!

Raw text visualization

Benedict Arnold's expedition to Quebec

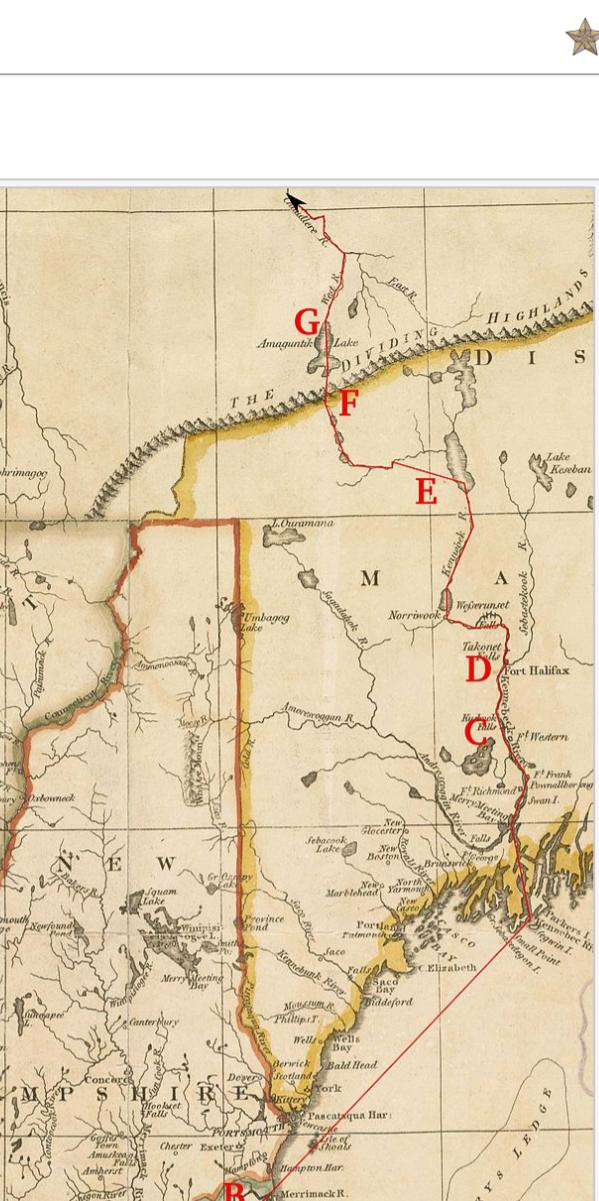
From Wikipedia, the free encyclopedia

In September 1775, early in the [American Revolutionary War](#), Colonel Benedict Arnold led a force of 1,100 [Continental Army](#) troops on an expedition from [Cambridge](#), [Massachusetts](#) to the gates of [Quebec City](#). The expedition was part of a two-pronged [invasion](#) of the British [Province of Quebec](#), and passed through the wilderness of what is now [Maine](#). The other expedition invaded Quebec from [Lake Champlain](#), led by [Richard Montgomery](#).

Unanticipated problems beset the expedition as soon as it left the last significant colonial outposts in Maine. The [portages](#) up the [Kennebec River](#) proved grueling, and the boats frequently leaked, ruining [gunpowder](#) and spoiling food supplies. More than a third of the men turned back before reaching the height of land between the Kennebec and [Chaudière](#) rivers. The areas on either side of the height of land were swampy tangles of lakes and streams, and the traversal was made more difficult by bad weather and inaccurate maps. Many of the troops lacked experience handling boats in [white water](#), which led to the destruction of more boats and supplies in the descent to the [Saint Lawrence River](#) via the fast-flowing Chaudière.

By the time that Arnold reached the French settlements above the Saint Lawrence River in November, his force was reduced to 600 starving men. They had traveled about 350 miles (560 km) through poorly charted wilderness, twice the distance that they had expected to cover. Arnold's troops crossed the Saint Lawrence on November 13 and 14, assisted by the local French-speaking [Canadiens](#), and attempted to put Quebec City under siege. Failing in this, they withdrew to [Point-aux-Trembles](#) until Montgomery arrived to lead an [unsuccessful attack](#) on the city. Arnold was rewarded for his effort in leading the expedition with a promotion to [brigadier general](#).

Arnold's route through northern Maine has been listed on the [National Register of Historic Places](#) as the **Arnold Trail to Quebec**, and some geographic features in the area bear names of expedition participants.



Rich text

Overview + details with minimap

Text datasets

Documents

Web pages, articles, logs, books, emails, comments, source code, tags

Collections of documents

Messages (chats, emails, letters), social profiles, publications



```
trace.log - Notepad
File Edit Format View Help

10/5/2010 3:56:54 PM, Category: General, Priority: 5, Severity: Information, "MirrorJob.Execute Complete", Machine: MOSSDEV,
10/5/2010 3:57:01 PM, Category: General, Priority: 5, Severity: Information, "MirrorJob.Execute Starting...", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetShadowRoot Starting...", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetSettings Starting...", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetSettings Complete", Machine: MOSSDEV,
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetShadowRoot Complete", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetLastStartTime Starting...", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetSettings Starting...", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetSettings Complete", Machine: MOSSDEV,
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetLastStartTime Complete", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.Recordstart Starting...", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetSettings Starting...", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetSettings Complete", Machine: MOSSDEV,
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.Recordstart Complete", Machine: MOSSDEV,
10/5/2010 3:57:01 PM, Category: General, Priority: 5, Severity: Information, "Processor.ProcessChanges Starting...", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "AuditLog.GetChanges Starting... - startDate = 2010
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "AuditLog.GetChanges - Got 0 Entries", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 1, Severity: verbose, "AuditLog.GetChanges Complete", Machine: MOSSDEV, R
10/5/2010 3:57:01 PM, Category: General, Priority: 5, Severity: Information, "Processor.ProcessChanges Queried Audit Log - G
10/5/2010 3:57:01 PM, Category: General, Priority: 5, Severity: Information, "Processor.ProcessChanges Complete", Machine: MOSSDEV
10/5/2010 3:57:01 PM, Category: General, Priority: 5, Severity: Information, "MirrorJob.Execute Complete", Machine: MOSSDEV,
10/5/2010 3:58:07 PM, Category: General, Priority: 5, Severity: Information, "MirrorJob.Execute Starting...", Machine: MOSSDEV
10/5/2010 3:58:07 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetShadowRoot Starting...", Machine: MOSSDEV
10/5/2010 3:58:07 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetSettings Starting...", Machine: MOSSDEV
10/5/2010 3:58:07 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetSettings Complete", Machine: MOSSDEV,
10/5/2010 3:58:07 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetShadowRoot Complete", Machine: MOSSDEV
10/5/2010 3:58:07 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetLastStartTime Starting...", Machine: MOSSDEV
10/5/2010 3:58:07 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetSettings Starting...", Machine: MOSSDEV
10/5/2010 3:58:07 PM, Category: General, Priority: 1, Severity: verbose, "MirrorJob.GetSettings Complete", Machine: MOSSDEV,
```



Challenges of text viz



Summarizing text?

High dimensionality, if possible summarize text with text. Choosing the right keywords?

Providing relevant context to assist the reader?

Need to understand the semantics

Show (or provide access to) the source text.

Ontologies extraction

Needs to understand the domain in addition of the previous points

In practice

Determine your analysis task

Find the best tools and models to match the task:

Read the litterature, look at other data viz, softwares, etc

Text as data

Words are ambiguous

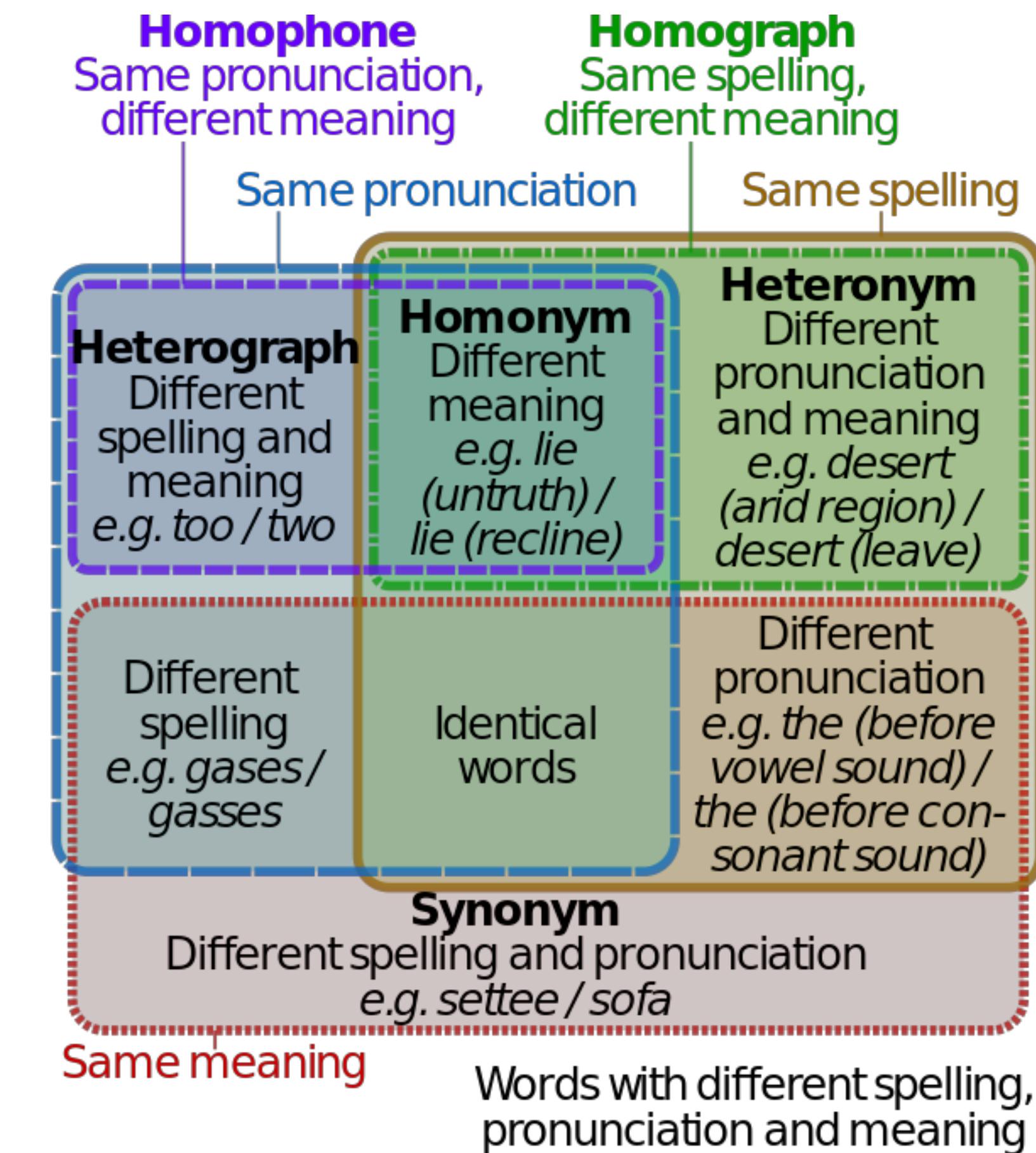
Cannot be used as nominal data: (20,000+) words

Cannot simply do an equality test

Some words are correlated: **Paris, New York, Madrid**

Ordered: **January, February, March**

Conjugated, tenses, antonyms, synonyms, etc.



Text processing pipeline

1. Tokenization

Segment text into terms: “a sentence” -> [“a”,
“sentence”]

Remove stop words

Remove numbers and symbols: #love, @potus, yo!!!!?

Define what to do with entities: Google, U.S.A

Default English stopwords list

This list is used in our [Page Analyzer](#) and [Article Analyzer](#) for English text, when you let it use the default stopwords list.

a	ourselves
about	out
above	over
after	own
again	same
against	shan't
all	she
am	she'd
an	she'll
and	she's
any	should
are	shouldn't
aren't	so
as	some
at	such
be	than
because	that
been	that's
before	the
being	their
below	theirs
between	them
both	themselves
but	then
by	there

Named Entity Recognition

Goal: label named entities in a text

James Bond -> PERSON

Germany -> COUNTRY

Rue du Petit Chêne -> ADDRESS

(555) 142-3432 -> PHONE NUMBER

How do these entities relate?

Example: Check co-occurrences in short term window (5 words)

Named entity Recognition

Google, headquartered in Mountain View, unveiled the new Android phone at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones. ANALYZE

[See supported languages](#)

Entities Sentiment Syntax

⟨[Google](#)⟩₁, headquartered in ⟨[Mountain View](#)⟩₆, unveiled the new ⟨[Android](#)⟩₄ ⟨[phone](#)⟩₃ at the ⟨[Consumer Electronic Show](#)⟩₇. ⟨[Sundar Pichai](#)⟩₅ said in his ⟨[keynote](#)⟩₉ that ⟨[users](#)⟩₂ love their new ⟨[Android](#)⟩₄ ⟨[phones](#)⟩₈.

1. Google	ORGANIZATION
Sentiment: Score 0 Magnitude 0	
Wikipedia Article	
Salience: 0.26	
2. users	PERSON
Sentiment: Score 0.7 Magnitude 1.5	
Salience: 0.15	
3. phone	CONSUMER GOOD
Sentiment: Score 0 Magnitude 0	
Salience: 0.13	
4. Android	CONSUMER GOOD
Sentiment: Score 0.4 Magnitude 0.9	
Wikipedia Article	

Token normalization

2. Stemming

Set of rules to remove characters from words (fast and simple)

"fishing", "fished", and "fisher" to the root word, "fish"

3. Lemmatization

Complex process to find the dictionary form of a word (needs a dictionary)

better -> good, went -> go

4. Ordered word mapping

From tokens to IDs: go -> 1, good -> 2

Tips

Use a library! [Gensim](#), [Spacy](#), [Tensorflow](#)

Beware of text input! [Gr8t m8t, u rok, c u l8tr](#)

Don't stem entities

[NATO \(North Atlantic Treaty Organization\)](#) -> [NAT \(Network Address Translation\)](#)

Reverse stemming when presenting to the user

[the boy's cars are different colors](#)

[the boy car be differ color](#)



Bag of words

N-gram with N=1, called unigram

Words are taken individually, no relationships, no ordering

A document \approx vector of term weights

each dimension corresponds to a term (10,000+)

each value represents the relevance (simple term counts)

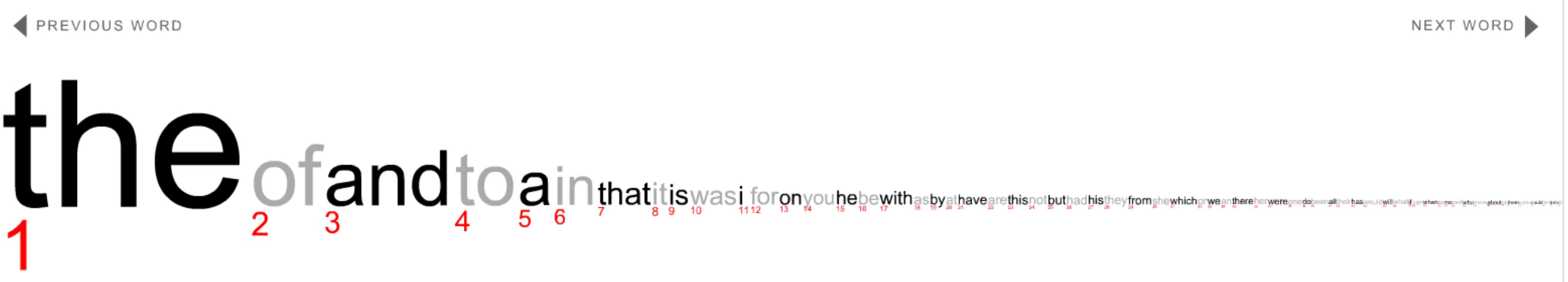
Aggregate into a document-term matrix

Document-term matrix

	Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	157	73	0	0	0	0
Brutus	4	157	0	1	0	0
Caesar	232	227	0	2	1	1
Calpurnia	0	10	0	0	0	0
Cleopatra	57	0	0	0	0	0
mercy	2	0	3	5	5	1
worser	2	0	1	1	1	0

Wordcount

WORDCOUNT



FIND WORD:

► BY RANK:

► REQUESTED WORD: THE

86800 WORDS IN ARCHIVE

RANK: 1

ABOUT WORDCOUNT

Tag cloud



Design critique

Can you interpret the visualization?

Do you trust the model?

Can you reason about the text?



Tag cloud

Pros

Can help with an initial query e.g. identifying the main words.

Cons

Bad visual encoding (size vs. position)!

No structure

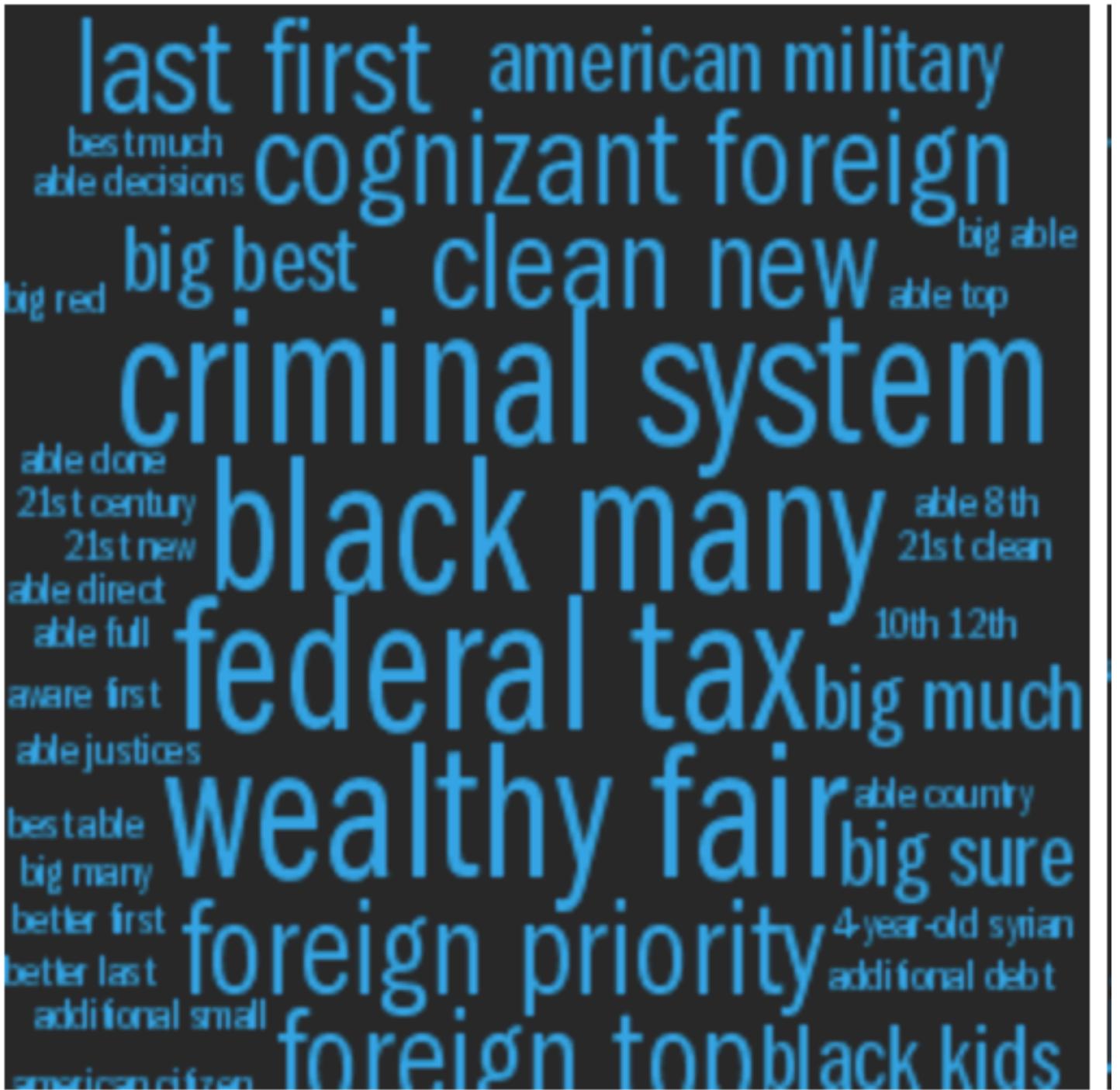
Inaccurate size encoding (long words are bigger)

Layout is unstable, difficult to compare two tag clouds

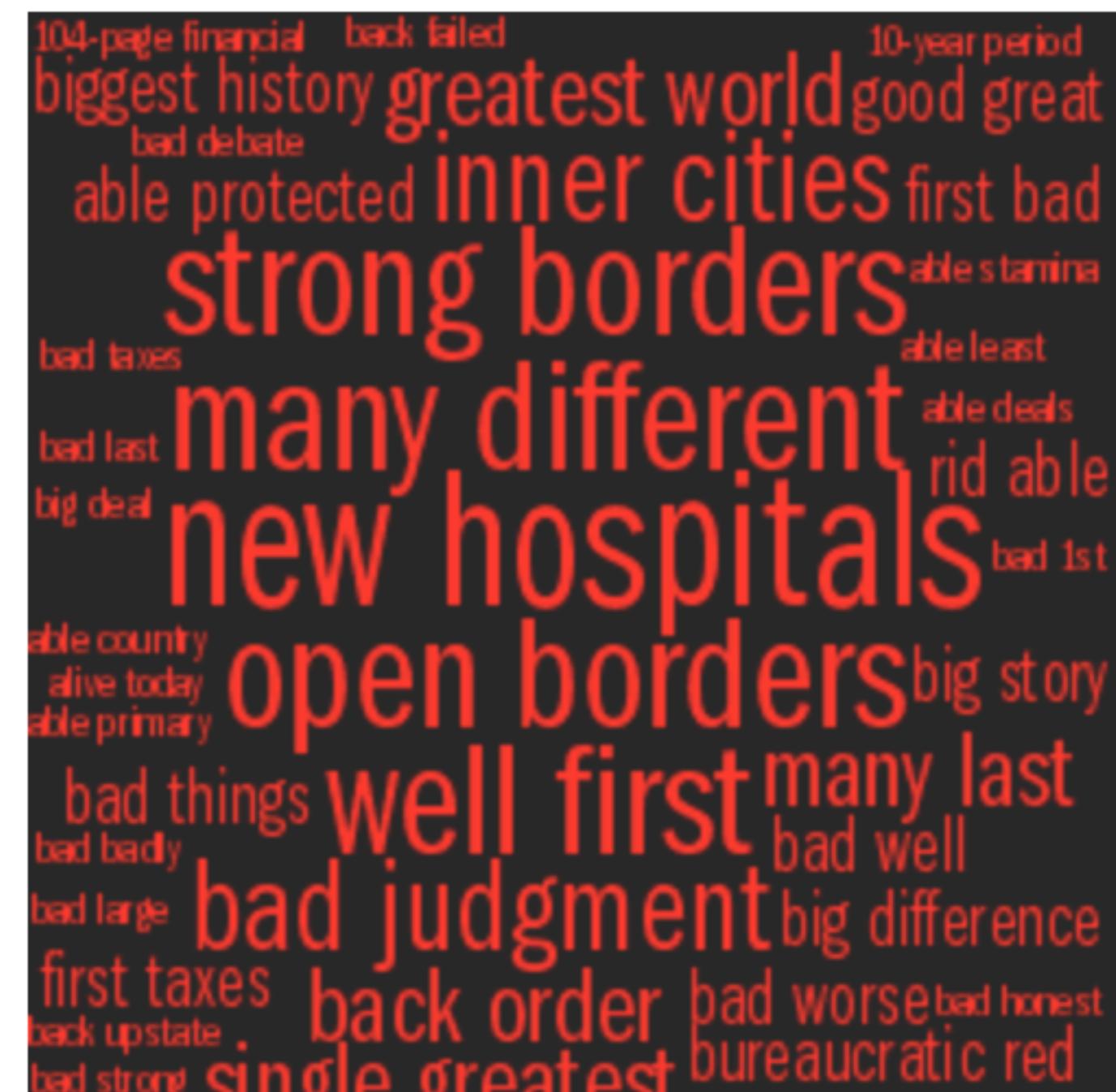


Words unique to each candidate and those used by both.

Clinton Trump both

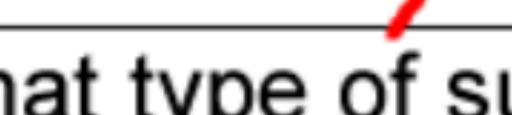


^ adjective/adjective by Hillary Clinton



^ adjective/adjective by Donald Trump

Adjective-noun pairs

'09 amazing around baked bar bass best chef delicious eat
elite everything favorite fish food fresh going hamachi
hawaiian hour line love mango minutes mussels name
night nigiri order people really restaurant roll
expensive or cheap? 
sake salmon sea seated service spicy stars sure
table think tuna 
"long wait" or "no wait"?  what type of sushi roll?

b) best sf
baked sea bass 
fresh fish 
sushi chef 
long wait 
baked mango 
small place 
delicious everything

Mentioned 63 times

possess sage of the halos wisdom , and know in advance sushi zone only accepts cash and the waits will be **long** and arduous .

yes , its a **long** wait , learn the master of zen if you want to eat here .

we came here early to try to avoid the **long** wait most people here talk about .

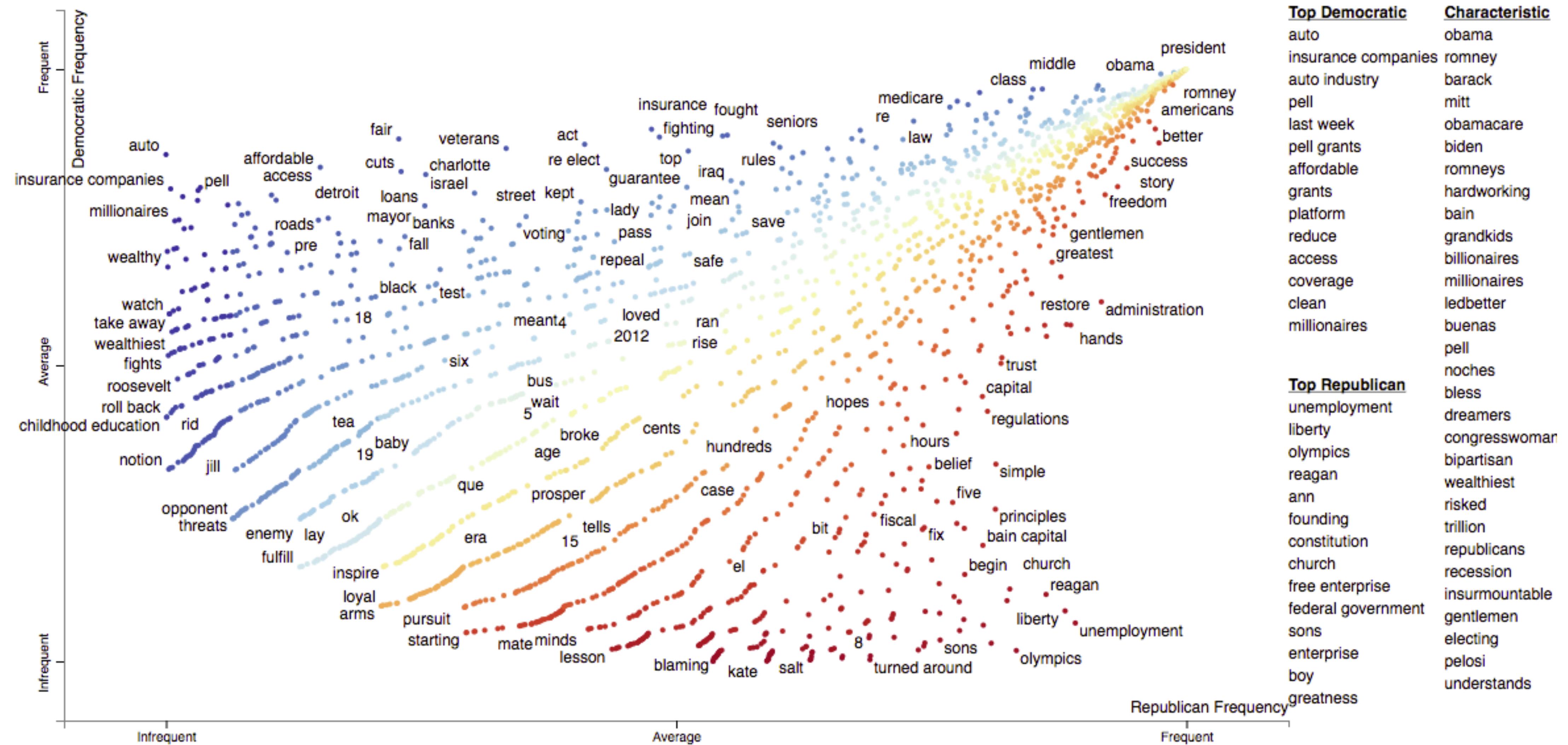
Tips

Use adjective-noun word pairs for reviews, more descriptive than bag of word model

Generate longer, more meaningful phrases

Show keyphrases within source text

Term frequency



Democratic document count: 123; word count: 76,861
Republican document count: 66; word count: 58,138

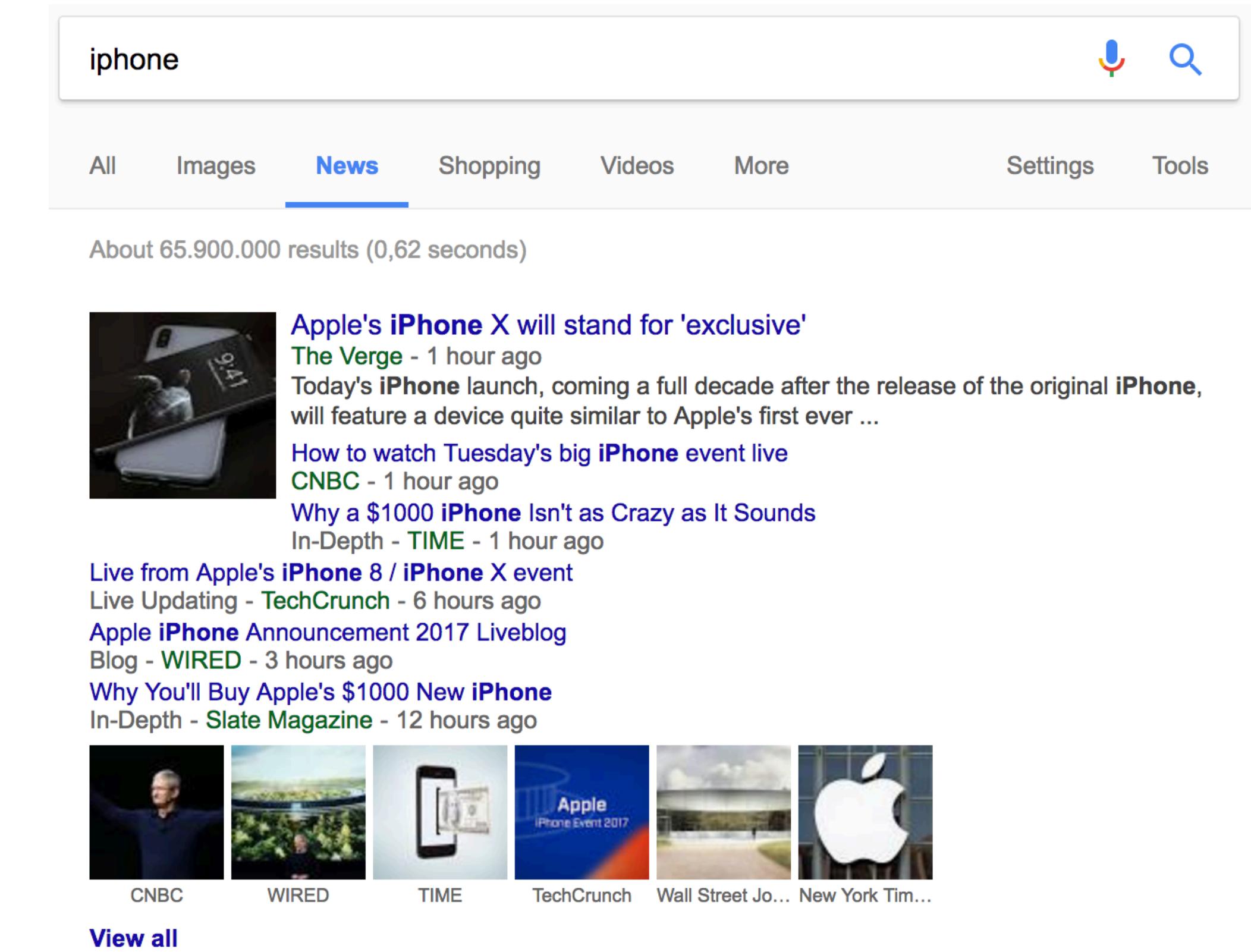
[Kessler 2017]

Document viz

Information retrieval

Match query string with documents

Provide context with picture and **query string in bold**



iphone

All Images **News** Shopping Videos More Settings Tools

About 65.900.000 results (0,62 seconds)

 Apple's iPhone X will stand for 'exclusive'
The Verge - 1 hour ago
Today's iPhone launch, coming a full decade after the release of the original iPhone, will feature a device quite similar to Apple's first ever ...

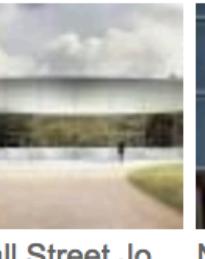
How to watch Tuesday's big iPhone event live
CNBC - 1 hour ago

Why a \$1000 iPhone Isn't as Crazy as It Sounds
In-Depth - TIME - 1 hour ago

Live from Apple's iPhone 8 / iPhone X event
Live Updating - TechCrunch - 6 hours ago

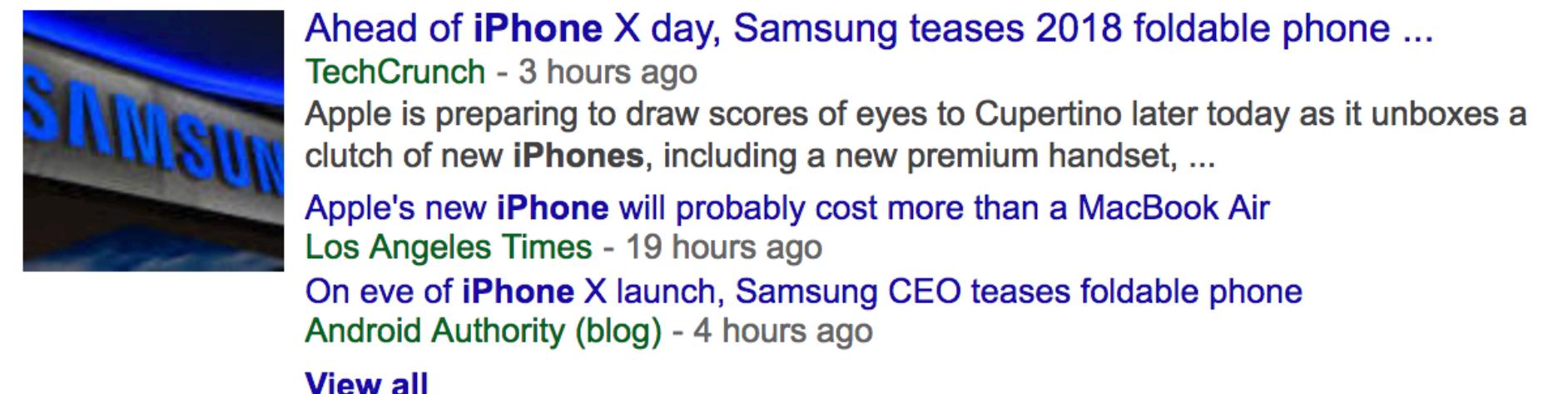
Apple iPhone Announcement 2017 Liveblog
Blog - WIRED - 3 hours ago

Why You'll Buy Apple's \$1000 New iPhone
In-Depth - Slate Magazine - 12 hours ago

CNBC WIRED TIME TechCrunch Wall Street Jo... New York Tim...

[View all](#)

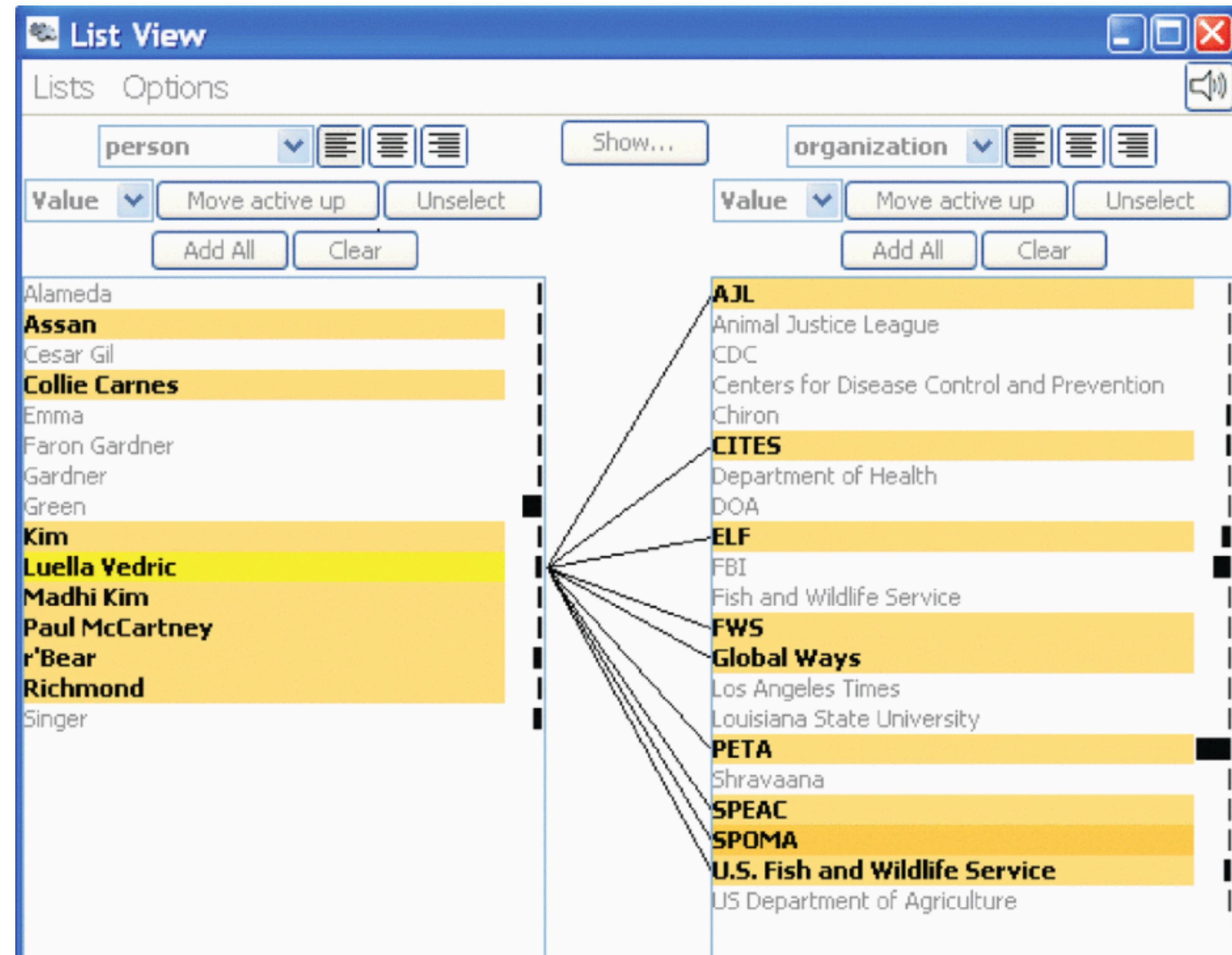


 Ahead of iPhone X day, Samsung teases 2018 foldable phone ...
TechCrunch - 3 hours ago
Apple is preparing to draw scores of eyes to Cupertino later today as it unboxes a clutch of new iPhones, including a new premium handset, ...

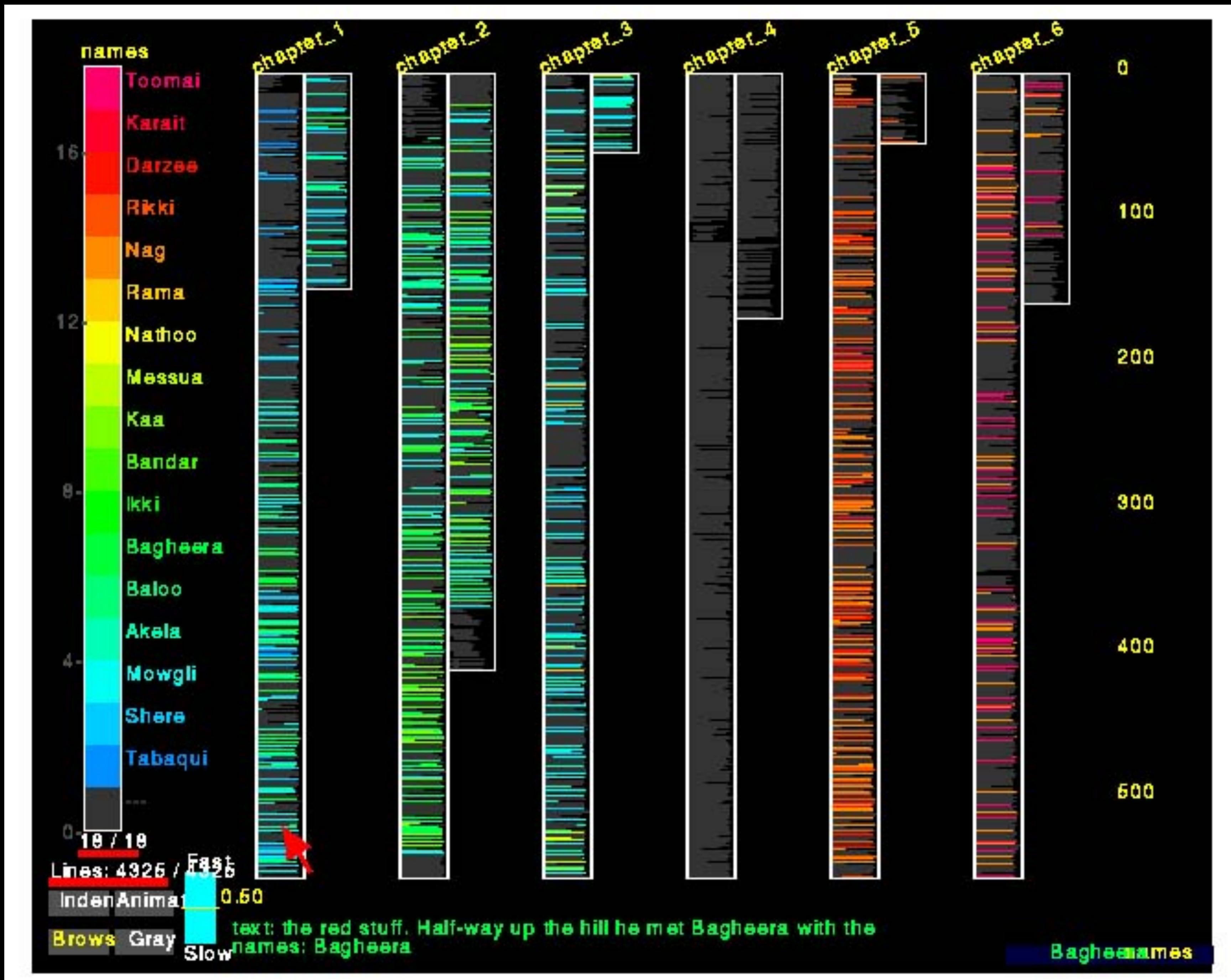
Apple's new iPhone will probably cost more than a MacBook Air
Los Angeles Times - 19 hours ago

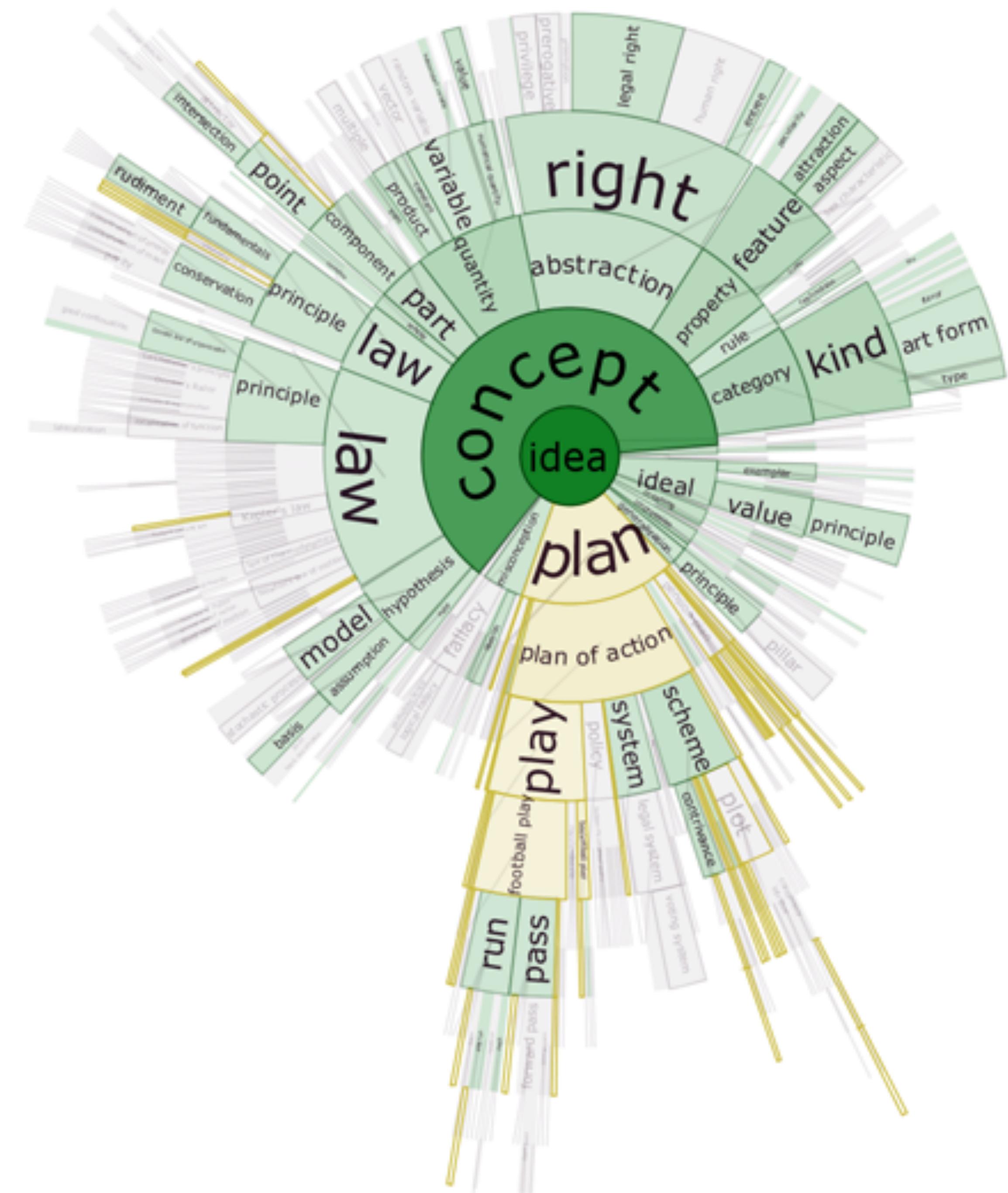
On eve of iPhone X launch, Samsung CEO teases foldable phone
Android Authority (blog) - 4 hours ago

[View all](#)



[Gorg et al., 2007]





[Collins, 2006]

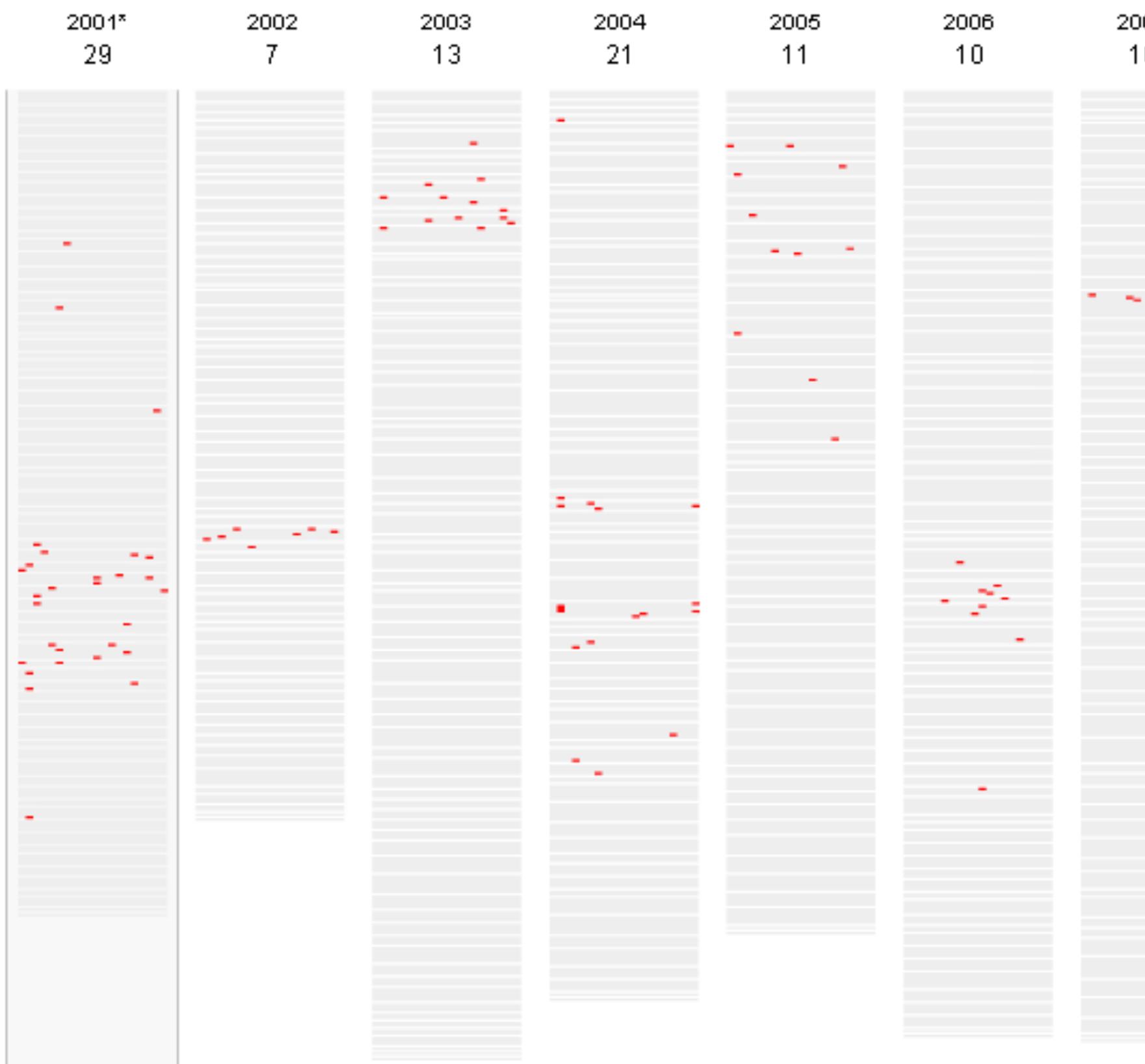
The 2007 State of the Union Address

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

Search

or choose a word here.

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



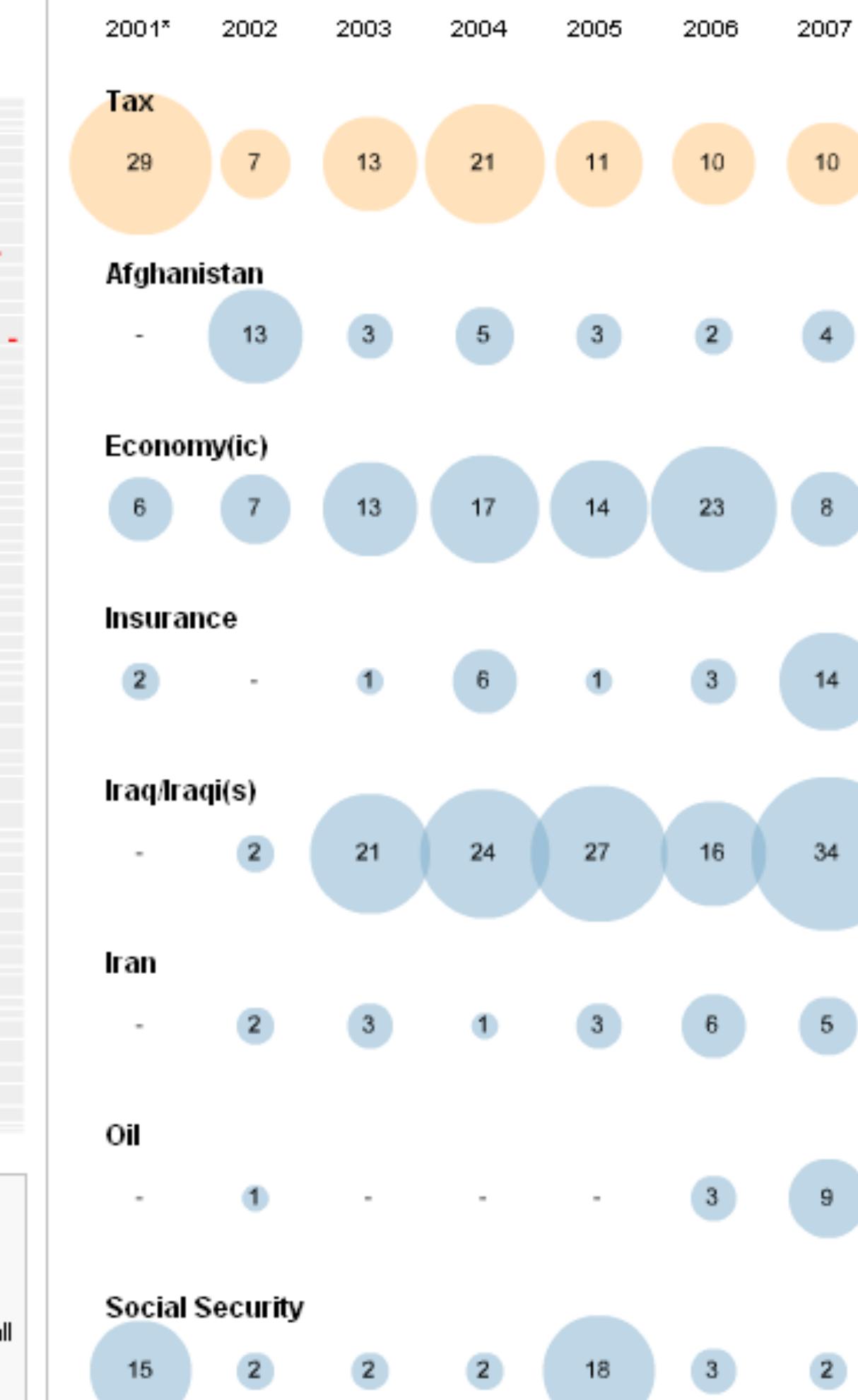
The word in context

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

Next Instance of 'Tax'

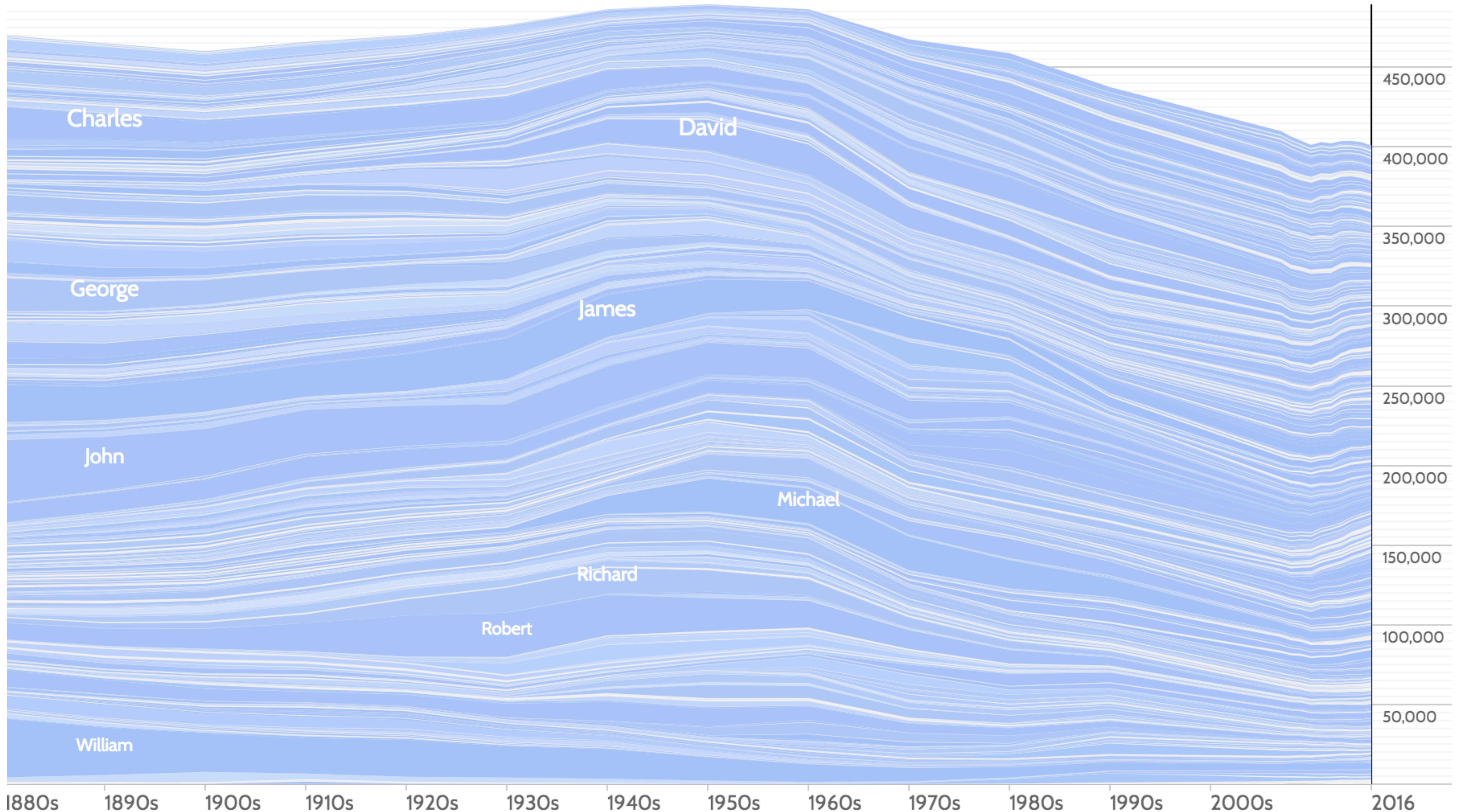
-- 2001 (Paragraph 14 of 73)

Compared with other words

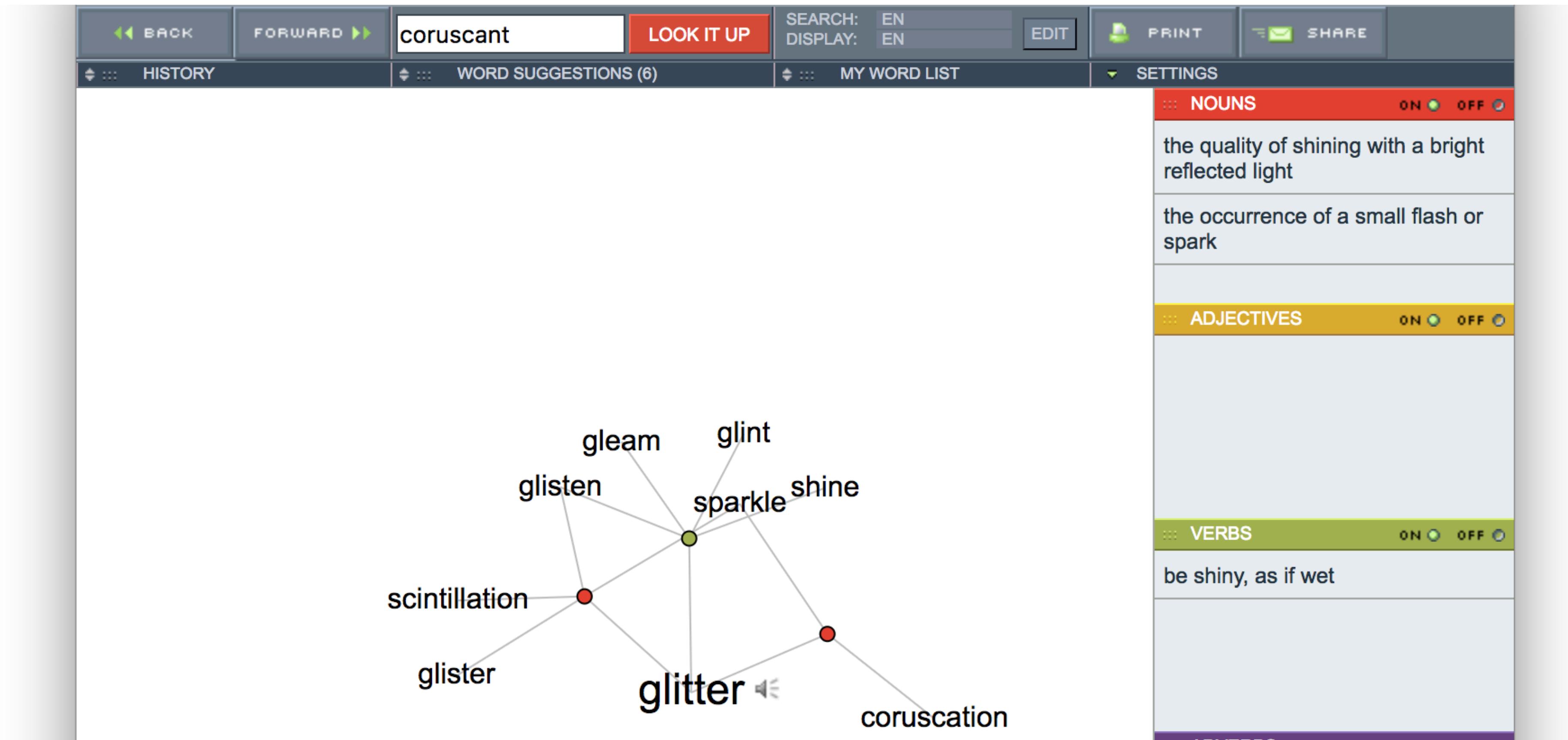


* As a newly elected president, Mr. Bush did not deliver a formal State of the Union address in 2001. His Feb. 27 speech to a joint session of Congress was analogous to the State of the Union, but without the title.

Ben Werschkul/The New York Times



Common local context



Down the Rabbit-Hole

Hide text

Show concordance

Show thesaurus lookup

Show story line

Text

Show only KWIC index (Key Word in Context)

Down the Rabbit-Hole

Rabbit with pink eyes ran close by **him** 

I think it so VERY much out of the way to hear the **Rabbit** say to
but when the **Rabbit** actually TOOK A WATCH OUT OF ITS WAISTCOAT-
before seen a rabbit with either a waistcoat-pocket, or a watch to
down a large rabbit-hole under the hedge.

The **rabbit-hole** went straight on like a tunnel for some way,
was another long passage, and the **White Rabbit** was still in
corner, but the **Rabbit** was no longer to be seen: she found
It was the **White Rabbit** returning, splendidly dressed, with a
that she was ready to ask help of any one, so, when the **Rabbit**
sir-' The **Rabbit** started violently, dropped the white kid
surprised to see that she had put on one of the **Rabbit's** little
herself. The **Rabbit** Sends in a Little Bill

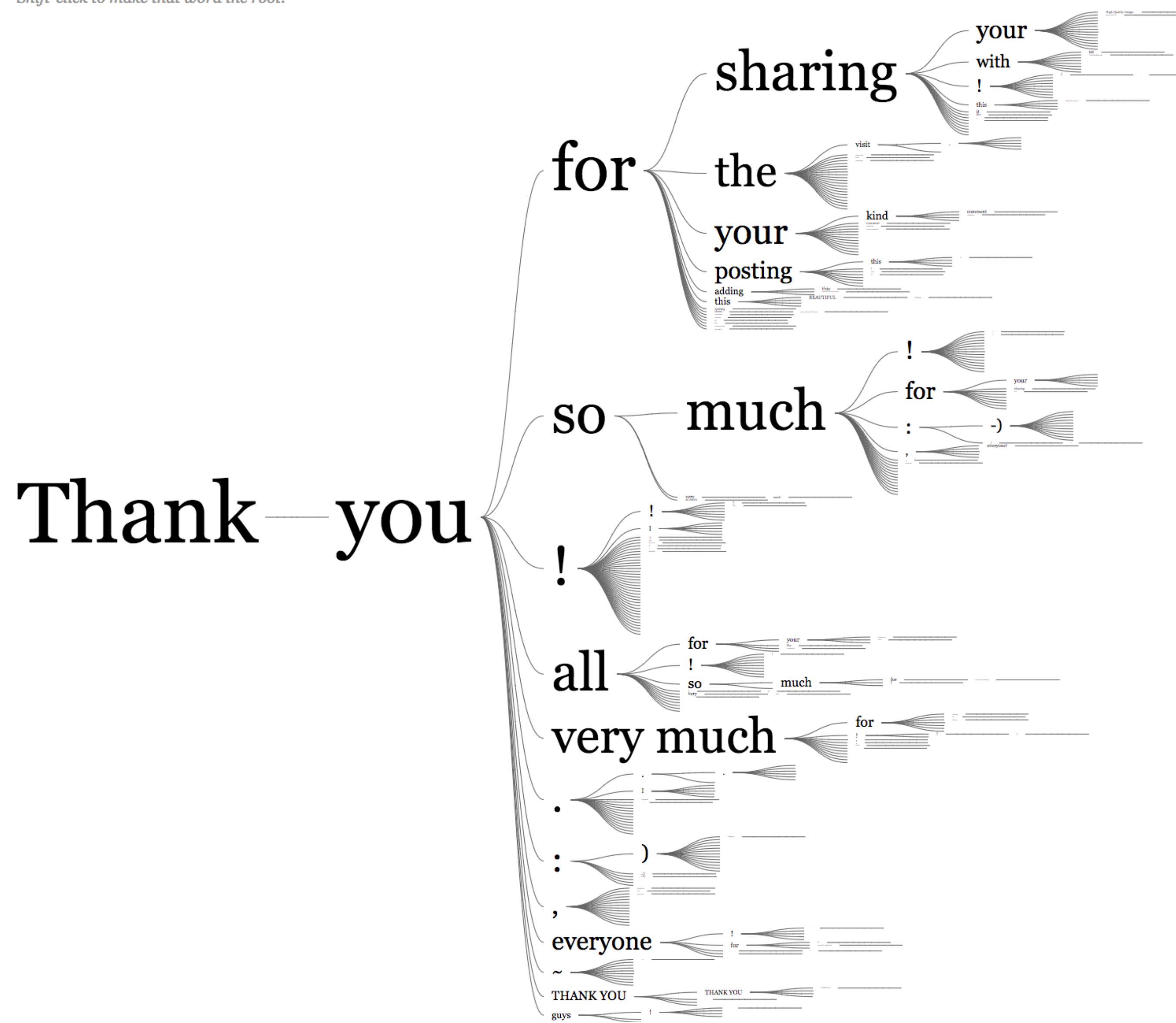
It was the **White Rabbit**, trotting slowly back again, and
Very soon the **Rabbit** noticed Alice, as she went hunting about,
messages for a rabbit. I suppose Dinah'll be sending me on
by mice and rabbits. I almost wish I hadn't gone down that
rabbit-hole--and yet--and yet--it's rather curious, you know,
stair. Alice knew it was the **Rabbit** coming to look for her, and
was now about a thousand times as large as the **Rabbit**, and had no
Presently the **Rabbit** came up to the door, and tried to open it:
lanced she heard the **Rabbit** just under the window, she suddenly
Digging for apples, indeed' said the **Rabbit** angrily. 'Here!
began moving about again, and Alice heard the **Rabbit** say, 'A
recognised the **White Rabbit**: it was talking in a hurried nervous
She was walking by the **White Rabbit**, who was peeping anxiously
'Hush! Hush!' said the **Rabbit** in a low, hurried tone. He
Did you say 'What a pity!?' the **Rabbit** asked.
She boxed the Queen's ears-- the **Rabbit** began. Alice gave a
little scream of laughter. 'Oh, hush!' the **Rabbit** whispered in a
she first saw the **White Rabbit**. She was a little nervous about
each side to guard him; and near the King was the **White Rabbit**,
she stopped hastily, for the **White Rabbit** cried out, 'Silence in
nuttered On this the **White Rabbit** blew three blasts on the trumpet, and
'Not yet, not yet!' the **Rabbit** hastily interrupted. 'There's
'Call the first witness,' said the King; and the **White Rabbit**

word tree

Thank you

reverse tree one phrase per line

Shift-click to make that word the root.



Seen @ Update Collection
Thank you for sharing your
High Quality Image.WOW
congratulations on your
accomplishments and nice
earnings.these are g o r g e
o u s ! greeen with envy....
.Cooool!!! Excellent!
congrats:) and great shot:)
Congratulations! This is
fabulous!lovely earings. And
congrats!i have those too!!
COOL shot! I love the
earings! I NEED some like
this!!!Awww those are so
cute!so cool!!!Cute !! :)
That's awesome
congratulations!! I've been
published three times, each
time feels like the first!
LOL. Great earrings I sooo
covet them.HI Helen...so
good to see you posting. I
would love a pair of those..
really so sweet!! Wonderful
presentation and focus.
Congratulations to you!!Nice
one Helen. I like them. --
Seen on your photo stream. (
?)Oh i Love these earings!
! Really need a pair right
now to give me a boost!

WordTree review

Pros

useful for n-gram visualisation $n > 1$

show recurrent concordance (common local context) of repeated structures

Cons

Starting word?

What about other type of structures (Syntactic, lexical) ?

Phrase nets

Look for specific linking patterns in the text: 'A and B', 'A at B', 'A of B', etc

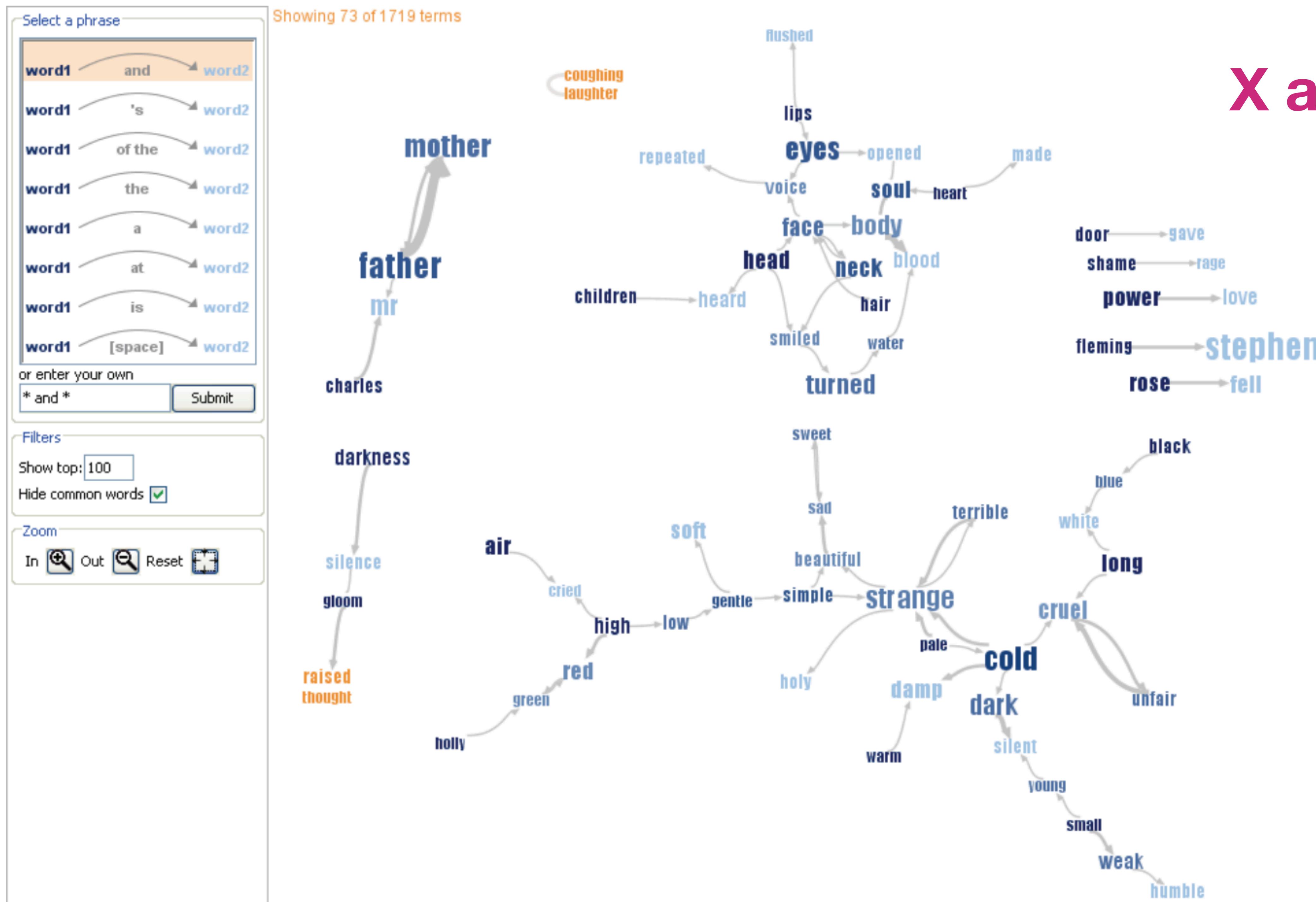
Visualize patterns in a node-link view

Node size is proportional to occurrences

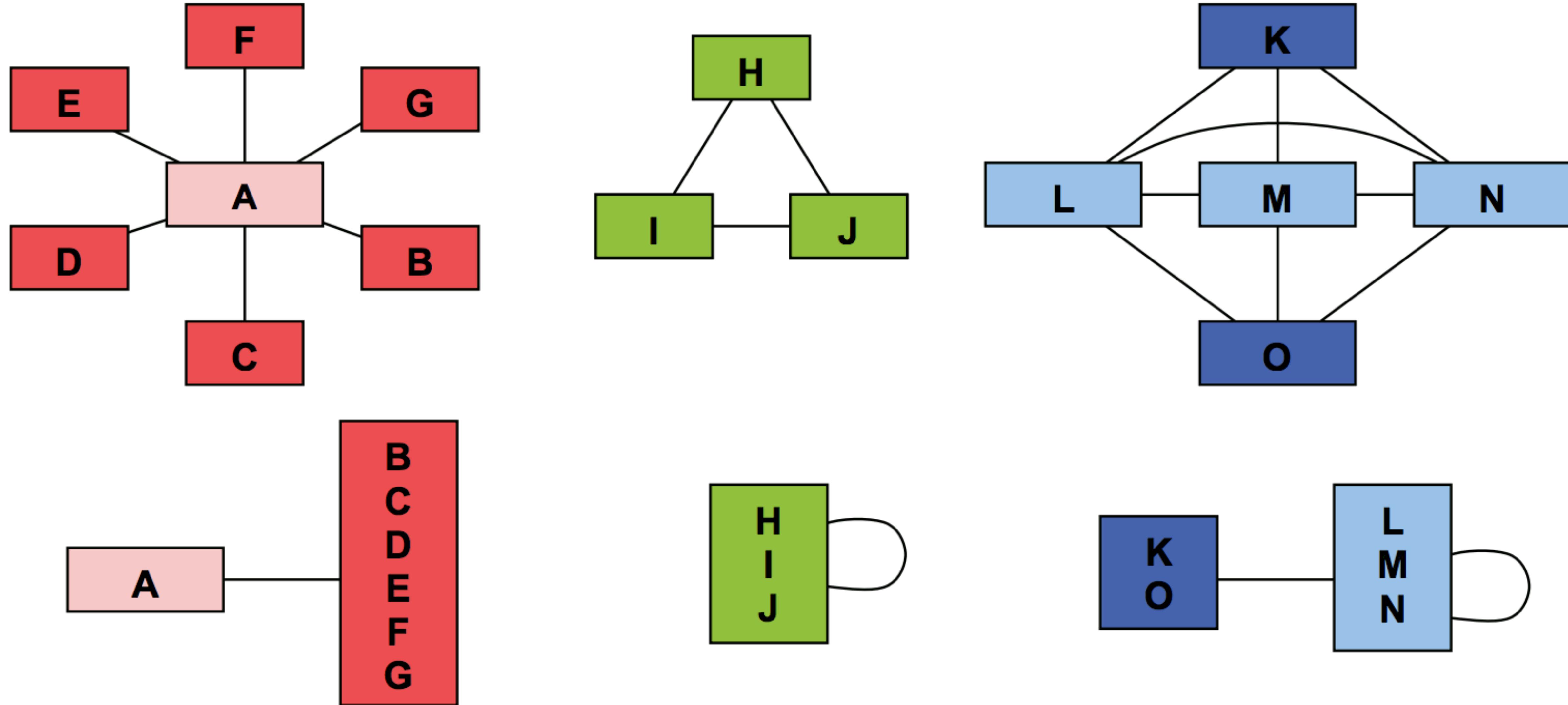
Edge direction follows the text pattern

[van Ham et al.]

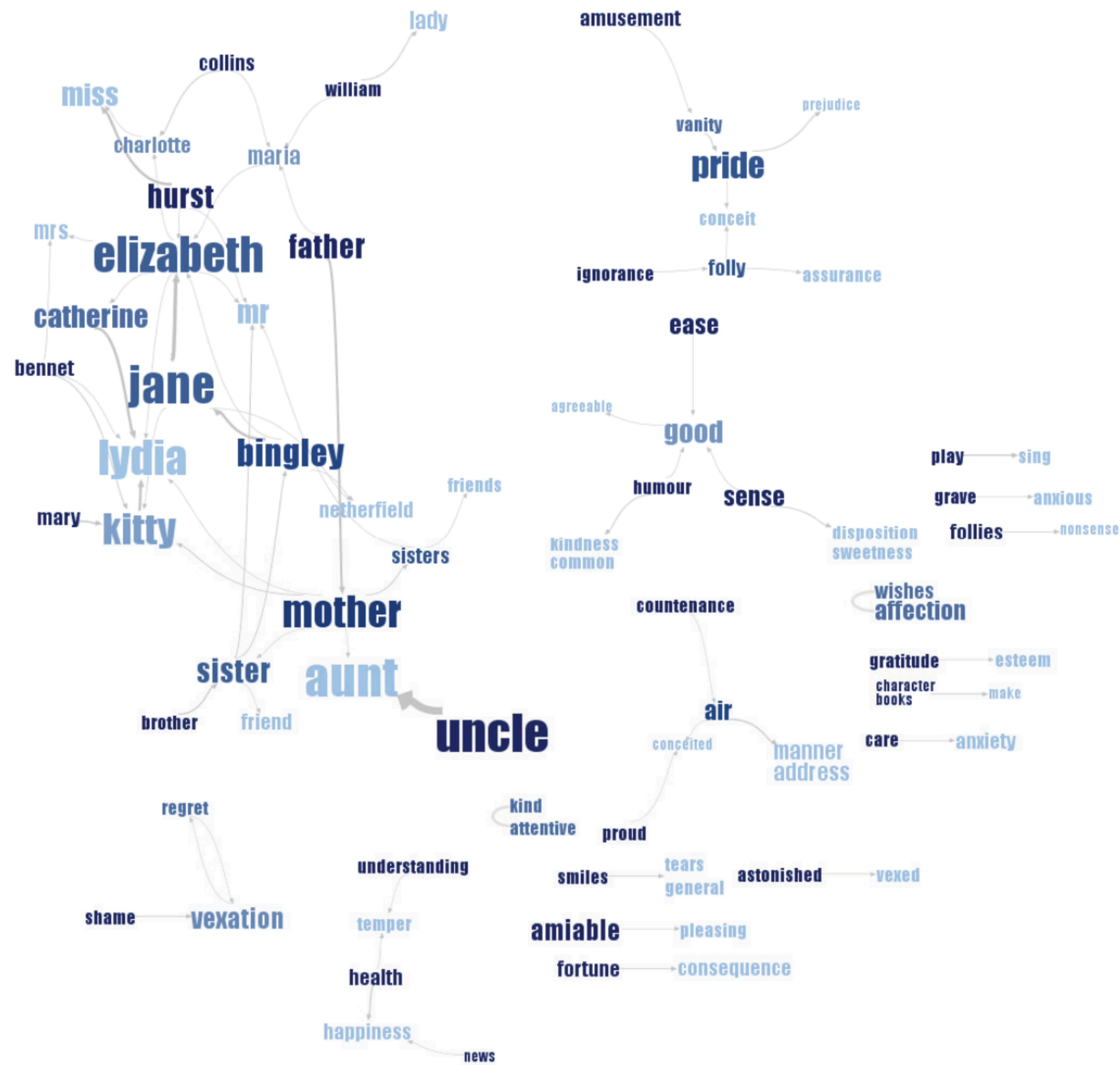
X and Y



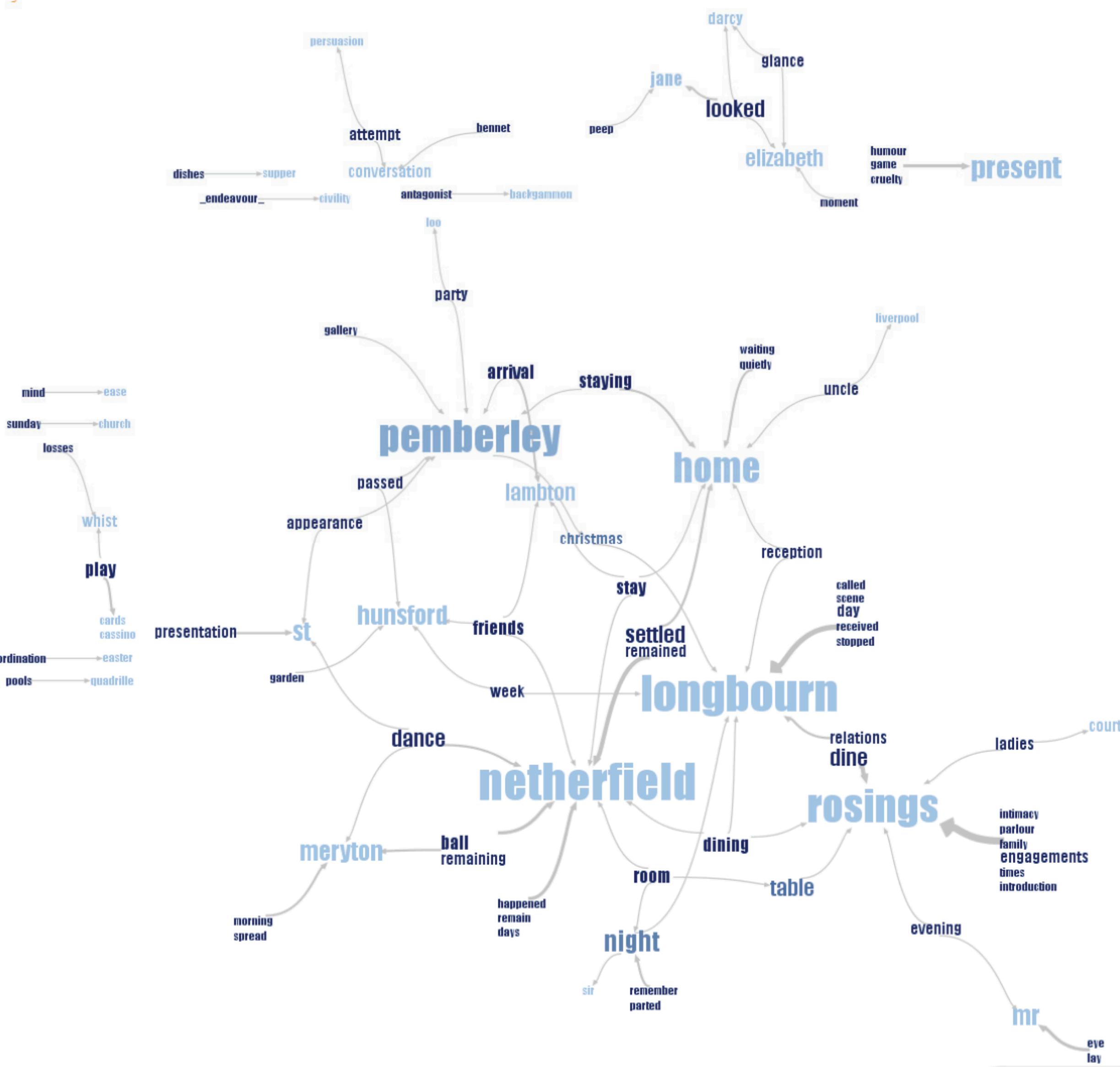
Edge compression



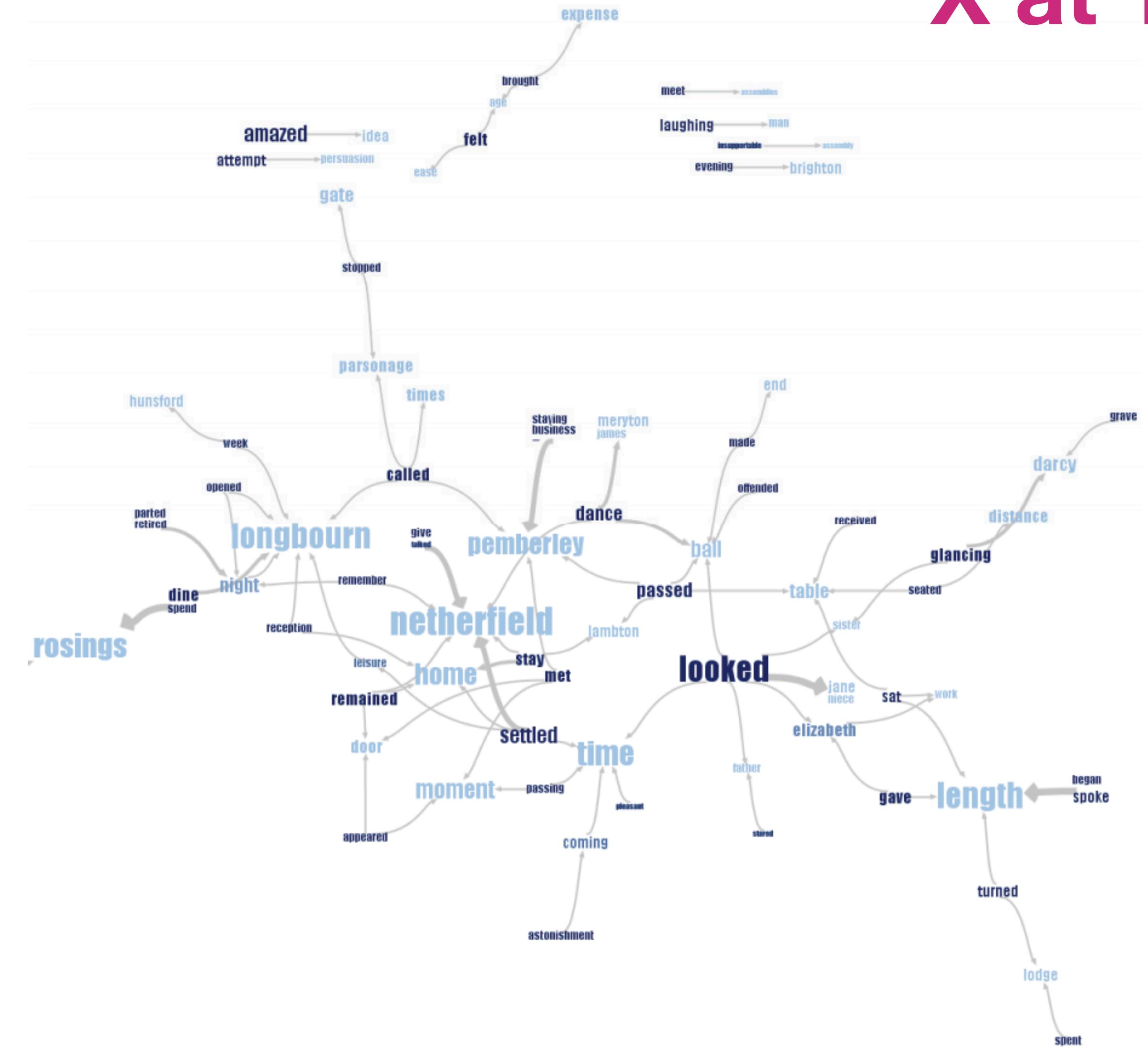
X and Y



X at Y



1 sec



24 hours computing power

Document content viz

Visual aspect: Word position, browsing, brush & link

Semantic aspect: Word sequence, hierarchy, clustering

Both: Spatial layout reflects semantic relationships

Interactions should support visual analysis + modifications of the model

Evolving documents

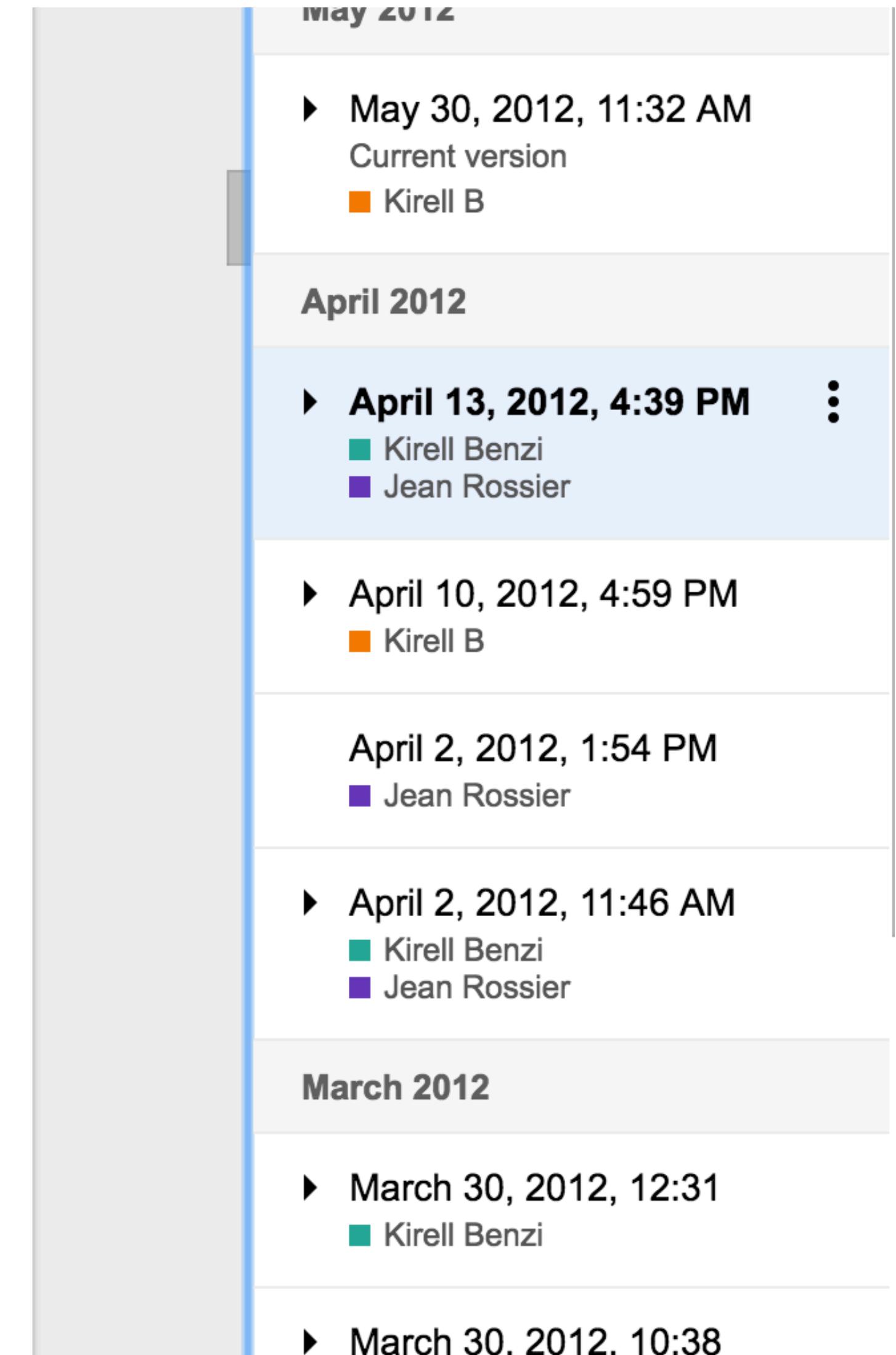
How to depict contributions?

Get urls

GET /api/urls

200 OK

```
{  
  "latest_server_api_version": 1,  
  "latest_features_api_version": 1,  
  "latest_client_version": 0.25,  
  "urls": [  
    { "key": "check_song",  
      "method": "GET",  
      "url": "api/v1/song/'songhash'",  
      "description": "Check if song is in database"  
    },  
    { "key": "check_song",  
      "method": "GET",  
      "url": "api/v1/song/'songhash'",  
      "description": "Check if song is in database"  
    },  
    ...  
  ]}  
  "latest_server_api_version": "1.0.0",  
  "latest_features_api_version": "1",  
  "latest_client_version": "1.0.0",
```



aspbern papers.pages (Word Processing)

View Sections Columns Style List Character Wrap Objects Comment Inspector Media Adjust Colors Fonts

Tracking Bubbles + Tracking: On Paused Accept Reject View Markup +

Kirk McElhearn ✓ X
Yesterday, 5:48 PM
Replaced: confidence >

Kirk McElhearn ✓ X
Yesterday, 5:46 PM
Added Text

Kirk McElhearn ✓ X
Yesterday, 5:47 PM
Replaced: advance >

Kirk McElhearn ✓ X
Yesterday, 5:48 PM
Replaced: invented >

Kirk McElhearn Today, 11:06 AM How about cut for this? ✎

Kirk McElhearn ✓ X
Today, 11:03 AM
Deleted: as a general thing >

Kirk McElhearn ✓ X
Yesterday, 5:47 PM
Replaced: throw off a >

Kirk McElhearn ✓ X
Yesterday, 5:47 PM
Added Text

Kirk McElhearn ✓ X
Yesterday, 5:47 PM
Deleted: conception >

The Aspern Papers
by Henry James

I

I had taken Mrs. Prest into my confidence; in truth, without her I should have made but little advance~~progress~~, for the fruitful idea in the whole business dropped from her friendly lips. It was she who invented~~proposed~~ the short cut, who severed the Gordian knot. It is not supposed to be the nature of women to rise-as-a-general-thing to the largest and most liberal view—I mean of a practical scheme; but it has struck me that they sometimes throw-off-a-show bold attitudes~~oneception~~—such as a man would not have risen to—with singular serenity. "Simply ask them to take you in on the footing of a lodger"—I don't think that unaided I should have risen to that. I was beating about the bush, trying to be ingenious, wondering by what combination of arts I might become an acquaintance, when she offered this happy suggestion

150% 1 of 39,145 Words Page 1 of 133 ▲ ▼ ◌

Git commits

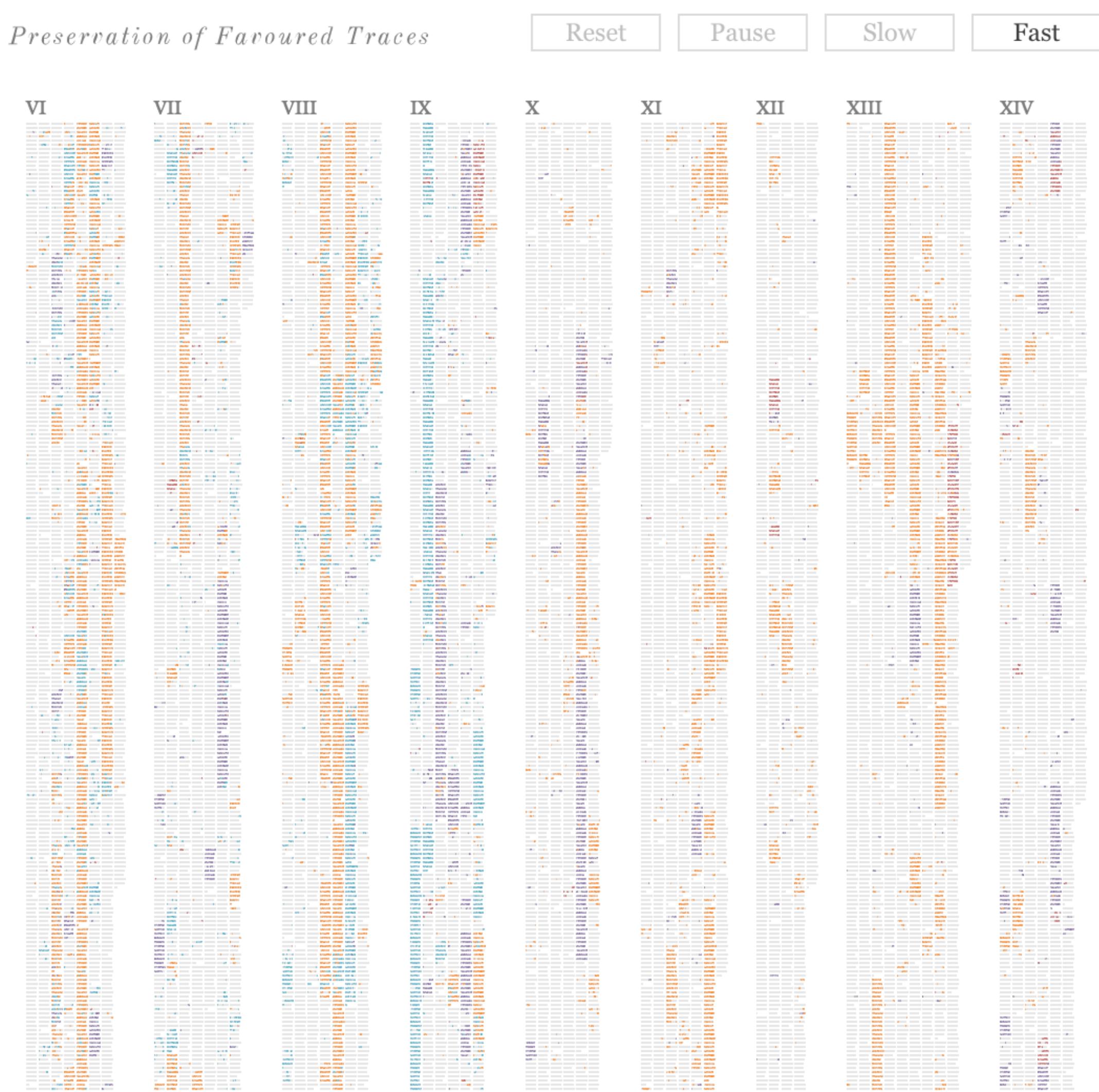
```
7 - <title>D3 Boilerplate</title>
7 + <title>D3 test charts</title>
8   8   <link rel="stylesheet" type="text/css" href="main.css">
9   9   </head>
10 10   <body>
11 -   <div id="viz"></div>
11 +   <div id="viz" class="svg-container"></div>
12 12   <script type="text/javascript" src="bundle.js"></script>
13 13   </body>
14 14 </html>
```

41  app/main.js View ▼

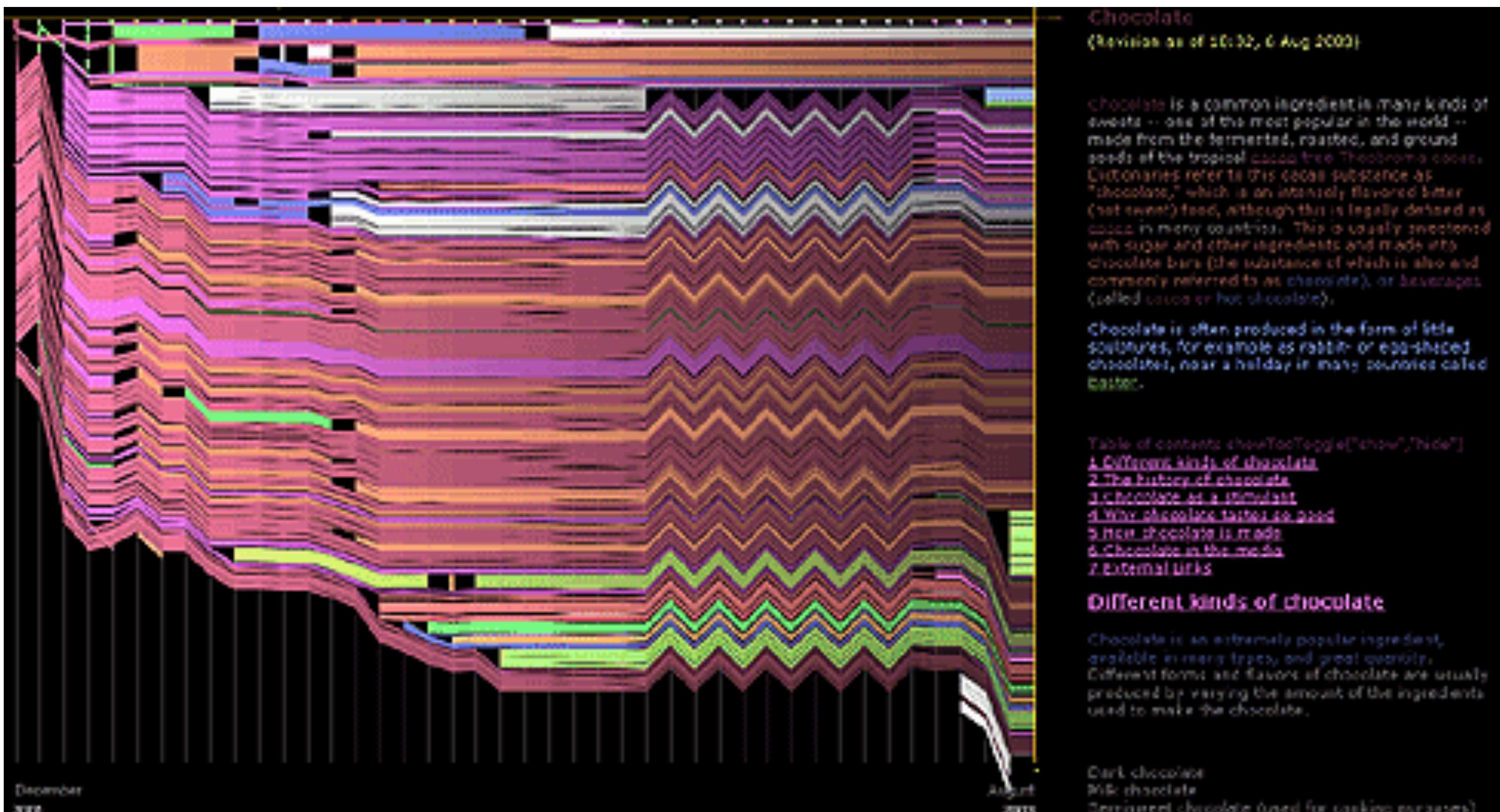
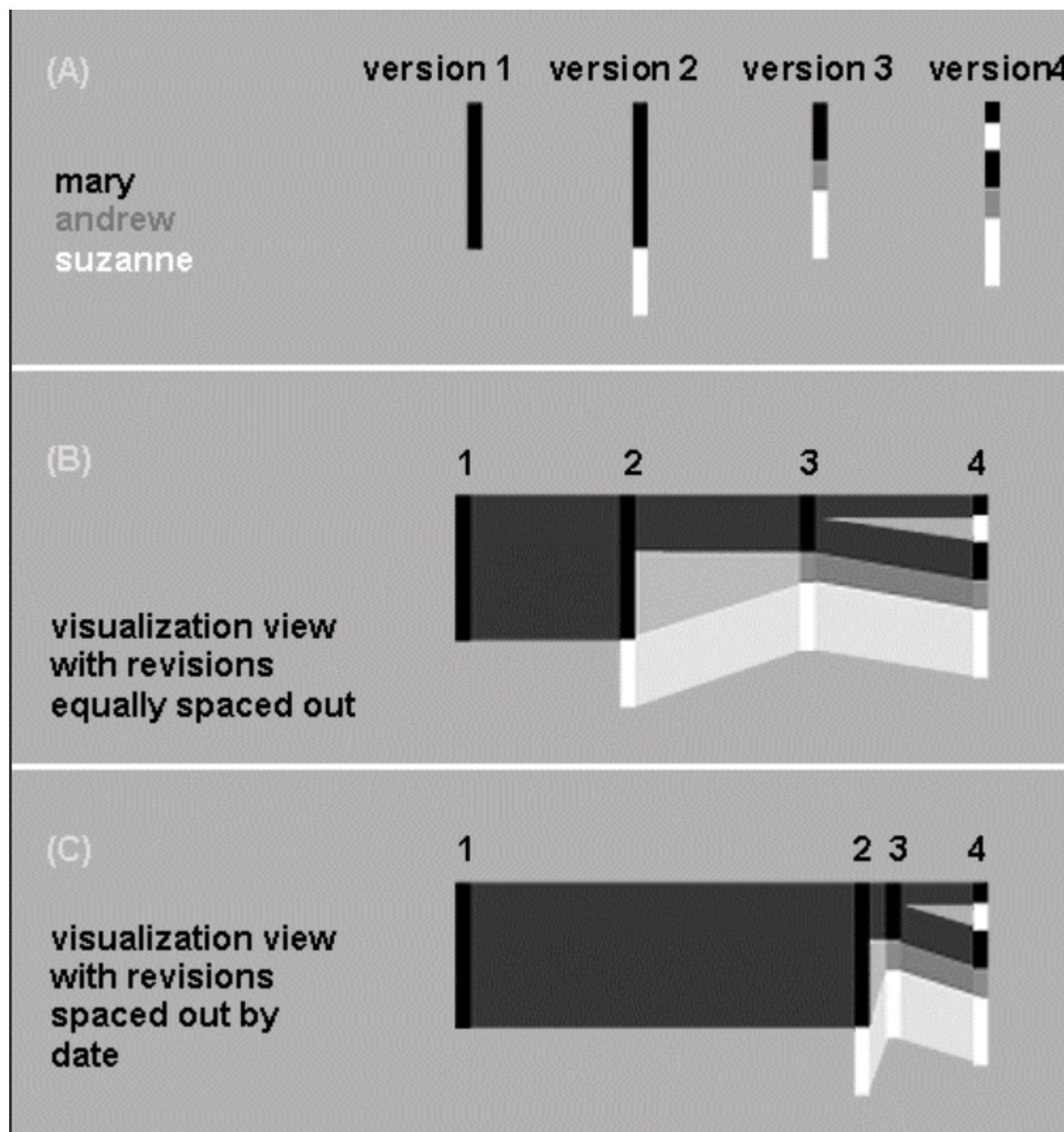
```
... ... @@ -1,21 +1,30 @@
1 -'use strict';
2 1
3 -require("./main.scss"); // will build CSS from SASS 3 file
4 -
5 -import _ from 'lodash'; // imports
6 2 import * as d3 from 'd3';
7 3
8 // Test ES6 and d3
9 -const data = [1, 2, 3];
10 -const [x, y] = [70, 100]; // Array destructuring
4 +// Local imports
5 +import { createCircleChart } from './circle'; // named import
6 +import makeBarChart from './bar'; // name is irrelevant since it is a default export
```

Animated traces

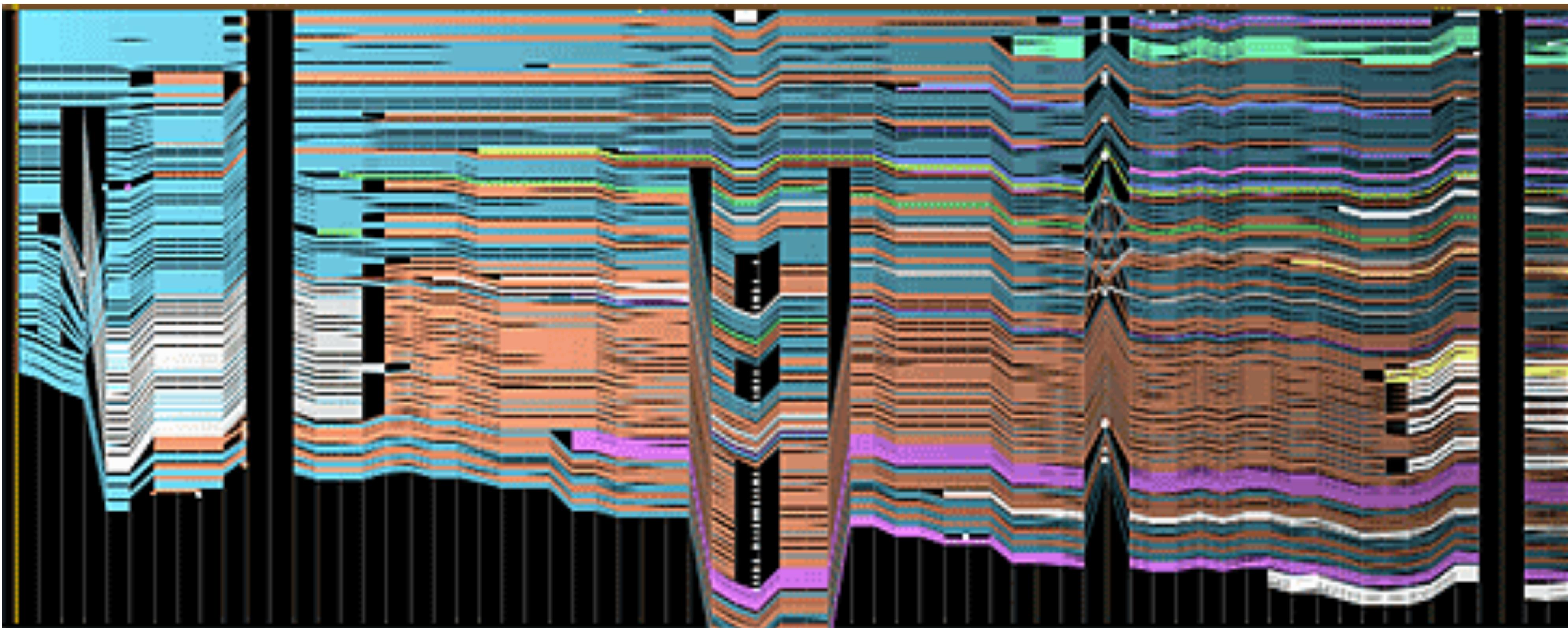
is to make up for much destruction at some period of life; and this period in the great majority of cases is an early one. If an animal can in any way protect its own eggs or young, a small number may be produced, and yet the average stock be fully



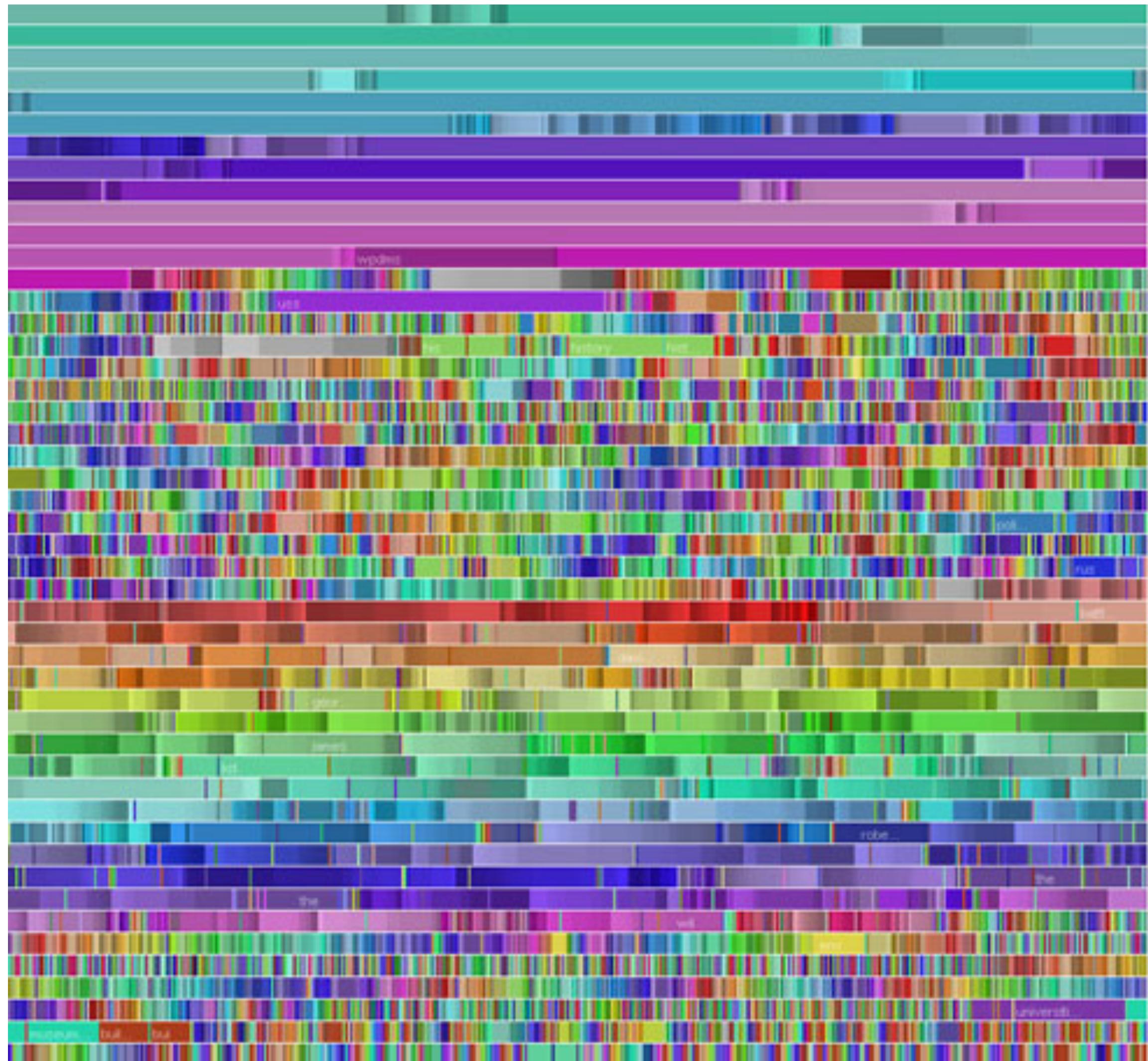
History flow

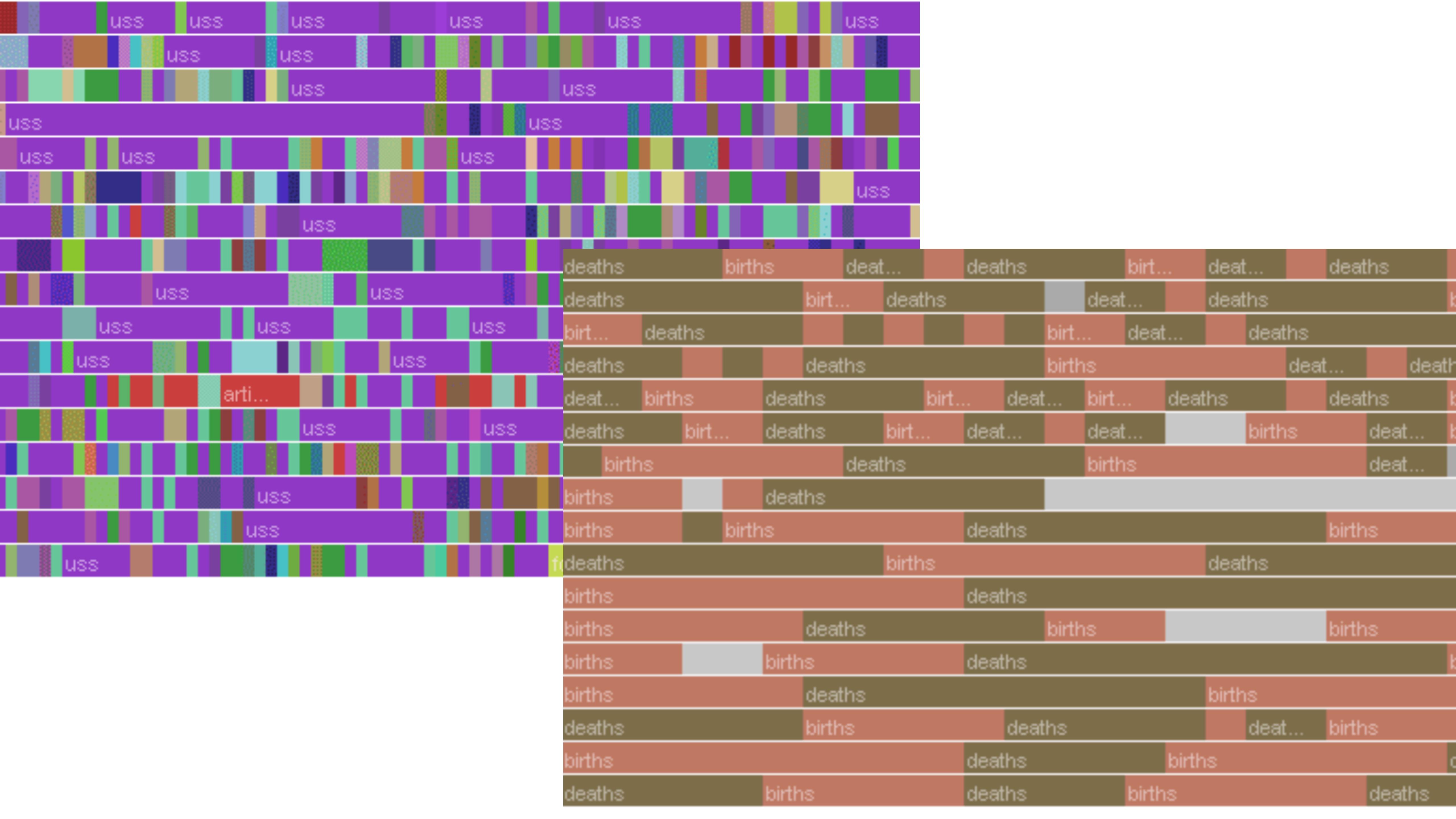


Visualize vandalism on Wikipedia



Chromogram





Conversation viz

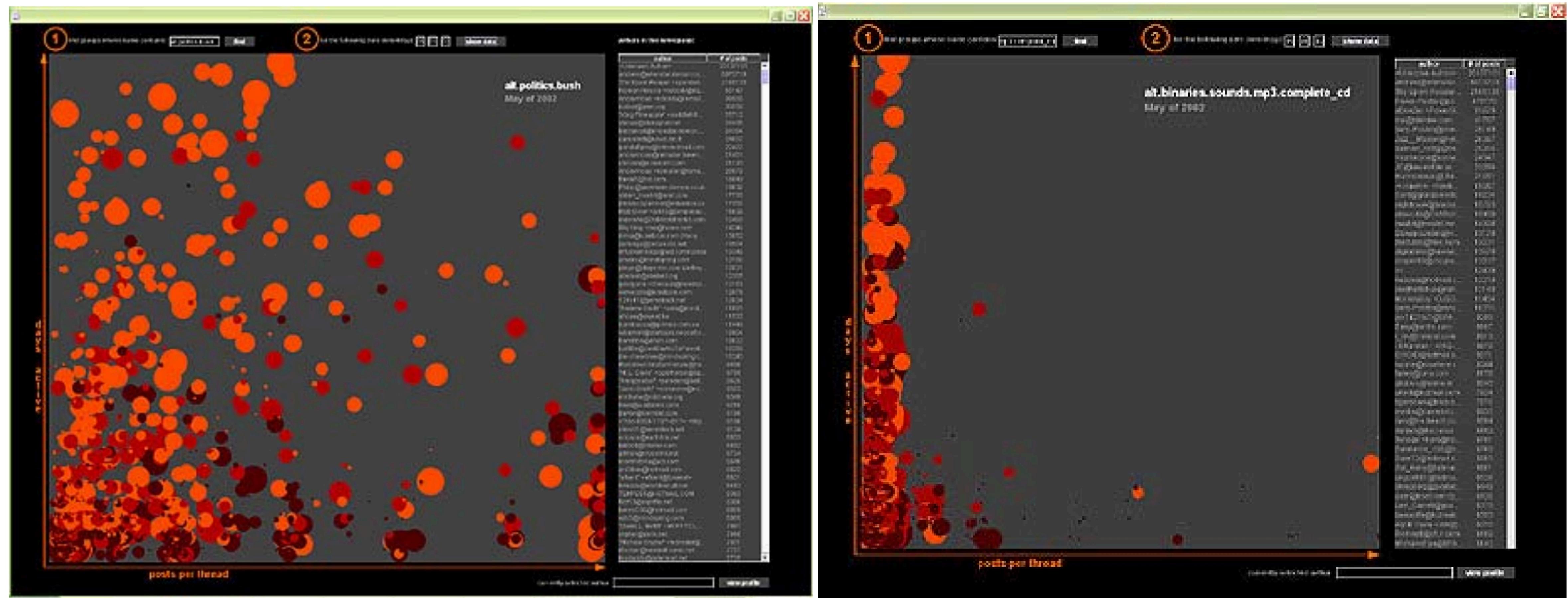
Senders, receivers (who)

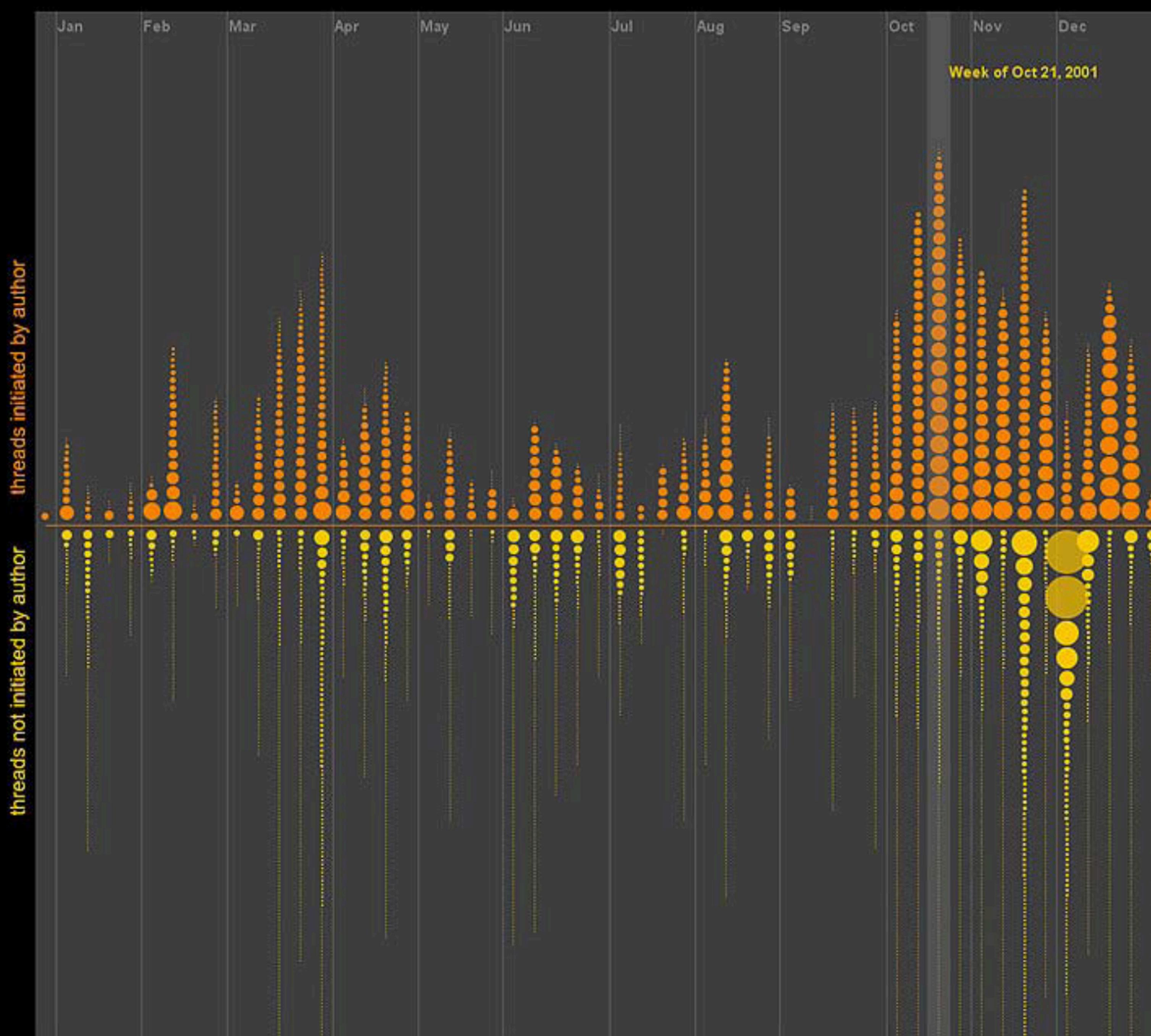
Content (what), temporal patterns (when)

Information flow: Who x Who x What x When

Newsgroup Crowds and Authorlines

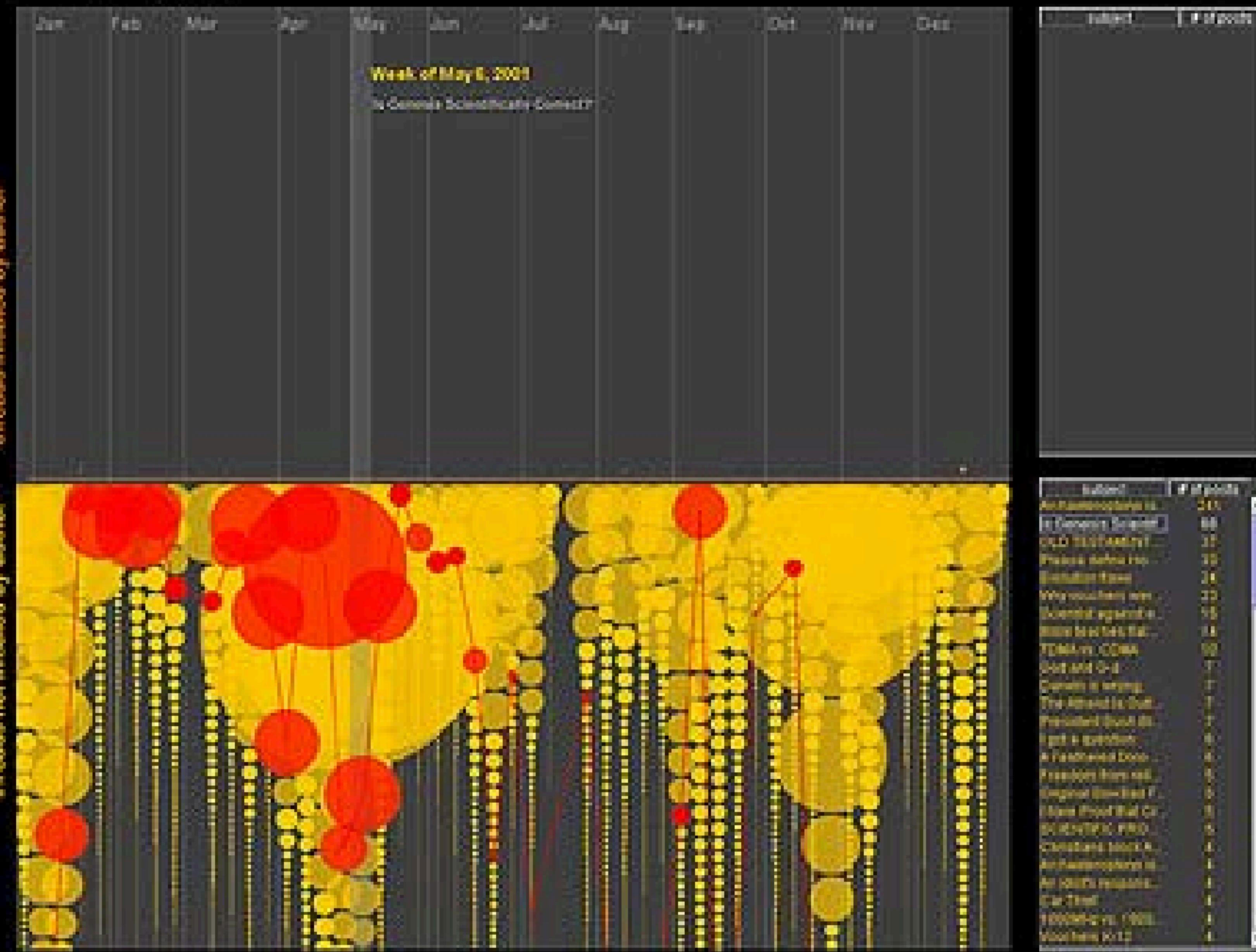
[Viegas & Smith]





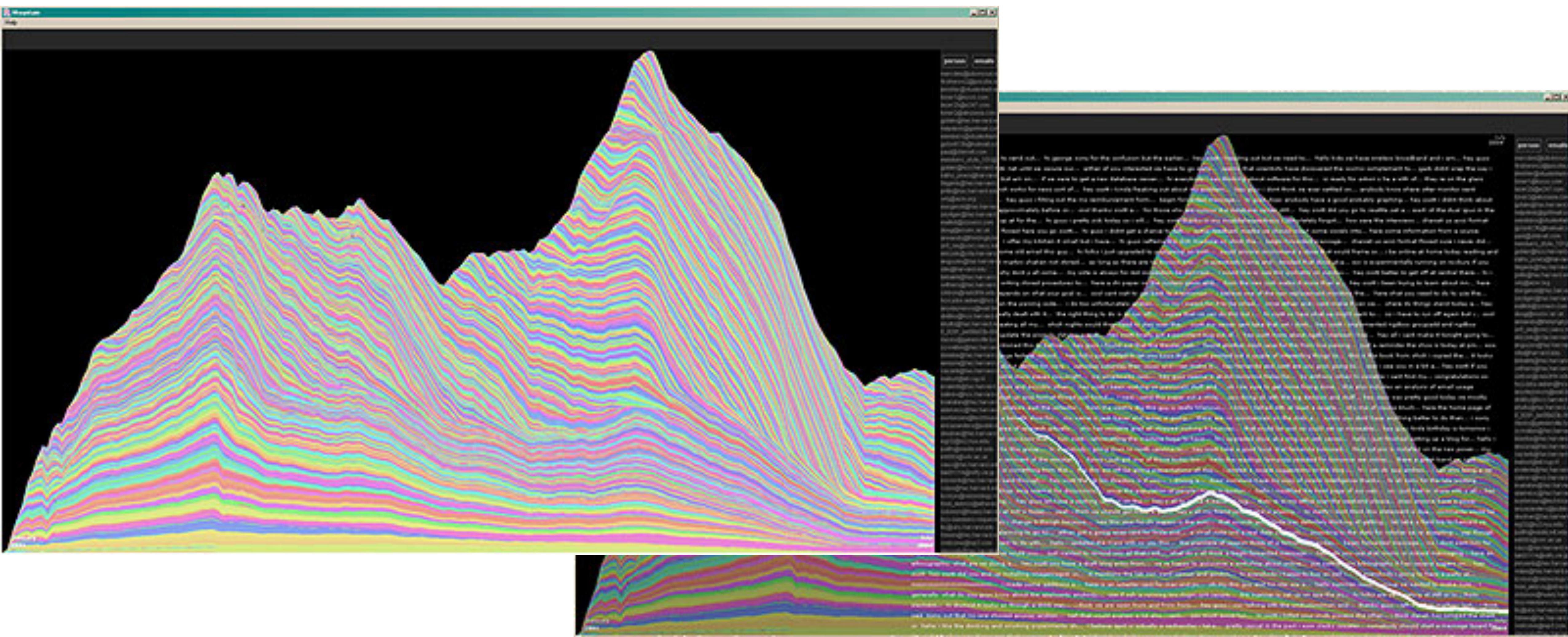
subject	# of posts
Wednesday Spooker ASF	21
WET #3 Anyone for breakfast	20
Sunny Side Up ASF)	18
Saturday Ensemble and WET	18
Oh no! Watch out! ASF	18
Thursday Combo-Post WET #...	16
The Yellow Rose Inn.....A gift to...	16
WET #1 JBP The First Time	16
We Love the Earth ASF	15
Monday Spooker "The Sight"	15
C'mon!!!	14
Theberge "Le Vent Se Lève"	14
Holiday Tog #3)	13
Spooker du Jour)	13
Beginning ASF Short and...	13
Second Try A Kalle for Suzy...	12
Come On a Safari With Me.....	11
Tuesday Spooker ASF	11
Curses, Foiled Again.....ASF	10
Halloween Togs Take Two)	9
Beauty of the Fury Jim Warren...	9
I thought I saw.....? ASF	7
Wednesday Evening at the Con...	4
Second Try A Kalle for Suzy...	2
Frank Was A Monster ASF...	1

subject	# of posts
Sunday Twofer ASF)	9
Chopsticks/A Jilly fake	8
Oh no! Trouble in Discworld!	7
WET.....your thirst! ASF	6
A pretty for you...Reposted fro...	5
Saturday Spooker ASF	5
Sample Previous install Upgr...	4
Tennessee weather tonite	4
WET - Well I am not smiling!	4
Somethin' mushy: <asf>	3
Getting seasonal with workin...	3
A Haunted House)	3
do you wonder what debi's be...	3
Question: Ethics of posters in...	3
For Jerry	3
Olu's Tribe - slightly rated	3
WET - Glass Bottles	3
Peace Train<ASF>	2
Arrival at Stewart Island II	2
WET 195. Wrap-up	2
Cat O'Lantern	2
I Put a Spell on You (Happy H...	2
Goodbye to Summer - A Timel...	2
Two Pumpkins In A Strange B...	2
Still Heading South II	2
WET- Frank Sinatra - The Man...	2
WET Autumn	2
Purple Martin ASF	2
Opposites Attract...	2
Time	2



Mountain (email)

[Viegas et al.]

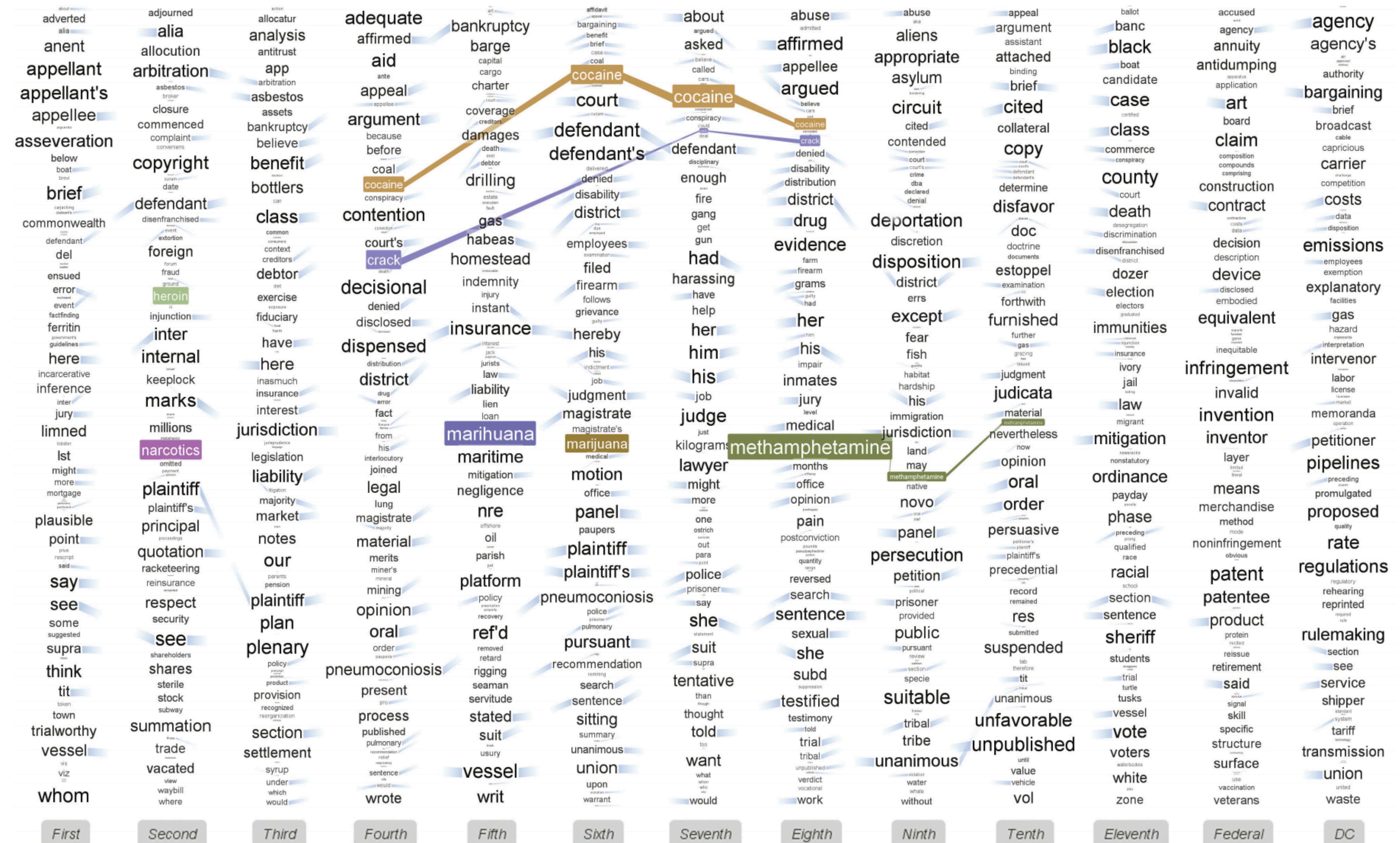


The mail

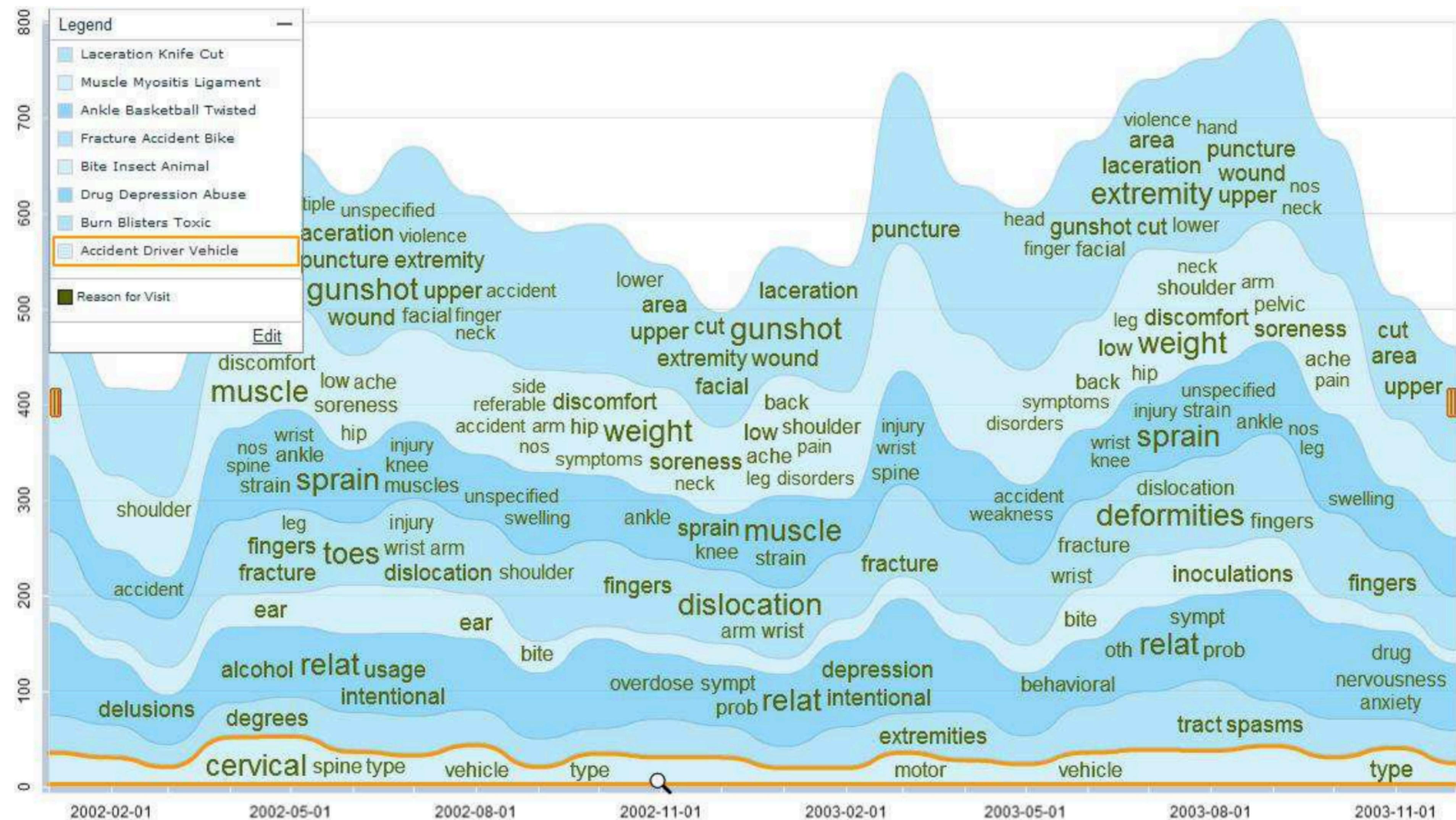
[Viegas]

Document collections

Parallel Tag clouds

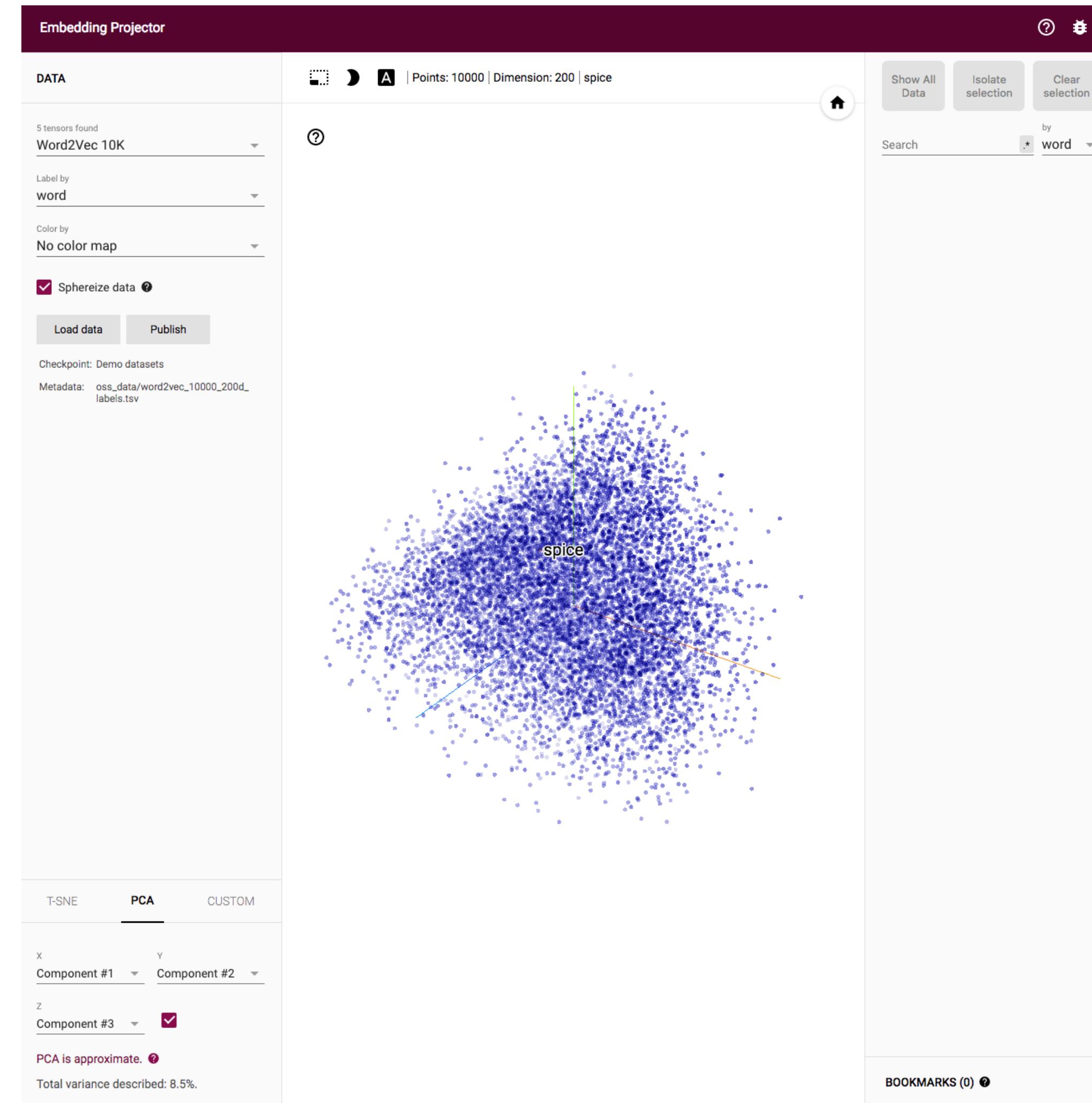


TIARA



[Wei et al.]

Document projections



[Google]

Topic modeling

Assume documents are a mixture of topics

Topics are (roughly) a set of co-occurring terms

Two common techniques:

Latent Semantic Analysis (LSA) or Latent Semantic Indexing (LSI)

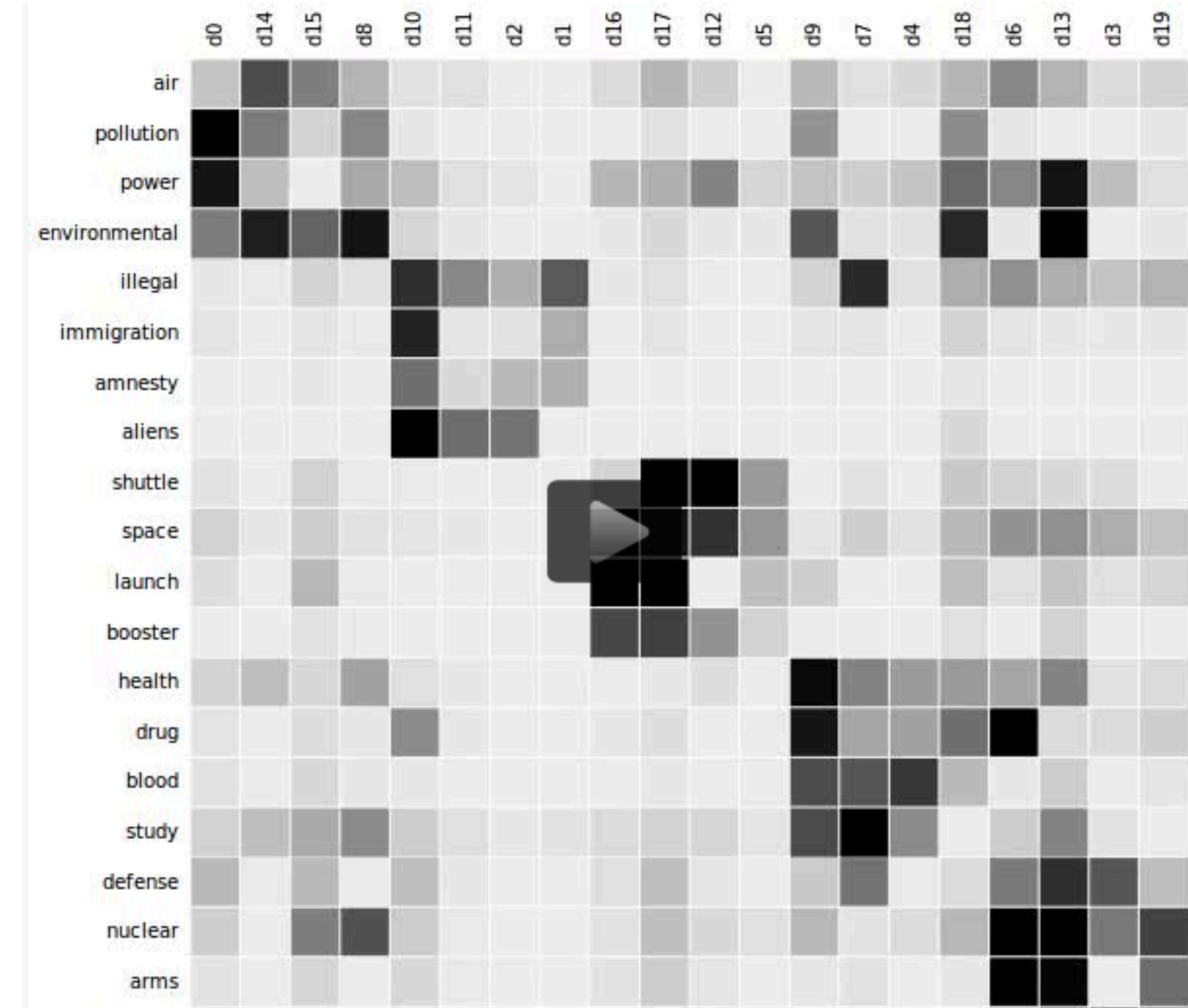
Latent Dirichlet Allocation (LDA)

LSA viz

LSA uses a **term-document matrix** which describes the occurrences of terms in documents

Weight of element proportional to the number of times the terms appear in each document (rare terms are upweighted to reflect their relative importance)

SVD to decompose the term-document matrix A into a term-concept matrix U , a singular value matrix S , and a concept-document matrix V in the form: $A = USV'$.

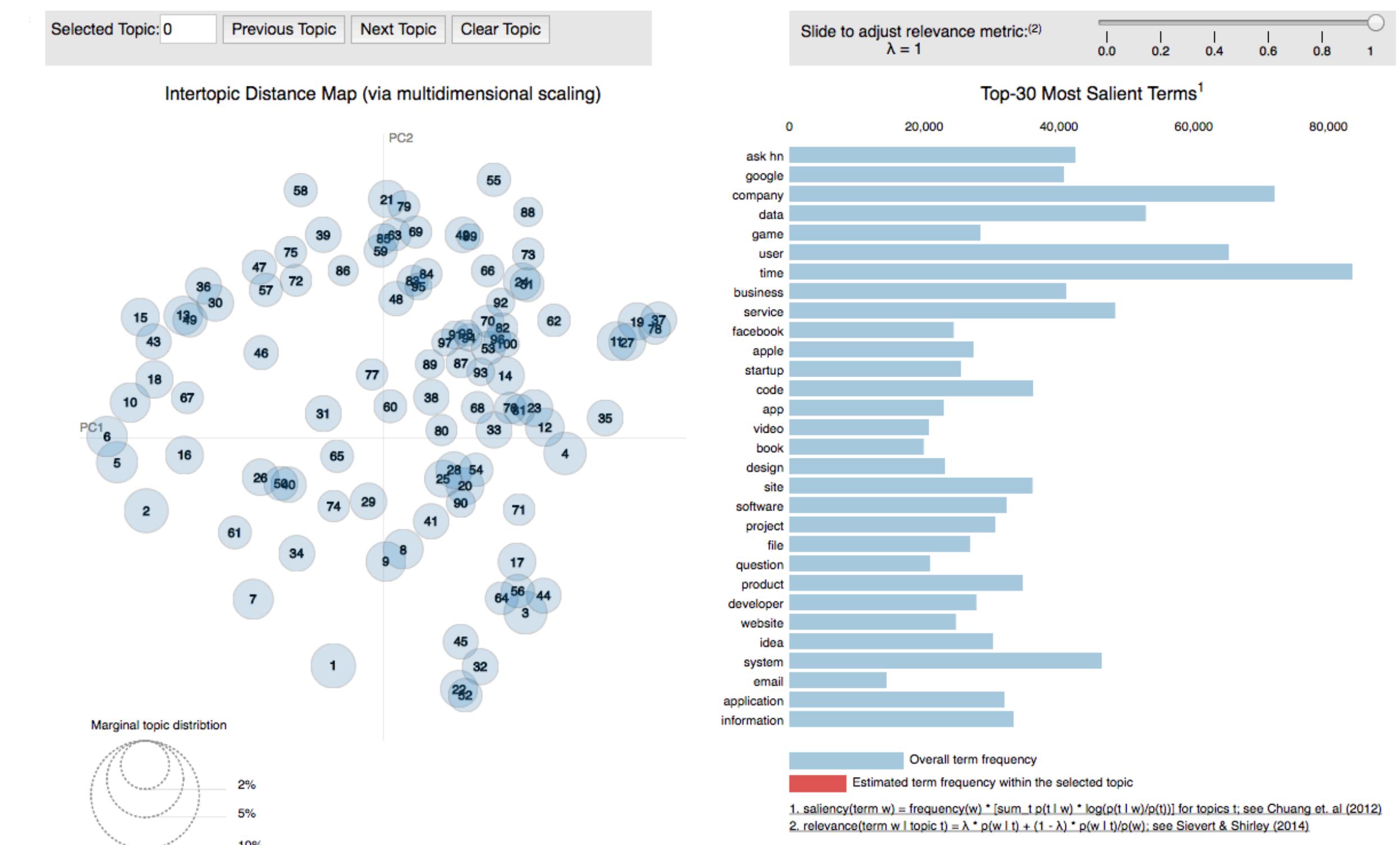


LDAviz

LDA is a generative statistical mode

LDA represents documents as mixtures of topics

LDA model might have topics that can be classified as CAT_related and DOG_related.



More examples

Text Visualization Browser

A Visual Survey of Text Visualization Techniques (IEEE PacificVis 2015 short paper)

Provided by ISOVIS group

[About](#) [Summary](#) [Add entry](#)

