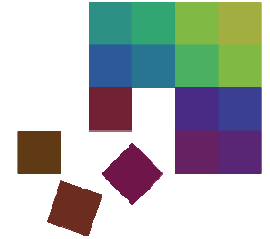


University of Konstanz
Data Analysis and Visualization Group



NLP and Visualization for Digital Humanities II

Guest Lecture: Dr. Christopher Collins
University of Ontario Institute of Technology
<http://vialab.ca>



Outline

- Review
- Tools for Supporting Specific DH Analysis Tasks
- Case Study: Metatation
- Case Study: VisArgue



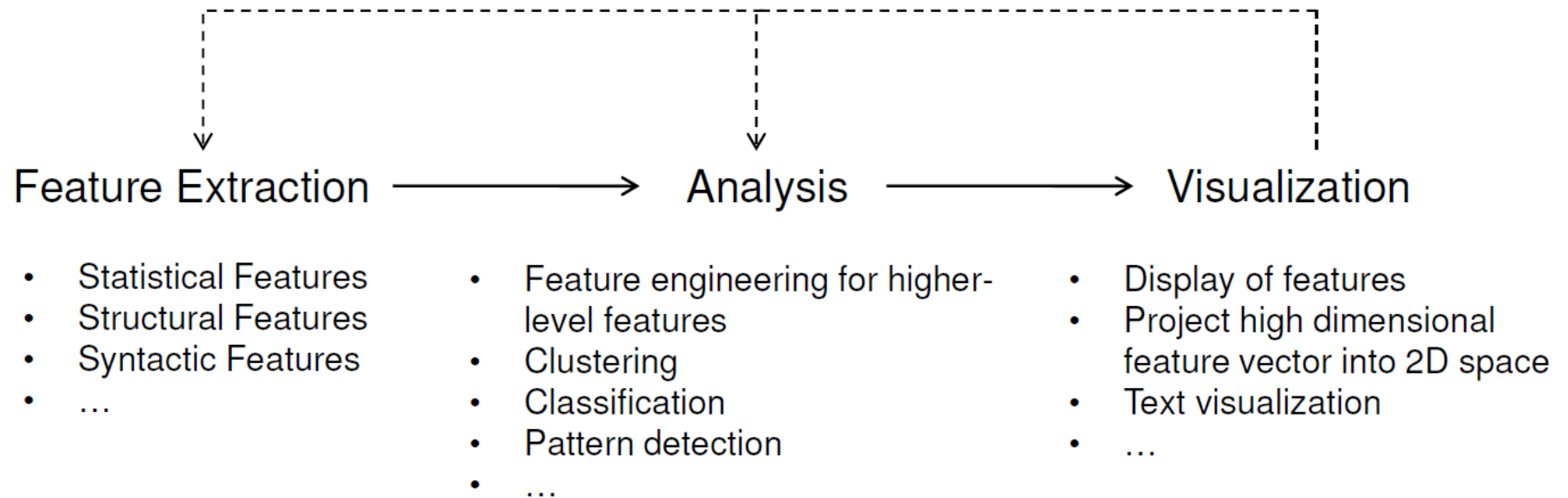
Difficulties Visualizing Text Data


- *Not* pre-attentive
 - Must foveate to read
- Abstract
 - Difficult to visualize
- Very high dimensionality
 - Tens to hundreds of thousands of features
- Compositional
 - Can be combined together in innumerable ways
- Subtle
 - Small differences matter

Why text is (deceptively) easy

- Text is easier when you have a lot of it
 - Web search is now usually conjunction
- Text has a lot of redundancy
 - An algorithm can:
 - Pull out “important” phrases
 - Find “meaningfully” related words
 - Create a “summary” from document
 - Group “related” documents

Stages of Document Visualization



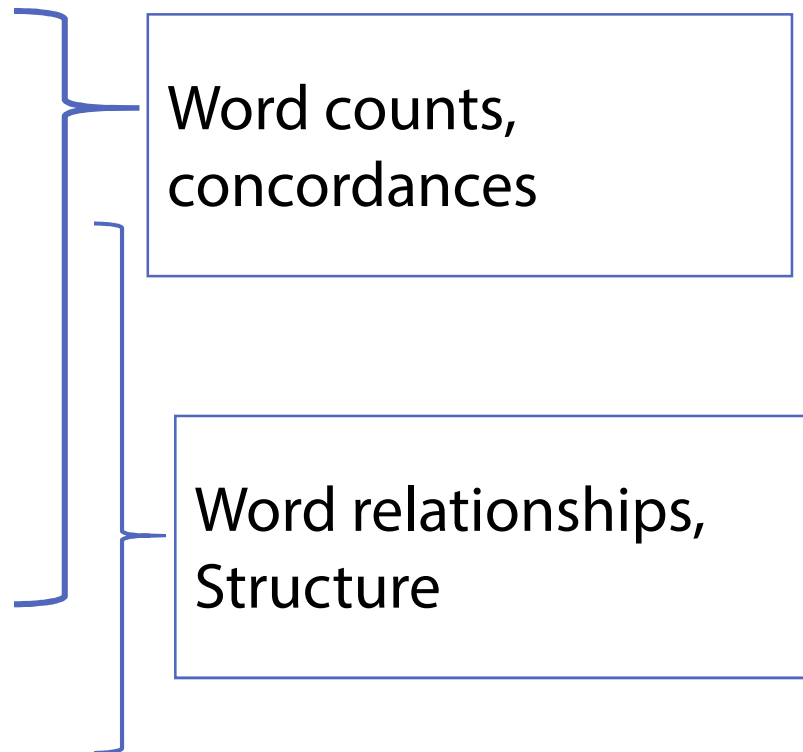


Text Visualization Examples

... with a Digital Humanities Bias

Tools – Lexical Data

- AntConc
- Wordandphrase.info
- Word Wanderer
- Text Arc
- Arc Diagram
- Word Tree
- Phrase Net



Concordance Lines (KWIC)

The screenshot shows the AntConc 3.4.4w (Windows) 2014 interface. The 'Corpus Files' pane on the left lists 'Alice in Wonderland.txt'. The main window displays the 'Concordance Hits' for the search term 'head', with 50 hits found. The concordance is shown in a table with columns for 'Hit', 'KWIC', and 'File'. The search term 'head' is highlighted in blue in the KWIC column. The 'Search Term' is 'head', and the 'Search Window Size' is 50. The 'Kwic Sort' options are checked for Level 1 (1R), Level 2 (2R), and Level 3 (3R). The 'Total No.' of hits is 1, and the 'Files Processed' is 1.

AntConc 3.4.4w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files

Alice in Wonderland.txt

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 50

Hit	KWIC	File
16	her head, she tried to get her head down to them, and was de	Alice in Won
17	VERY nearly at the top of his head. But at any rate he might a	Alice in Won
18	g out, straight at the Footman's head: it just grazed his nose, and	Alice in Won
19	said the Duchess, 'chop off her head!' Alice glanced rather anxi	Alice in Won
20	lbows on it, and talking over its head. 'Very uncomfortable for th	Alice in Won
21	nose. The Dormouse shook its head impatiently, and said, with	Alice in Won
22	on't!' the Hatter said, tossing his head contemptuously. 'I dare sa	Alice in Won
23	ice asked. The Hatter shook his head mournfully. 'Not !' he repl	Alice in Won
24	murdering the time! Off with his head!' 'How dreadfully savage	Alice in Won
25	A bright idea came into Alice's head. 'Is that the reason so man	Alice in Won
26	ot!' said the Queen, tossing her head impatiently; and, turning to	Alice in Won
27	d beast, screamed 'Off with her head! Off--' 'Nonsense!' said Al	Alice in Won
28	ve the hedgehog a blow with its head, it WOULD twist itself roun	Alice in Won
29	ghing: and when she had got its head down, and was going to be	Alice in Won

Search Term ☒ Words ☐ Case ☐ Regex

Search Window Size

head Advanced 50

Start Stop Sort

Kwic Sort

☒ Level 1 1R ☒ Level 2 2R ☒ Level 3 3R

Clone Results

Keyword Plots

Queen



No. of Hits = 75

File Length (in chars) = 163817

Alice



No. of Hits = 403

File Length (in chars) = 163817

Cat



No. of Hits = 37

File Length (in chars) = 163817

N-grams

AntConc 3.4.4w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files

Alice in Wonderland w

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 210

Hit	KWIC	File
1	not get dry very soon. 'Ahem!' said the Mouse with an importa	Alice in Won
2	cia and Northumbria--" 'Ugh!' said the Lory, with a shiver. 'I be	Alice in Won
3	h a shiver. 'I beg your pardon!' said the Mouse, frowning, but ve	Alice in Won
4	olitely: 'Did you speak?' 'Not II' said the Lory hastily. 'I thought	Alice in Won
5	lory hastily. 'I thought you did,' said the Mouse. '--I proceed. "Ed	Alice in Won
6	it advisable--" 'Found WHAT?' said the Duck. 'Found IT,' the M	Alice in Won
7	ell enough, when I find a thing,' said the Duck: 'it's generally a fr	Alice in Won
8	to dry me at all.' 'In that case,' said the Dodo solemnly, rising to	Alice in Won
9	tic remedies--" 'Speak English!' said the Eaglet. 'I don't know the	Alice in Won
10	dibly. 'What I was going to say,' said the Dodo in an offended to	Alice in Won
11	inclined to say anything. 'Why,' said the Dodo, 'the best way to e	Alice in Won
12	es asked. 'Why, SHE, of course,' said the Dodo, pointing to Alice	Alice in Won
13	have a prize herself, you know,' said the Mouse. 'Of course,' the	Alice in Won
14	Alice sadly. 'Hand it over here,' said the Dodo. Then they all cro	Alice in Won

Search Term ☒ Words ☐ Case ☐ Regex

Search Window Size

said the Advanced 50

Start Stop Sort

Kwic Sort

☒ Level 1 1R ☒ Level 2 2R ☒ Level 3 3R

Clone Results

Total No.
1

Files Processed

Collocates

AntConc 3.4.4w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files
Alice in Wonderland w
Through the Looking G

Concordance Concordance Plot File View Clusters/N-Grams **Collocates** Word List Keyword List

Total No. of Collocate Types: 660 **Total No. of Collocate Tokens:** 2600

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	2	0	2	7.64075	furiously
2	2	0	2	6.64075	wringing
3	1	1	0	6.64075	wonders
4	1	1	0	6.64075	warning
5	1	1	0	6.64075	vicious
6	1	1	0	6.64075	tinkling
7	1	1	0	6.64075	thursdays
8	1	0	1	6.64075	stroked
9	1	0	1	6.64075	spied
10	2	1	1	6.64075	shyly
11	1	0	1	6.64075	shrieked
12	1	0	1	6.64075	seizing
13	1	1	0	6.64075	scrubbing
14	1	1	0	6.64075	roared

Search Term ☒ Words ☐ Case ☐ Regex

queen

Start Stop Sort

Sort by ☐ Invert Order
Sort by Stat

Window Span ☐ Same
From... 5L To... 5R

Min. Collocate Frequency
1

Clone Results

Total No.
2
Files Processed

Keywords

AntConc 3.4.4w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files

Alice in Wonderland w
Through the Looking G

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List **Keyword List**

Types Before Cut: 3651 **Types After Cut:** 3312 **Search Hits:** 0

Rank	Freq	Keyness	Keyword
1	853	6930.483	alice
2	935	2982.574	said
3	260	1891.394	queen
4	245	795.826	little
5	171	572.442	went
6	129	541.539	king
7	117	520.425	quite
8	160	491.822	thought
9	81	482.544	tone
10	56	469.858	hatter
11	76	468.107	cried
12	55	461.468	gryphon
13	53	444.687	dumpty
14	53	444.687	humpty

Search Term ☐ Words ☐ Case ☐ Regex **Hit Location** Search Only 0

Start Stop Sort **Reference Corpus** ☒ Loaded **Clone Results**

Sort by ☐ Invert Order Sort by Keyness

Total No. 2
Files Processed

Investigate Words in Context

- External context: From Corpus of Contemporary American English

SYNONYMS (click to see) [?]

bored	
1969	tired
5948	bored
6858	weary
10514	dissatisfied
16929	jaded
overused	
1969	tired
10153	stale
15830	corny
16929	jaded
18836	worn-out
22157	trite
28794	hackneyed
31543	overused
weary	
6858	weary
7551	sleepy
8200	exhausted
10432	shattered
12889	drained
14479	spent
17822	drowsy
18836	worn-out
27093	whacked
27841	fatigued
51734	dog-tired

TIRED_j (RANK 1969, FREQ 19931)

	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
CLICK BAR TO LIMIT					
STORED	32	81	41	39	9
MORE	3403	8959	3702	3106	761

CONCORDANCE LINES

	GENRE				
			Sort	Sort	Sort
1	SPOK	to remember everything that I have to do . It gets	tired	after	a while . I get tired after a
2	MAG	came up here , " she said , " was how	tired	all	the men looked . Little by litt
3	FIC	you . That 's it . That 's why I get	tired	all	the time . " " No , " he said ,
4	MAG	, and I started in on the other . I was	tired	and	cold and clumsy , and befor
5	MAG	led back to sit . Captive in the corridor , growing	tired	and	cramped . Asian realized he
6	MAG	all-day , late-July Ozfest tour date in Boston , Ozzy is	tired	and	cranky . having barely gotte
7	FIC	temporarily . Yvonne agreed to have the baby because she was	tired	and	depressed . and she did n't
8	FIC	and went from the apartment buildings surrounding the vendors ;	tired	and	dirty and coming home fron
9	MAG	and started jogging on the treadmill . Instead of feeling	tired	and	exhausted . surprisingly , I'
10	NEWS	her seat on a segregated bus in Montgomery because she was	tired	and	fed up -- - what if the authc
11	FIC	meat market . Claude paid for four nights . He was	tired	and	had no choice . To get to th
12	ACAD	of his ill-health and pleaded that he was " sick ,	tired	and	headed for death . I ask tha
13	MAG	night before he died , Kile told his brother he was	tired	and	his right shoulder hurt . Une
14	MAG	and his Fool arrive at Regan 's castle . They are	tired	and	hungry from their journey .
15	NEWS	and lay down , I 'll be fine . I 'm	tired	and	I 'm honked off . " # Sexag
16	NEWS	defensive end Chuck Smith . Smith said " my legs are	tired	and	my knees hurt " but said he
17	ACAD	# take simply held his head in his hands and looked	tired	and	was tired then backed up

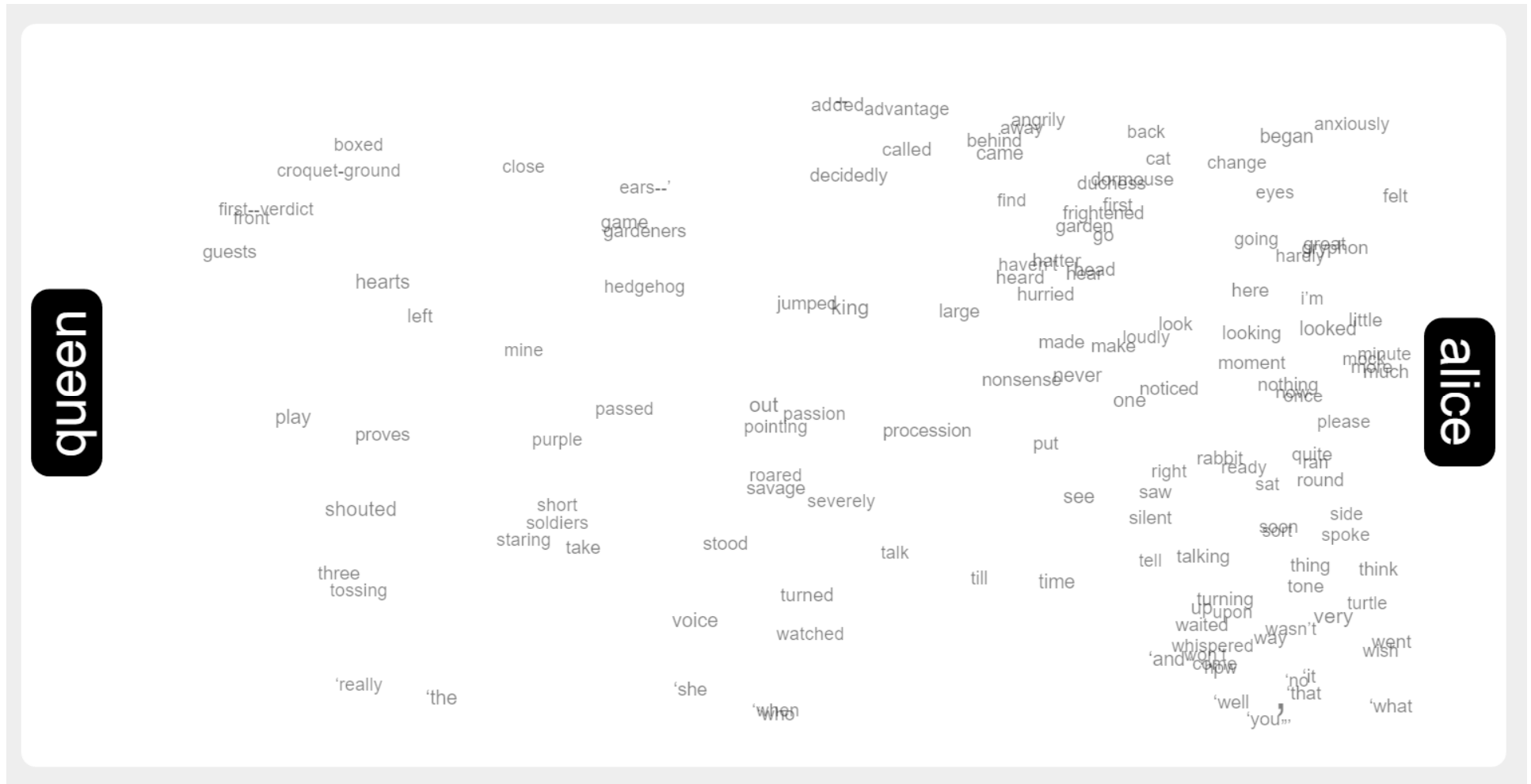
DEFINITIONS (WORDNET)

1. depleted of strength or energy 2. repeated too often

COLLOCATES (click to see with TIRED)

noun eye, voice, face, leg, muscle, smile, excuse, joke **misc** get, so, too, look, feel, sick, very

Word Wanderer



Help
About

Hide to

Download thesaurus

Read

Lewis Carroll

CHAPTER I

Down the Rabbit-Hole

So she was considering in her **swimming** (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

W. Bradford Paley, 2002 , <http://www.textarc.org/>

Alice in Wonderland - fixed & Through the Looking Glass

Search

Comments

Read

↑

28

↓

Of course it is. Its called WABE, you know, because it goes a long way before it, and a long way behind it--

And a long way beyond it on each side, Alice added.

Exactly so. Well, then, MIMSY is flimsy and miserable (theres another portmanteau for you). And a BOROGOVE is a thin shabby-looking bird with its feathers sticking out all round--something like a live mop.

And then MOME RATHS? said Alice. Im afraid Im giving you a great deal of trouble.

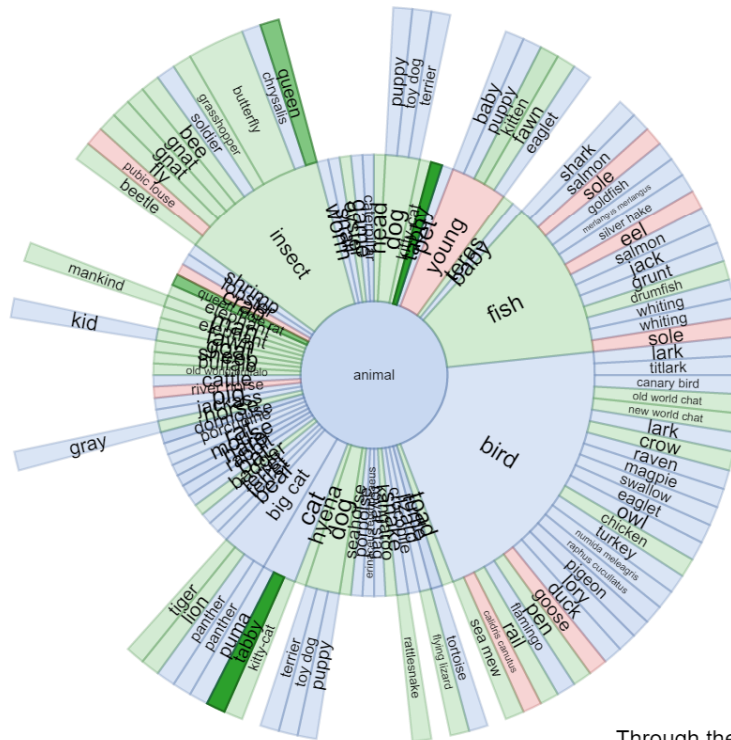
Well, a RATH is a sort of green pig: but MOME Im not certain about. I think its short for from home--meaning that theyd lost their way, you know.

And what does OUTGRABE mean?

Well, OUTGRABING is something between bellowing and whistling, with a kind of sneeze in the middle: however, youll hear it done, maybe--down in the wood yonder--and when youve

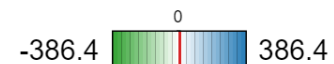
Font Size:

Small

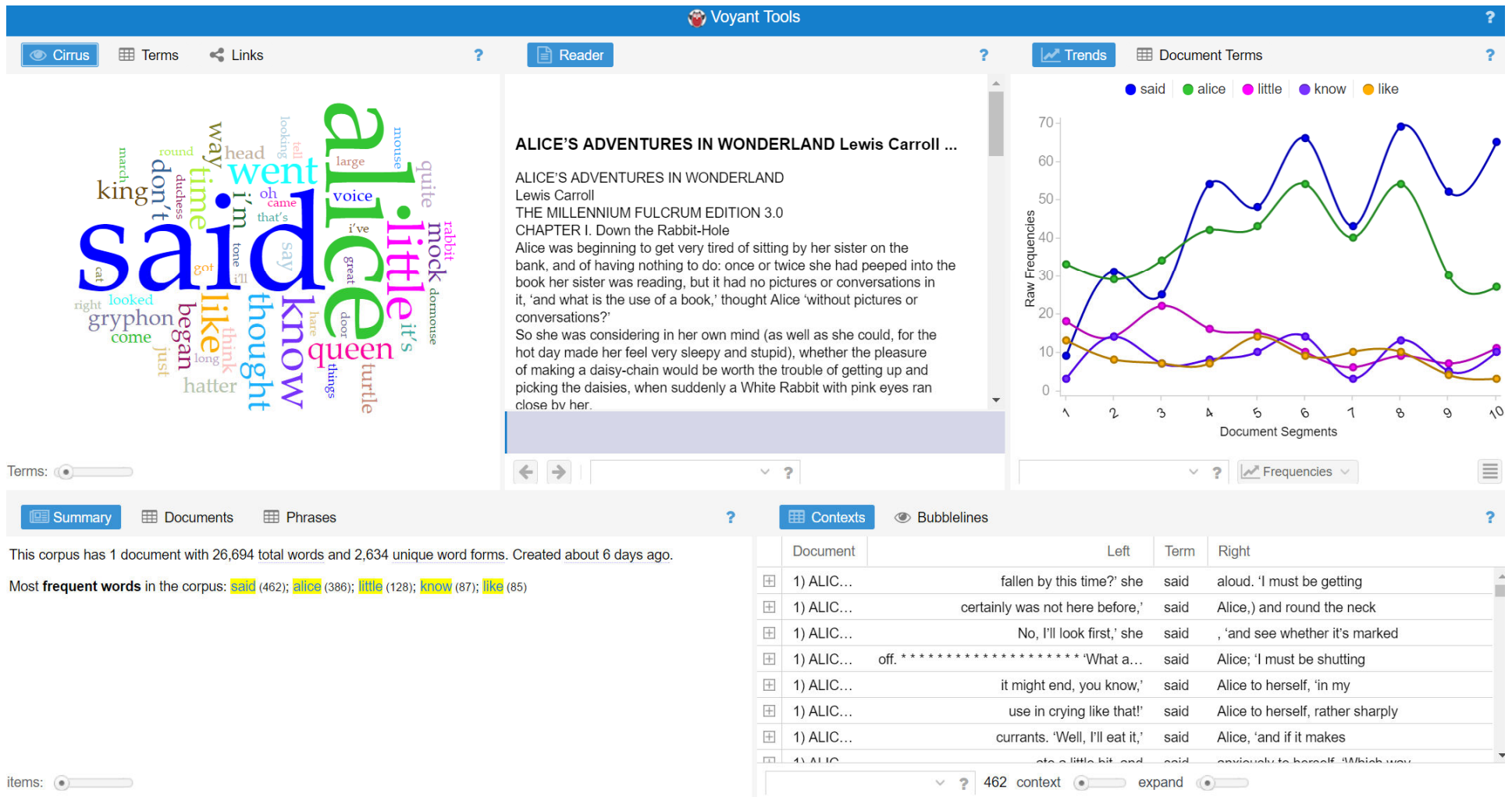


Tweedledee
 the White
 Menai Alice didn
 Dinah hadn paint and me
 KING AND QUEEN
 Queen the Queen head
 patting great Rome UN Red
 London e Cat bank horse
 mind Mary King whole
 Queen Twinkle Like
 Kitty minute Alice
 Alice couldn Waiter
 Second Maine
 Tweedledum

Through the Looking Glass Alice in Wonderland - fixed



Voyant Tools



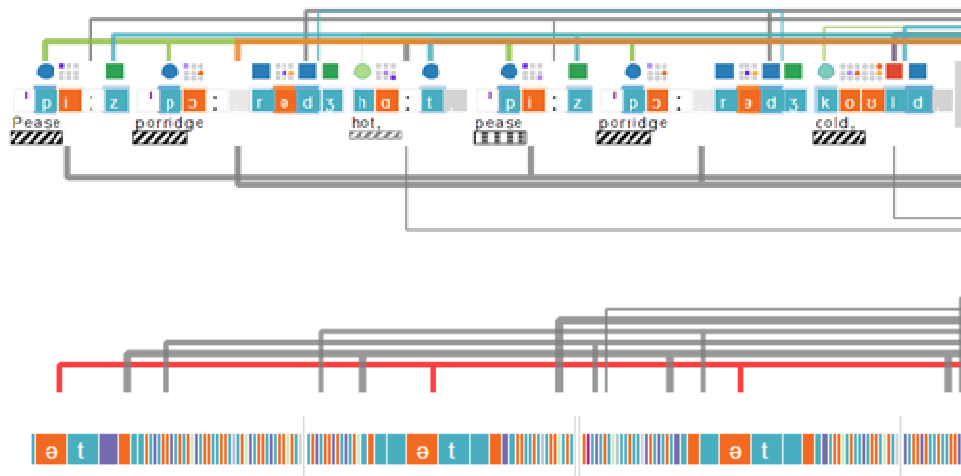


Visualizing Other Document Features

- Sonic Patterns
- Sentence Complexity / Readability
- Sentiment
- Edits and Variations

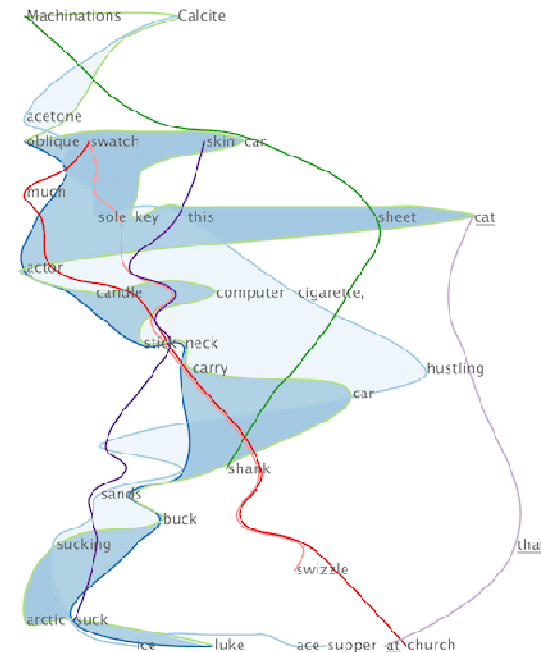
Visualizing Sonic Patterns in Text

Poemviewer



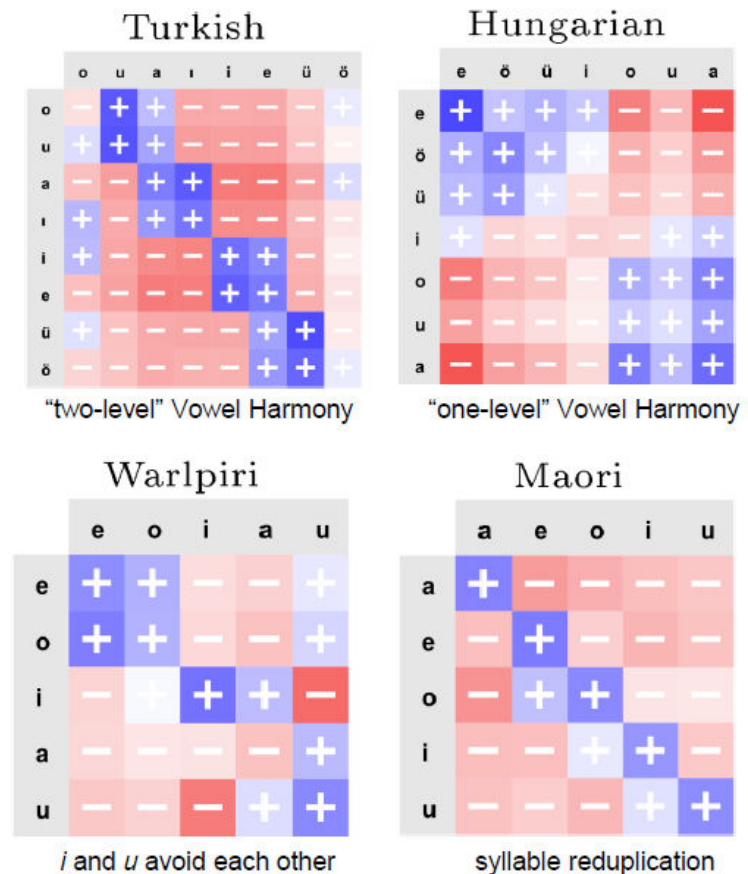
Rule-based Visual Mappings - with a Case Study on Poetry Visualization. A. Abdul-Rahman, J. Lein, K. Coles, E. Maguire, M. Meyer, M. Wynne, C. R. Johnson, A. Trefethen, and M. Chen. In *Computer Graphics Forum*, 32(3):381-390, 2013.

Poemage



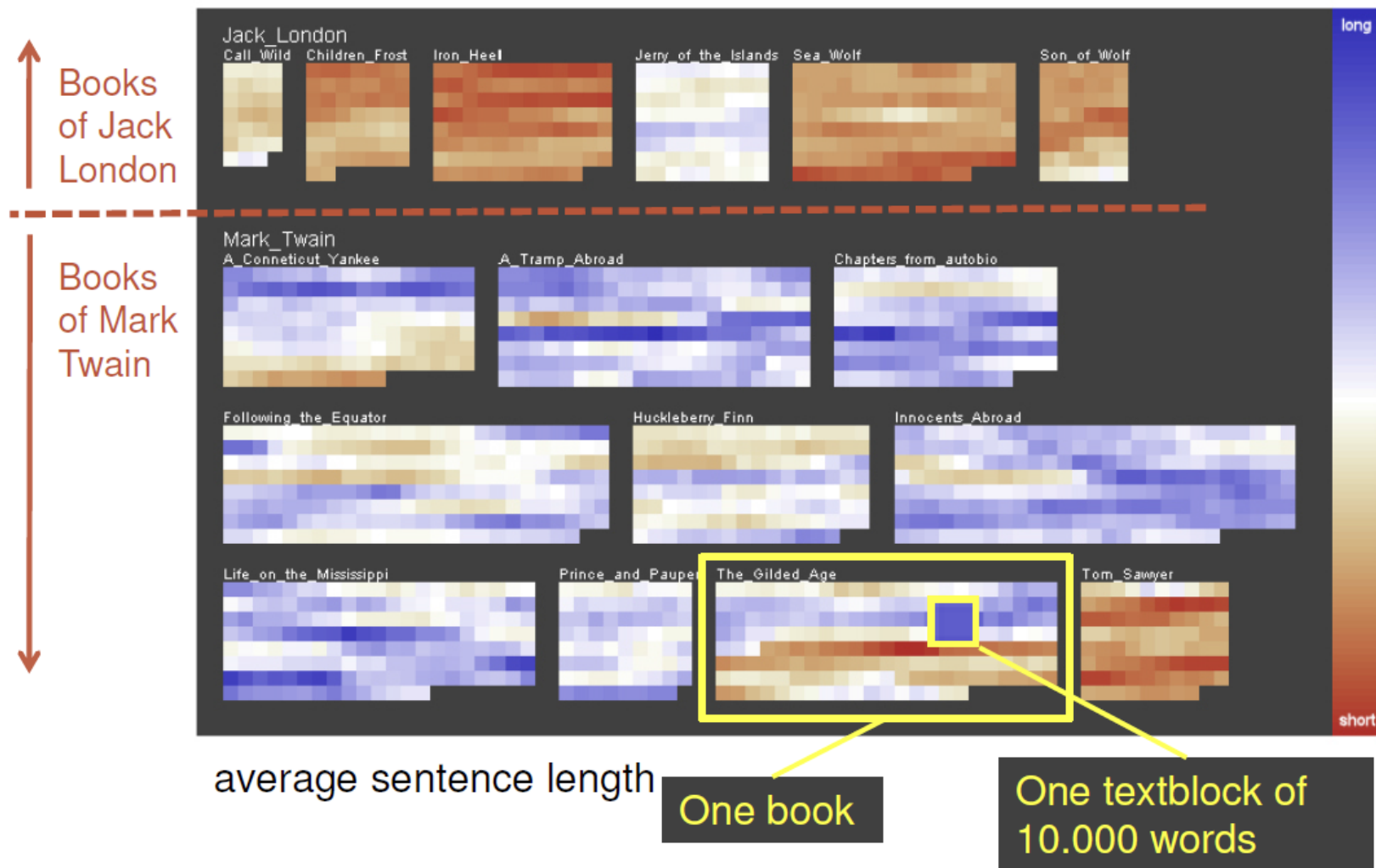
N. McCurdy, J. Lein, K. Coles, M. Meyer. Poemage: Visualizing the Sonic Topology of a Poem. *IEEE TVCG*, pages 439-448, January 2016.

Vowel Harmony: Cross-linguistic Comparison of Complex Language Features



Comparative visual analysis of cross-linguistic features. Proceedings of the International Symposium on Visual Analytics Science and Technology (EuroVAST 2010).

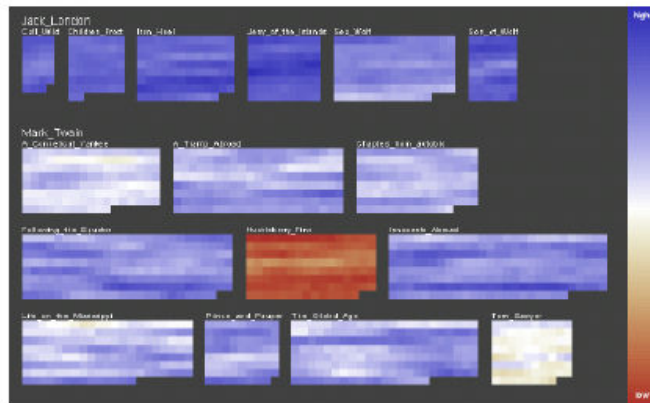
Literature Fingerprinting



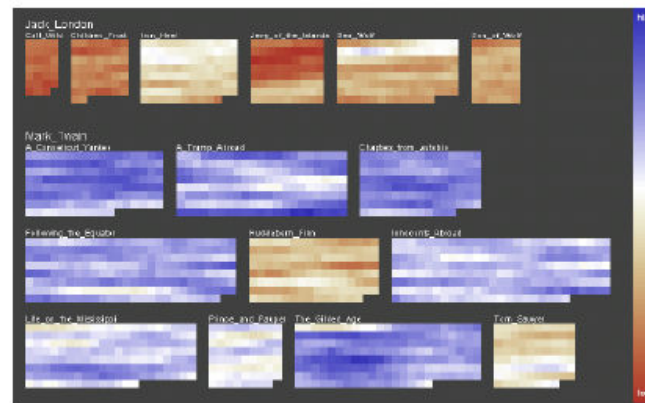
Daniel A. Keim and Daniela Oelke. 2007. Literature Fingerprinting: A New Method for Visual Literary Analysis. In *Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology (VAST '07)*. IEEE Computer Society, Washington, DC, USA, 115-122.

Features

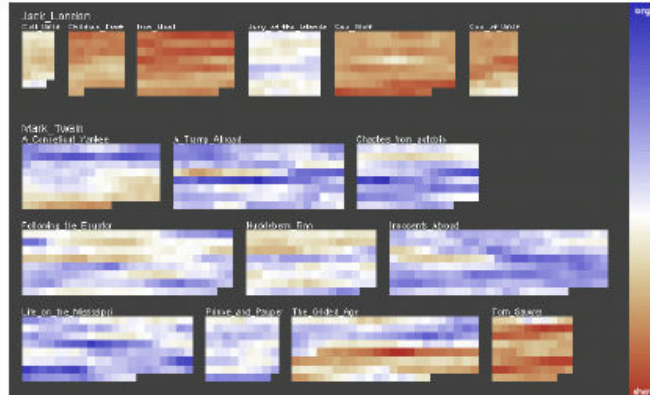
- Hapax legomena: Word's that appear exactly once
- Hapax dislegomena: appear twice
- Function words
- Simpson's index: measure of diversity/repetition
- Sentence length
- Word length
- ...



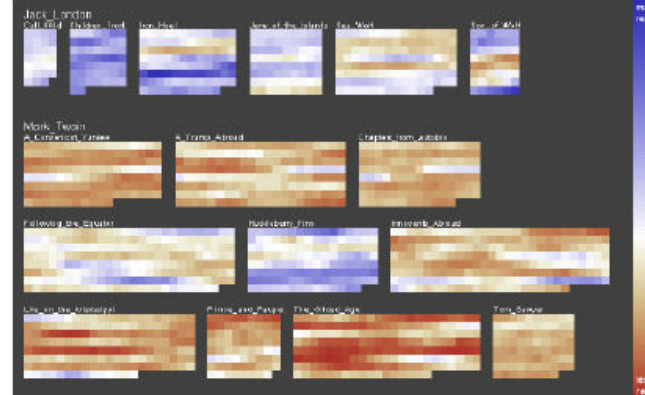
(a) Function words (First Dimension after PCA)



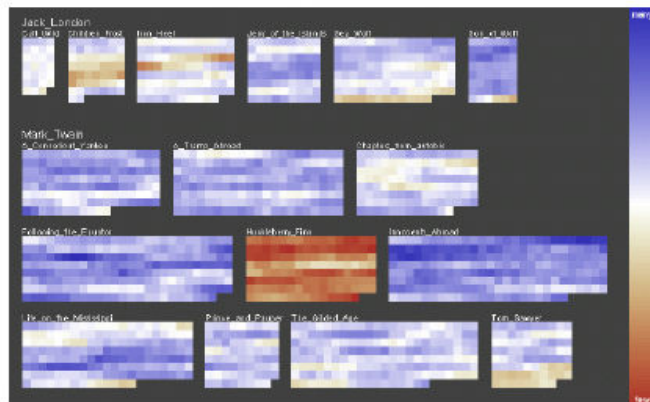
(b) Function words (Second Dimension after PCA)



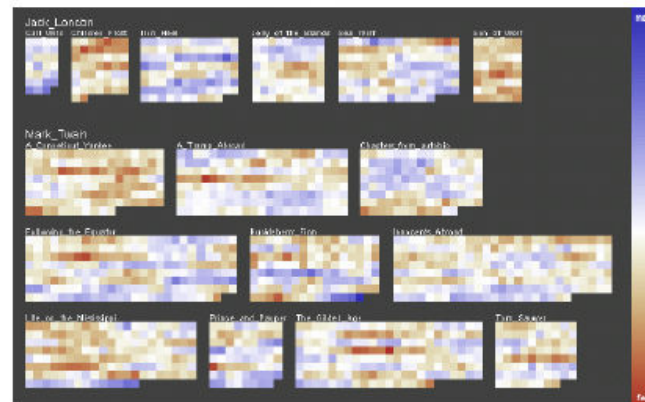
(c) Average sentence length



(d) Simpson's Index



(e) Hapax Legomena



(f) Hapax Dislegomena

Different measures for authorship attribution are tested on books of Mark Twain (last three rows) and Jack London (first row).

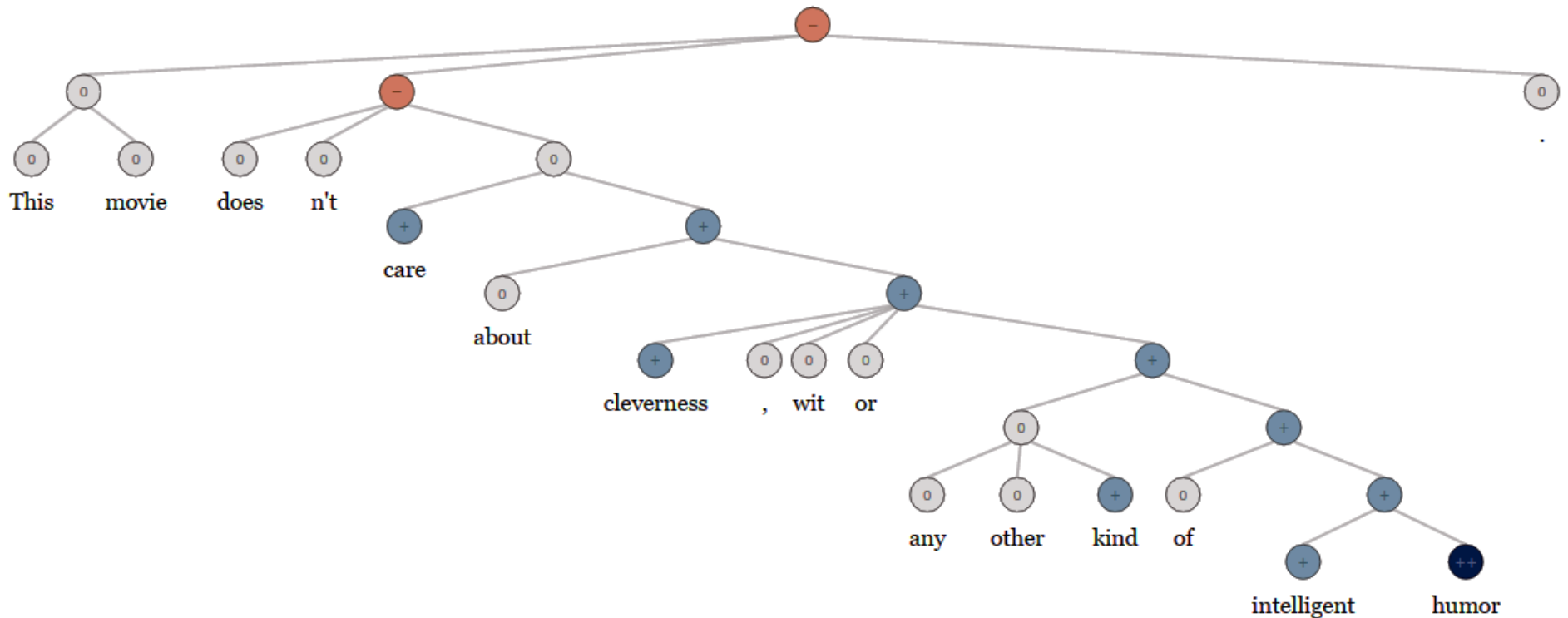
If a feature is able to discriminate between the two authors the books in the first row visually set apart from the rest of the books.



D. Oelke, D. Spretke, A. Stoffel and D. A. Keim.
Visual Readability Analysis: How to Make Your Writings Easier to Read.
IEEE Transactions on Visualization and Computer Graphics, 18(5):662-674, 2012.

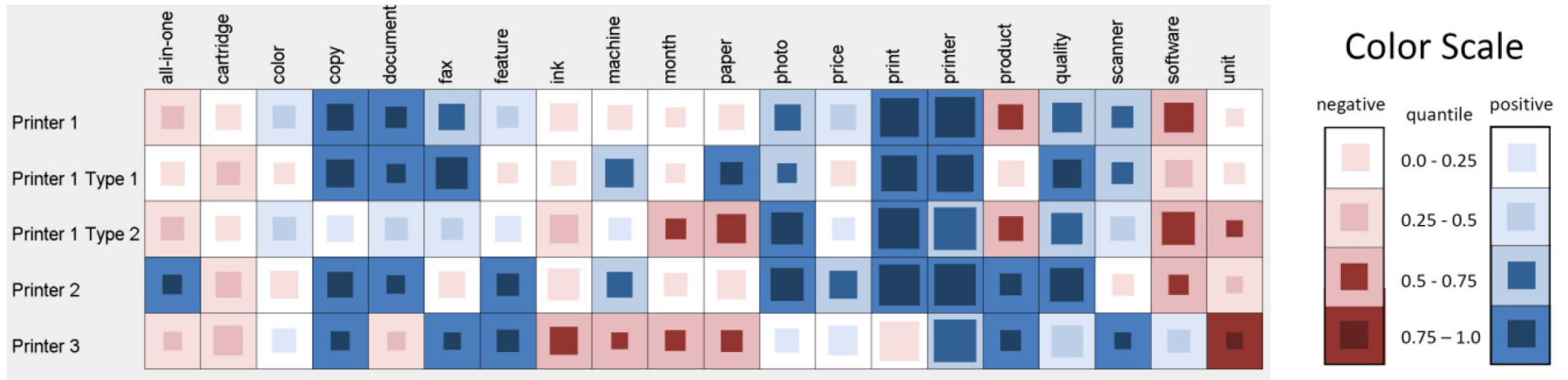
Sentiment Analysis

- Dictionary approaches (emotion word lists)
- Parsing approaches (addresses negation)



<http://nlp.stanford.edu/sentiment/>

Sentiment Analysis

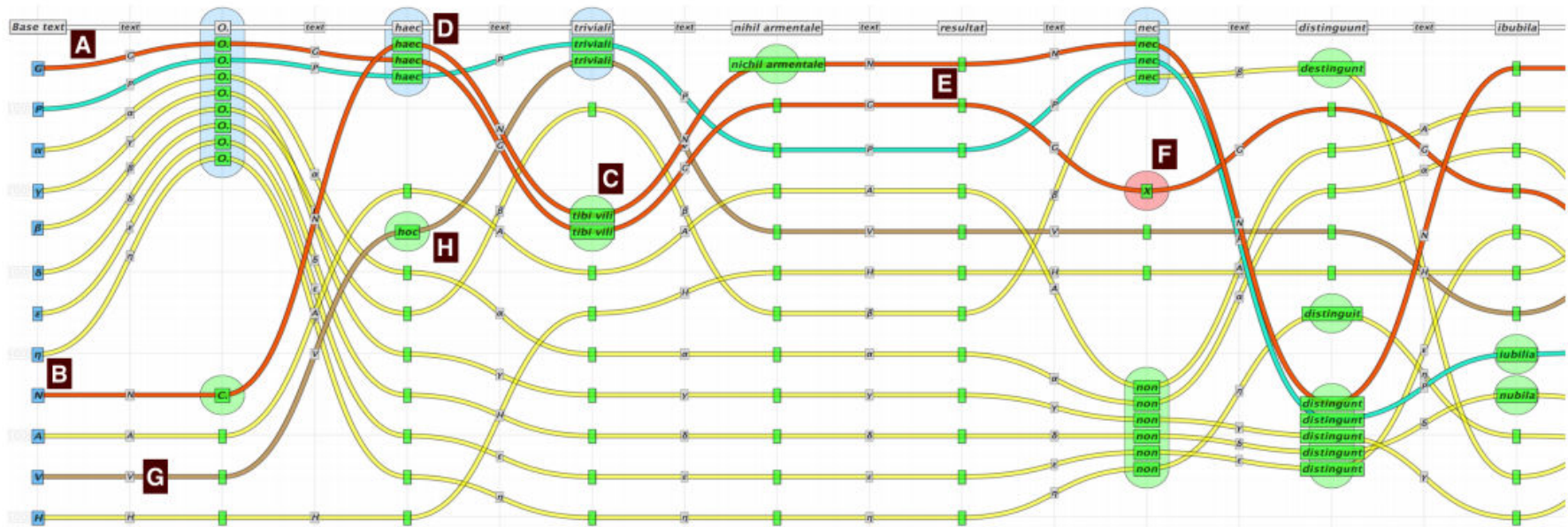


Larger inside square = more comments

D. Oelke, M. C. Hao, C. Rohrdantz, D. A. Keim, U. Dayal, L.-E. Haug and H. Janetzko.
Visual Opinion Analysis of Customer Feedback Data. *Proc. IEEE VAST*, pages 187-194, 2009.

Document Variations

- Variations in *editions* or edits of documents
- Used in historical analysis



Shejuti Silvia, Ronak Etemadpour, June Abbas, Sam Huskey, and Chris Weaver. "Visualizing Variation in Classical Text with Force Directed Storylines". *Proceedings of the Workshop on Visualization for the Digital Humanities*, Baltimore, MD, October 2016.



Tools - Document Collections

- Visualizing Document Collections
 - Citations and Document Connections
 - Clustering
 - Galaxies
 - Themescapes
 - Annotating and Curating
 - Overview Project
 - Timeline Curator
- Visualizing Corpus Trends
 - Culturomics

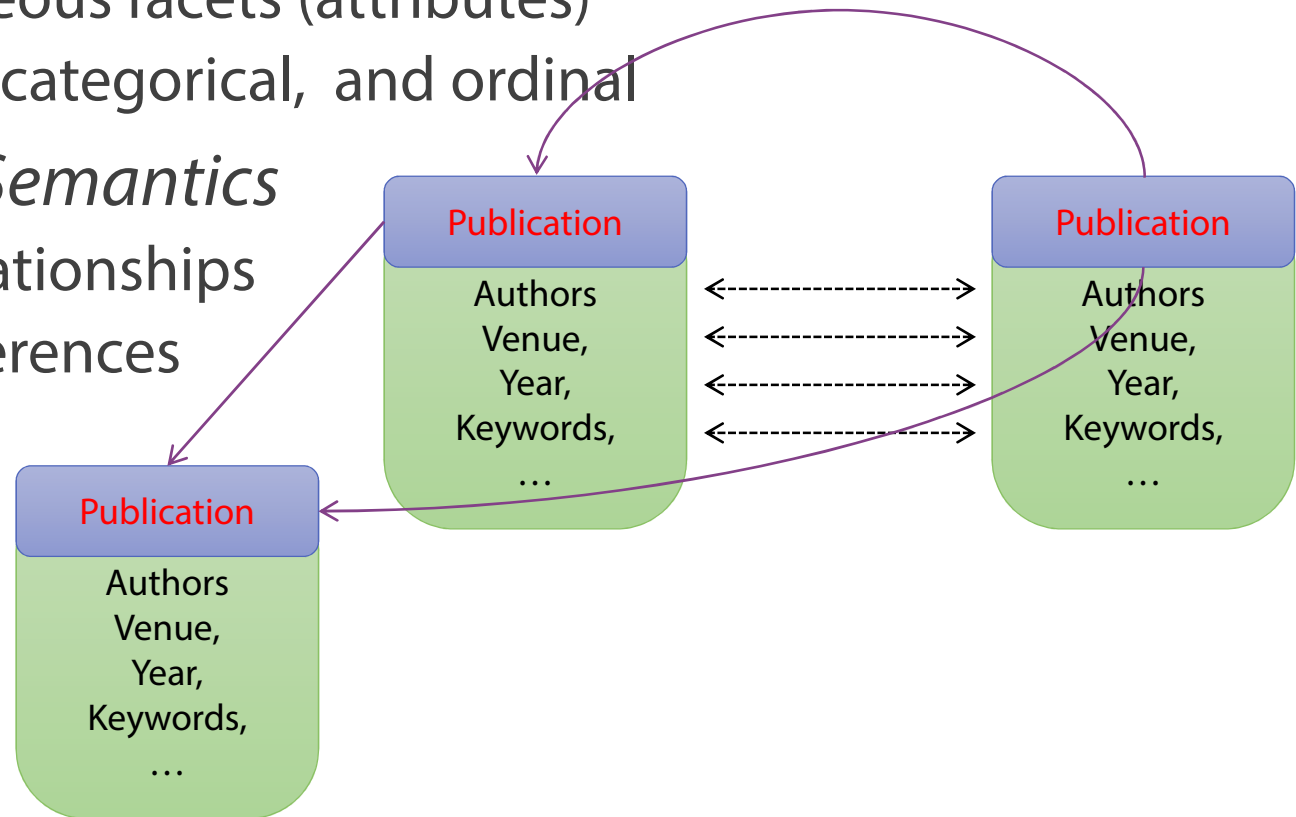
Bibliometric Data

- Large in *Size*
- Rich in *Structure*
 - Heterogeneous facets (attributes)
 - Numerical, categorical, and ordinal
- Complex in *Semantics*



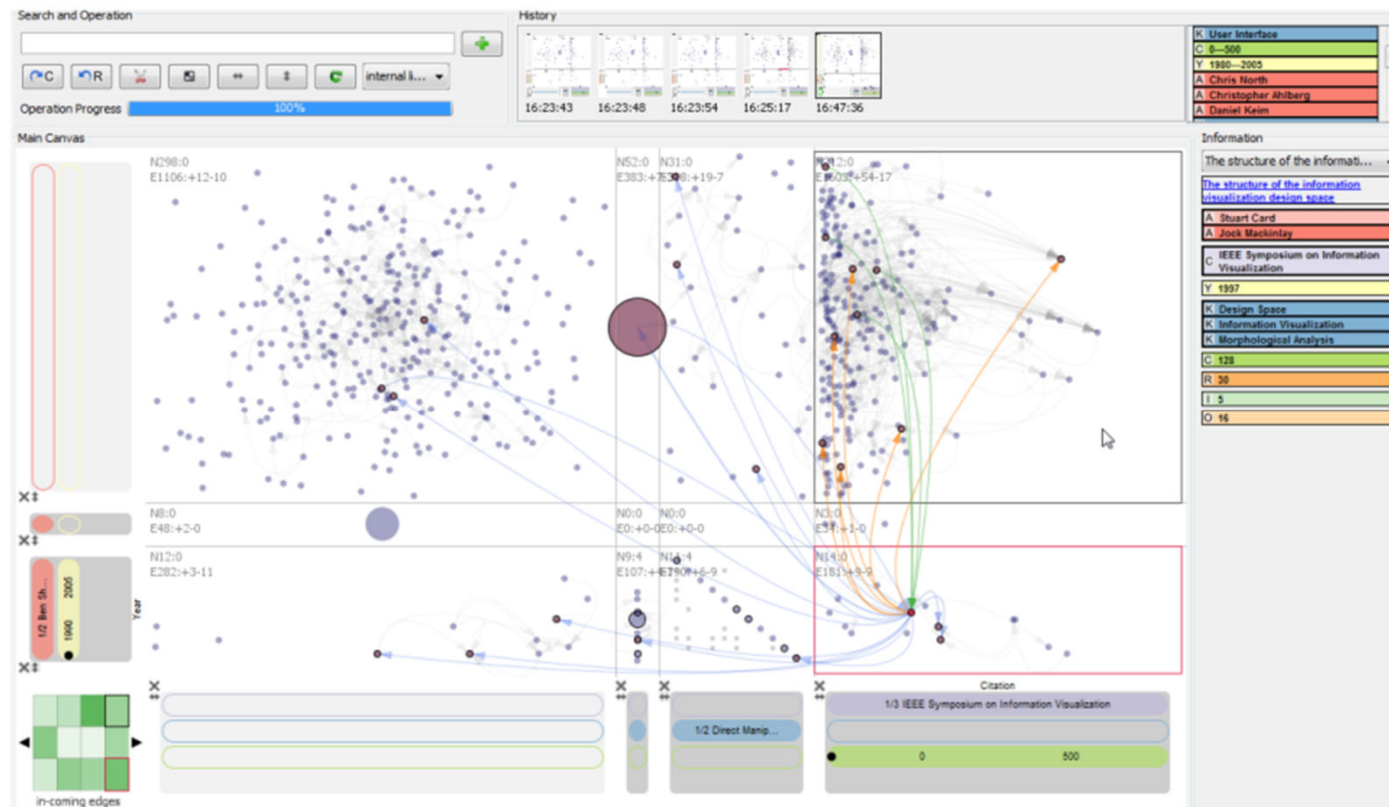
Bibliometric Data

- Large in *Size*
- Rich in *Structure*
 - Heterogeneous facets (attributes)
 - Numerical, categorical, and ordinal
- Complex in *Semantics*
 - Implicit relationships
 - Explicit references



Citation Visualizations

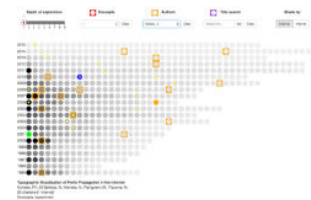
- Citations provide *explicit relations*
- Keywords, authors, etc. provide *implicit relations*



J. Zhao, C. Collins, F. Chevalier, and R. Balakrishnan, "Interactive Exploration of Implicit and Explicit Relations in Faceted Datasets," *IEEE Trans. on Visualization and Computer Graphics (Proc. of the IEEE Conf. on Visual Analytics Science and Technology (VAST))*, vol. 19, iss. 12, pp. 2080-2089, 2013.

Citation Visualizations

- Citation research by John Stasko's group



CiteVis

This visualization presents all the conference papers, their total numbers of citations, and citations between papers. It draws from the original [CiteVis technique](#), but adds the notion of cascading influence through paper generations.

Go!



Author's matrix

This visualization presents citations from specific authors to other authors. It contains only the most referenced authors, and it supports reordering of authors on axes to help explore the data.

Go!



Author's statistics

This webpage shows statistics for specific authors including their most frequent co-authors, the most frequent concepts of their papers, and sets of other authors that cite all of authors in each set (cliques).

Go!



Co-citations network

This visualization presents a node-link network diagram, with nodes representing authors and edges between authors if they cite each other at least once.

Go!

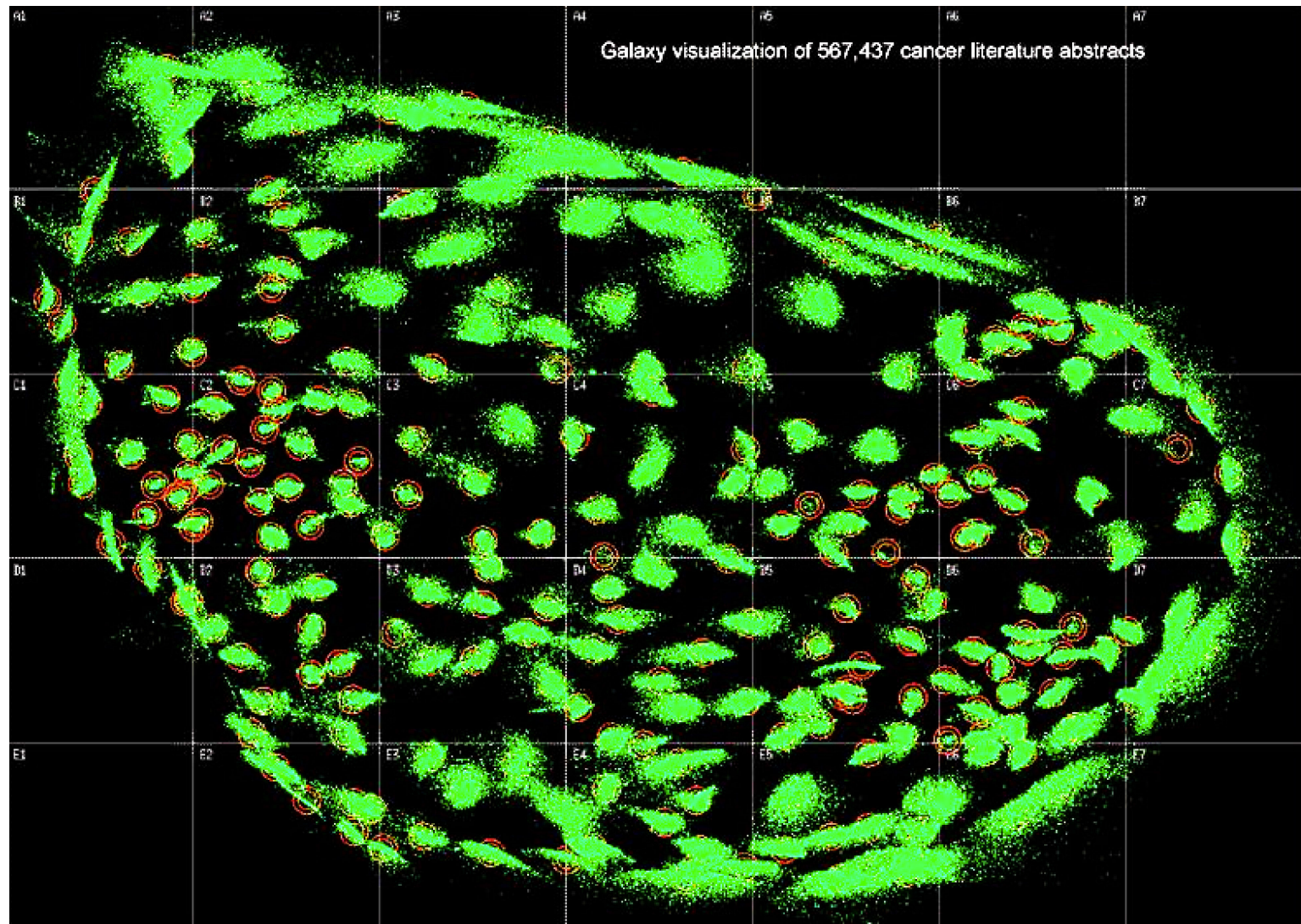


Concepts

This visualization presents common concepts (keywords) across all the titles and abstracts of papers, and how frequently each concept co-occurs with other concepts.

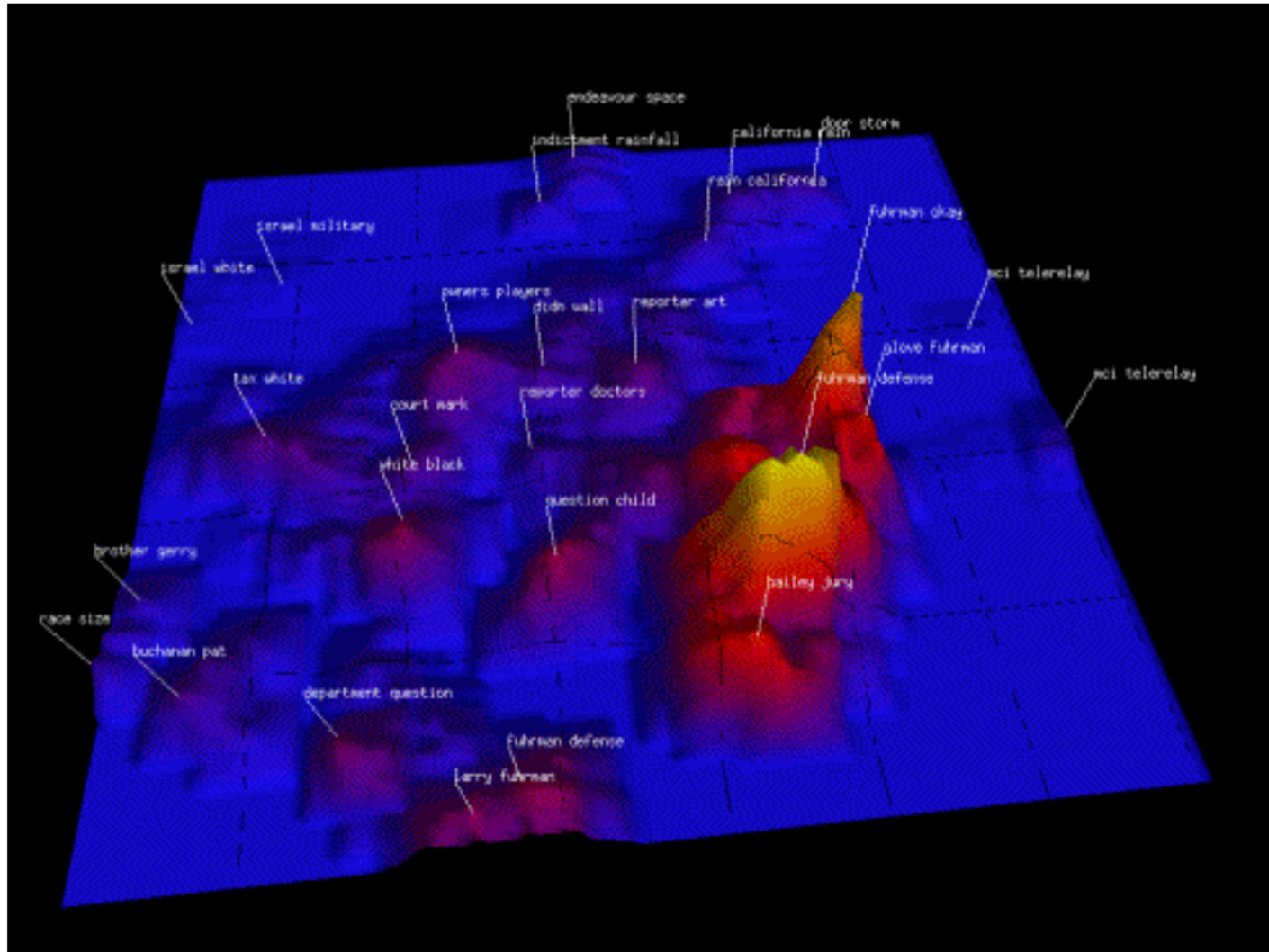
Go!

Clustering - Galaxies



James A. Wise et al. 1995. Visualizing the non-visual: spatial analysis and interaction with information for text documents. In *Proc. of the IEEE Symp. on Information Visualization*, pages 51–58.

Clustering - Themescape



James A. Wise et al. 1995. Visualizing the non-visual: spatial analysis and interaction with information for text documents. In *Proc. of the IEEE Symp. on Information Visualization*, pages 51–58.

Overview Project - Entities

Search for only these:			Entity	count	docs
<input type="checkbox"/>	Companies by Suffix		huma	487	352
			Huma Xian		
<input type="checkbox"/>	Geonames: Cities		israel	246	64
			State of Israel		
<input checked="" type="checkbox"/>	Geonames: Countries	Any country or disputed country, plus variations and translations, from geonames.org	libya	245	93
			Libya		
<input checked="" type="checkbox"/>	Geonames: Political Boundaries	About 300,000 countries and administrative regions worldwide, plus variations and translations, from geonames.org	syria	163	39
			Syrian Arab Republic		
<input type="checkbox"/>	Numbers		tripoli	145	47
			Tripoli		
... then remove any of these:			israeli	144	53
			State of Israel		
<input checked="" type="checkbox"/>	English: Google Books words	The most common 50,000 uncapitalized words in English books, according to Google Books (CC BY 3.0)	united states	119	75
			United States		
<input checked="" type="checkbox"/>	English: stop words	Extremely common English words	obama	104	47
			Obama-machi		
<input checked="" type="checkbox"/>	Numbers		washington	100	58
			Washington County		

Overview Project

- “Multisearch” is customized to the workflow of journalists – brainstorming.

Searching all documents

×

[16,000 tweets about drones \(16,005\) ⓘ](#) [Word Cloud ⓘ](#) [Xpermental: Word Co-occurrence ⓘ](#) [Entities ⓘ](#) [Multisearch ⓘ](#)

[Add view ▾](#)

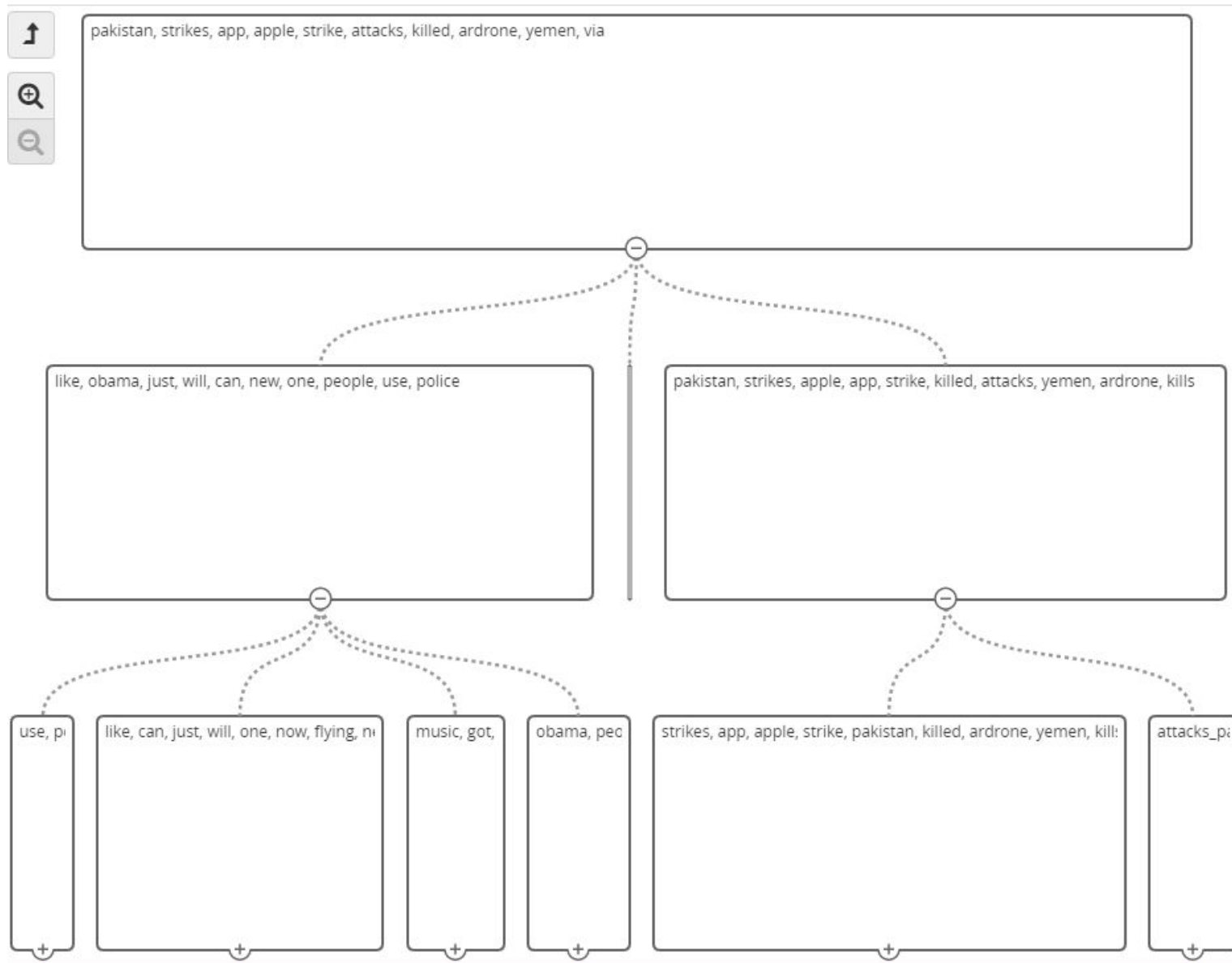
Searches

↑ Query	Count	
isis	0	
iraq	37	
pakistan	1,981	
africa	80	
canada	8	
toronto	0	⌵ Set as filter ✎ Edit 🗑 Delete

[Cut and paste searches](#)

Overview – Document Clustering

Add view ▾



Timeline Curator

TimeLineCurator v0.2

✕

✕

Insert your text here

Title

1 2 3 4 5 6

Content

Since the days of the fur trade, British Columbia's economy has been based on natural resources, particularly fishing, logging and mining. From the canneries to the mills and mines, B.C.'s resource sector was increasingly the domain of large commercial interests.

With industrialization and economic growth, workers arrived to join in the seemingly boundless prosperity. Increasingly, these workers came from Asia as well as Europe. The mix of cultures and diversity was a source of strength, but also, often, of conflict. The early part of the 20th century was a time of great change and talk between immigrants and the First Nations, all of whom found their lives changing rapidly.

Rise of the labour movement[edit]

The dominance of the economy by big business was accompanied by an often militant labour movement. The first major sympathy strike was in 1903 when railway employees struck against the CPR for union recognition. Labour leader Frank Rogers was killed while picketing at the docks by CPR police during that strike, becoming the British Columbia movement's first martyr.[29] Canada's first general strike occurred following the death of another labour leader, Ginger Goodwin, in 1918, at the Cumberland coal mines on Vancouver Island.[30] A lull in industrial tensions through the later 1920s came to an abrupt end with the Great Depression. Most of the 1930s strikes were led by Communist Party organizers.[31] That strike wave peaked in 1935 when unemployed men flooded the city to protest conditions in the relief camps run by the military in remote areas throughout the province. After two tense months of daily and disruptive protesting, the relief camp strikers decided to take their grievances to the federal government and embarked on the On-to-Ottawa Trek.[32] but their commandeered train was met by a gatling gun at Hatzic, just east of Mission City, and the strikers arrested and interned in work camps for the duration of the Depression.[33]

Race and ethnic relations[edit]

The Komagata Maru and HMCS Rainbow

During the 20th century, many immigrant groups arrived in British Columbia and today, Vancouver is the second most ethnically diverse city in Canada, only behind Toronto. In 1886, a head tax was imposed on the Chinese, which reached as much as \$500 per person to enter Canada by 1904. By 1923 the government passed the Chinese Immigration Act, which prohibited all Chinese immigration until 1947. Sikhs had to face an amended Immigration Act in 1908 that required Sikhs to have \$200 on arrival in Canada, and immigration would be allowed only if the passenger had arrived by continuous journey from India, which was impossible. Perhaps the most famous incident of anti-Sikh racism in B.C. was in 1914 when the Komagata Maru arrived in Vancouver harbour with 376 Sikhs aboard, of whom only 20 were allowed entry. The Komagata Maru spent two months in harbour while the Khalsa Society went through the courts to appeal their case. The Khalsa Society also kept the passengers on the Komagata Maru alive during those two months. When the case was lost, HMCS Rainbow, a Royal Canadian Navy cruiser, escorted the Komagata Maru out to sea while thousands of Caucasians cheered from the seawall of Stanley Park.

During the Second World War, security concerns following the bombing of Pearl Harbor and Canada's entry into the war versus Japan led to controversial measures. The local Japanese-Canadian population was openly discriminated against, being put in internment camps. The Pacific Coast Militia Rangers were formed in 1942 in order to provide an armed presence on the coast in addition to the pre-war fortress garrisons, which were expanded after hostilities. Japanese military attacks against BC amounted to a small number of parachute bombs released from great distance away and by the middle of 1942 the threat of direct attack diminished following defeat at the Battle of Midway by US forces.

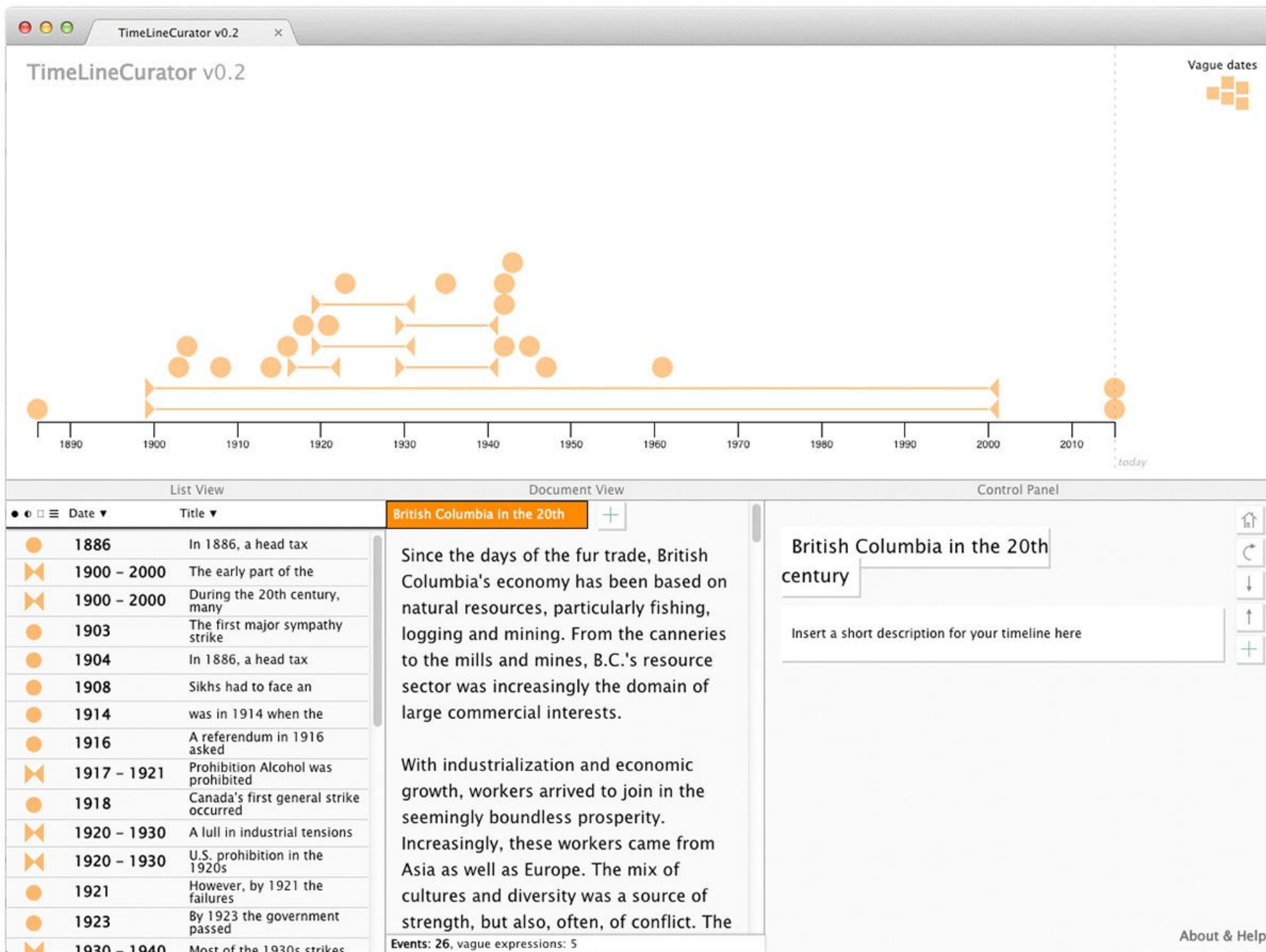
Prohibition[edit]

Which date does "today" refer to in the document (optional):

Go!

Or choose from an example file

About & Help



TimeLineCurator v0.2

TimeLineCurator v0.2

Vague dates

List View

Document View

Control Panel

Date ▼	Title ▼
1935	That strike wave peaked in
1942	The Pacific Coast Militia Rangers
1942	Japanese military attacks against BC
1942	World War II contributions
1943	The Rocky Mountain Rangers sent
1945	World War II contributions
1947	By 1923 the government passed
1961	Columbia River Treaty In
19 Feb 2015	However, by 1921 the failures
19 Feb 2015	During the 20th century, many
???	Prohibition Alcohol was prohibited
???	After two tense months of
???	The Khalsa Society also kept
???	The Komagata Maru spent two
???	After two tense months of

British Columbia In the 20th

Keefe

Manohy, were awarded the Victoria Cross for actions with BC-based regiments in Italy.

Columbia River Treaty[edit]

In 1961, British Columbia ratified the Columbia River Treaty which required the building of three large dams in British Columbia in return for financial compensation related to U.S. hydroelectric power production enabled by the dams. The dams flooded large areas within British Columbia, but would prove to be a very stable and renewable source of power for the province.

Events: 26, vague expressions: 5

Columbia River Treaty

Content:

British Columbia ratified the Columbia River Treaty which required the building of three large dams in British Columbia in return for financial compensation related to U.S. hydroelectric power production enabled by the dams.

Add Media

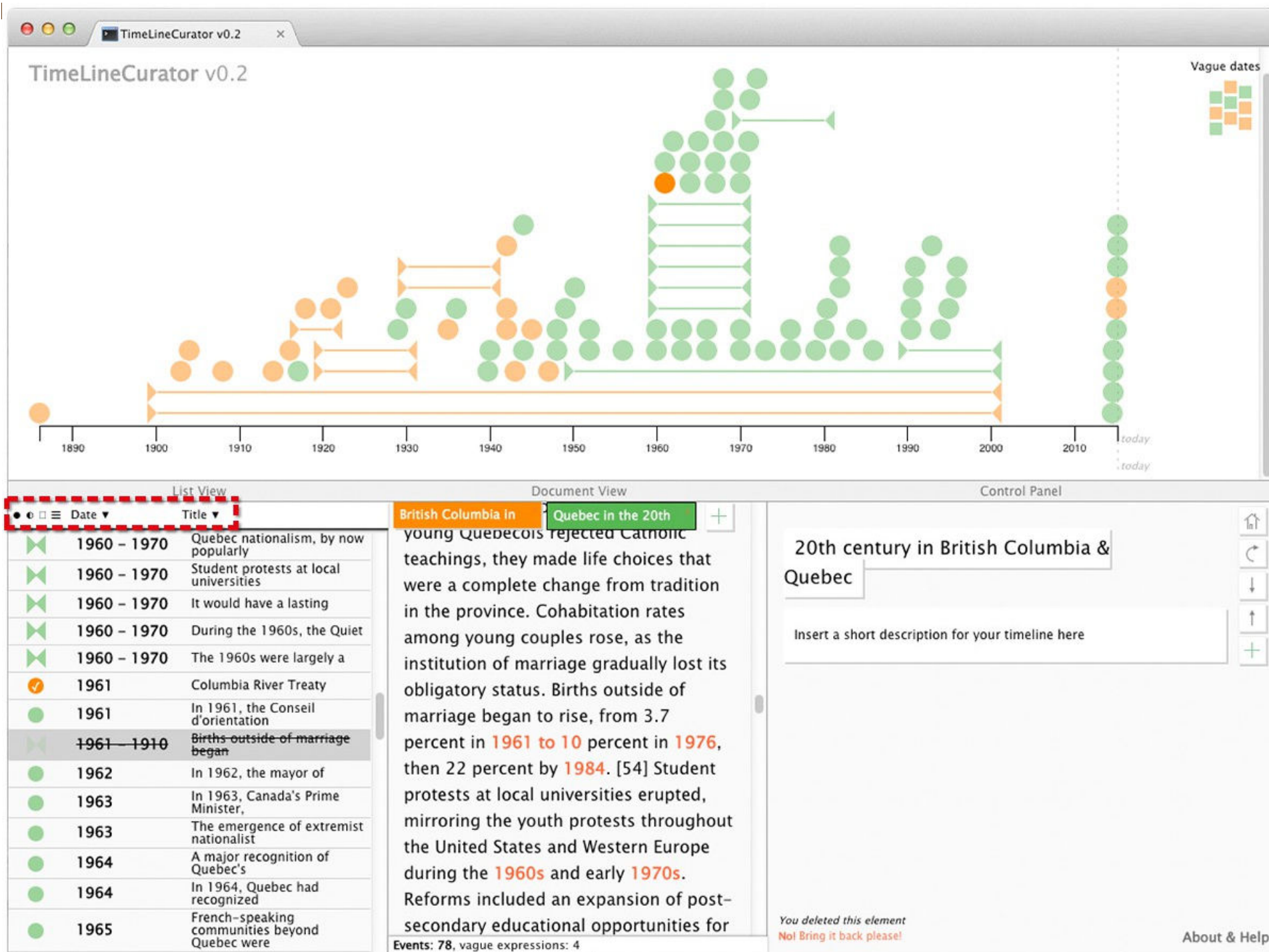
<https://upload.wikimedia.c>

Wikipedia

Columbia River Treaty

Track: 1 2 3 4 5 6

About & Help



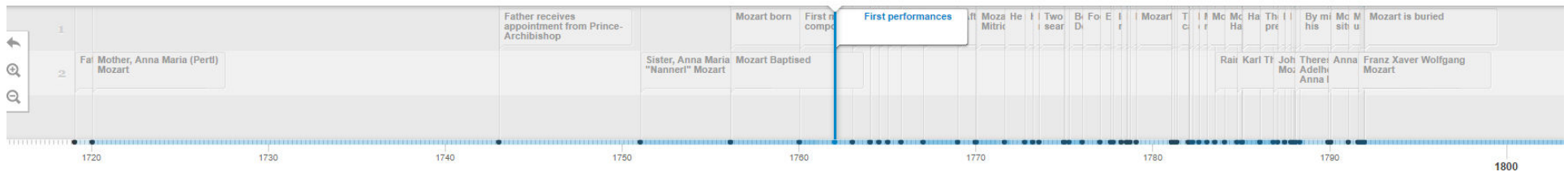


Mozart painted by Greuze, 1764

1762

First performances

These began with an exhibition, in 1762, at the court of the Prince-elector Maximilian III of Bavaria in Munich, and at the Imperial Court in Vienna and Prague.



Topic Modelling?!

“In machine learning and natural language processing, a topic model is a type of statistical model for **discovering the abstract "topics"** that occur in a **collection of documents.**”

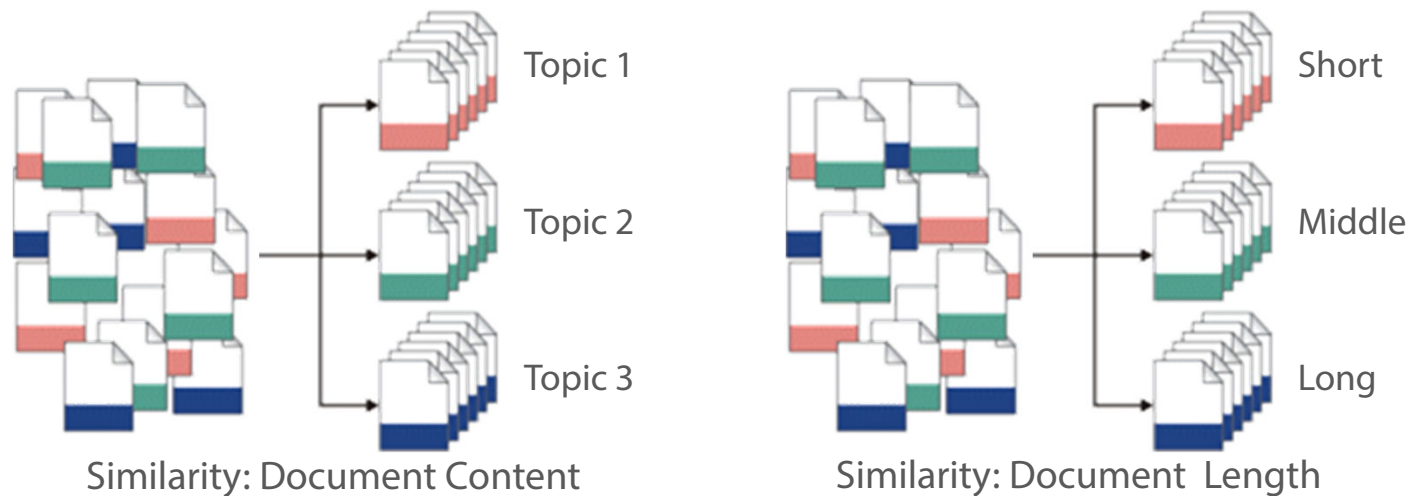
- Wikipedia

Topic Modelling = Clustering of documents into topics

- | | | |
|-----------------|------------|------------------|
| • Hierarchical | vs. | Non-hierarchical |
| • Partial | vs. | Agglomerative |
| • Deterministic | vs. | Probabilistic |
| • Incremental | vs. | Batch |
| • Supervised | vs. | Unsupervised |

Text Clustering $\not\equiv$ Topic Modeling

Clustering = finding groups of *similar objects* in the data



Topic Modeling

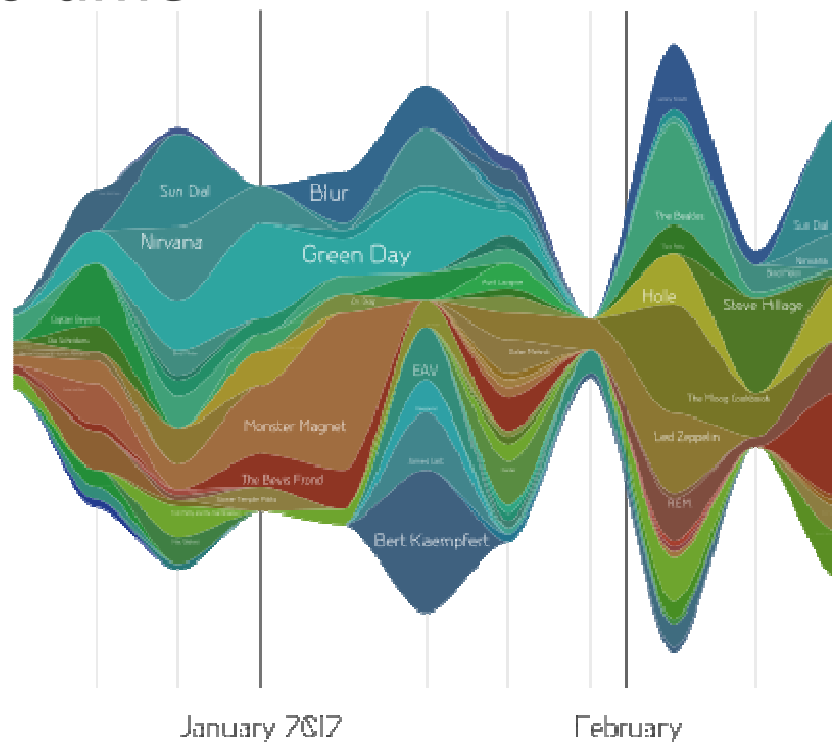
= discovering the abstract topics that occur in a documents collection

= clustering documents (with respect to their content) into topics



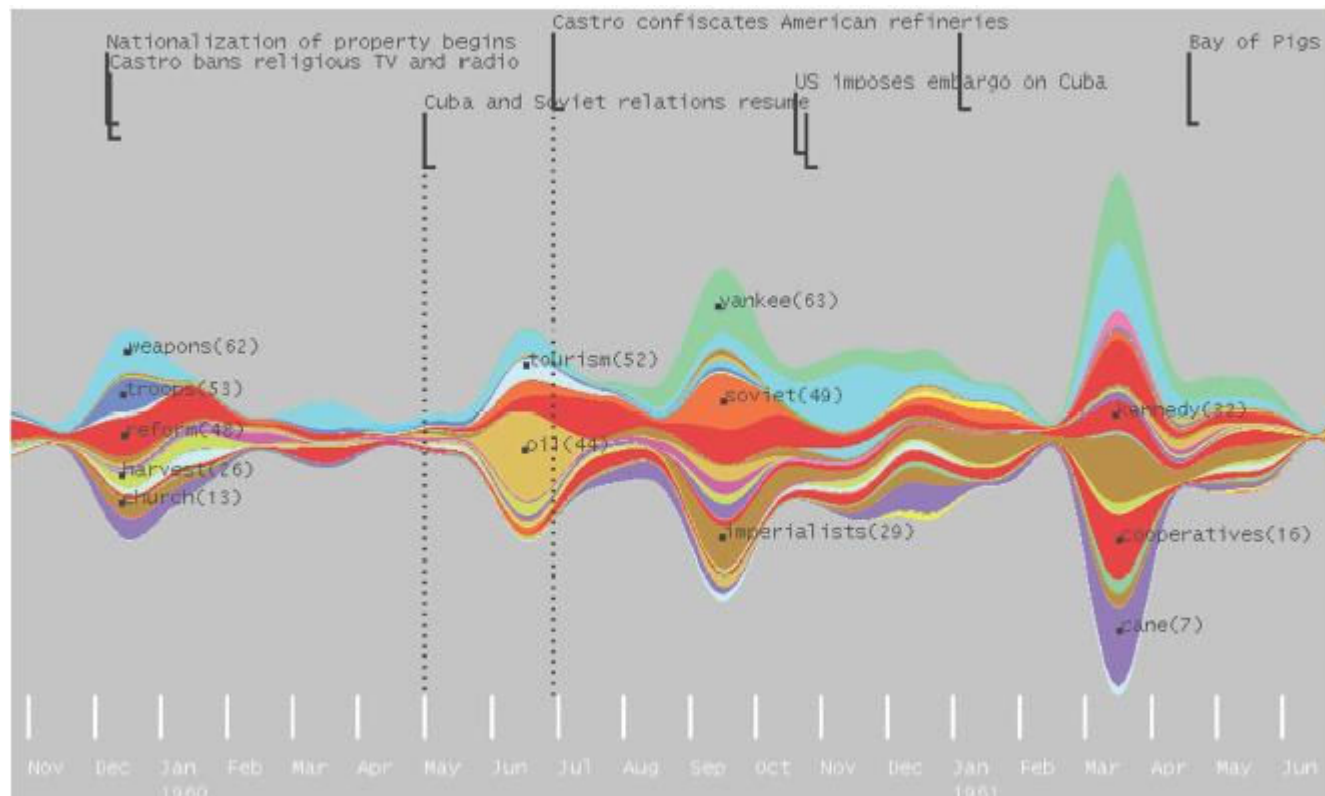
Aside: Stacked area graph / Stream graph

- Streams represent data element prominence over time
- Overall flow shows data volume



Theme River

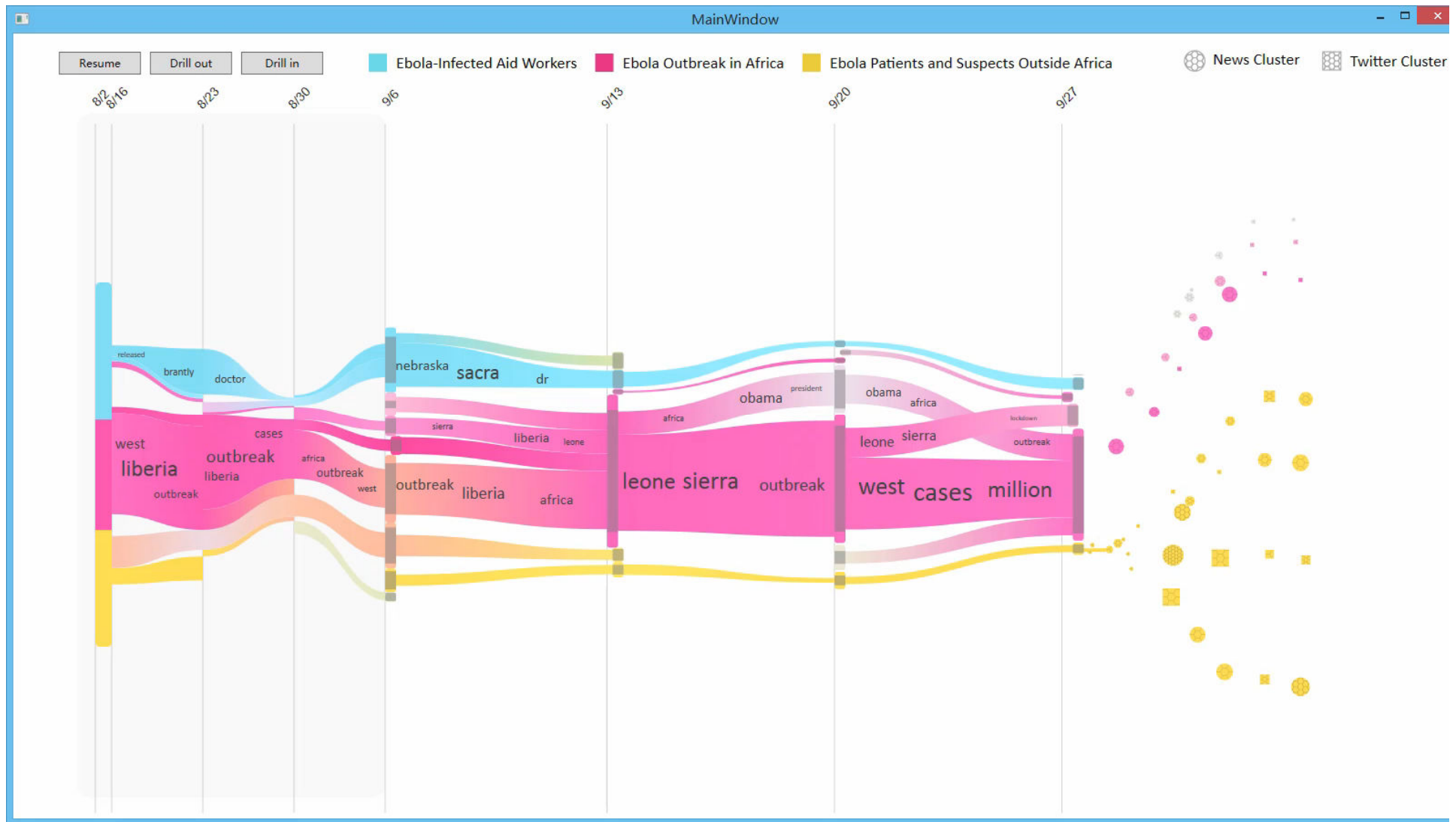
- Canonical example of stream graphs
- News themes over time



Susan Havre, IEEE Computer Society, *ThemeRiver: Visualizing Thematic Changes in Large Document Collections*, Vol.8. No.1., January-March 2002

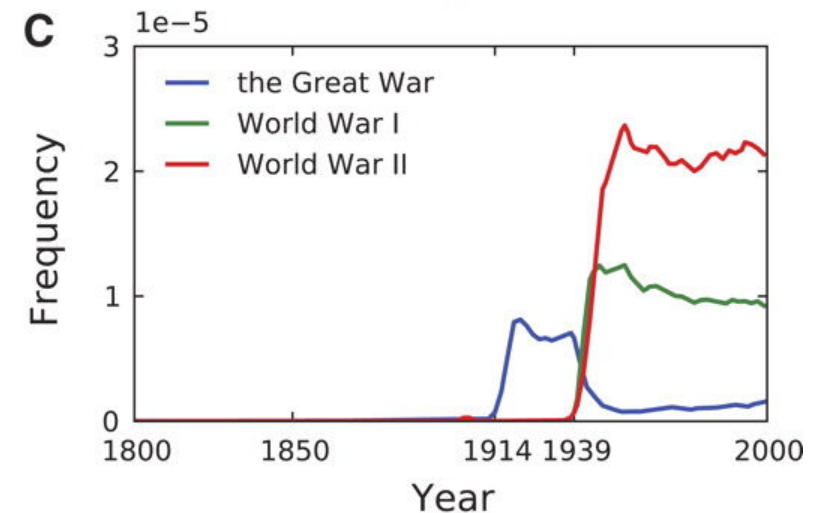
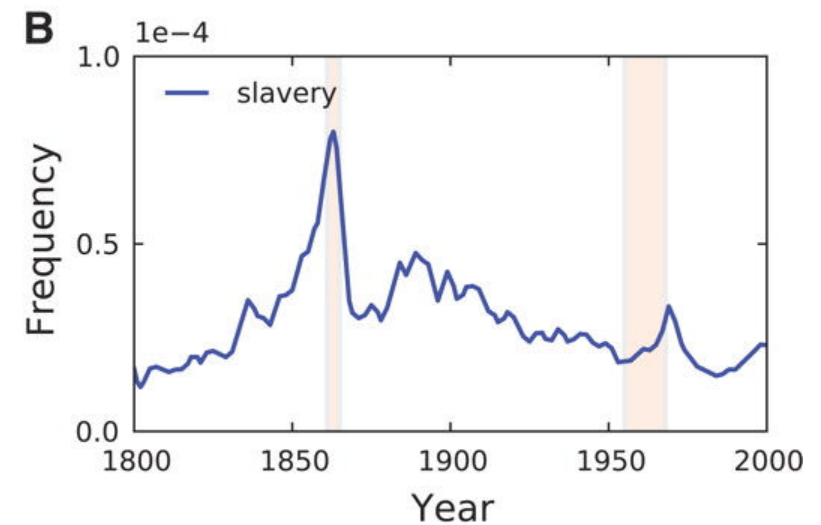
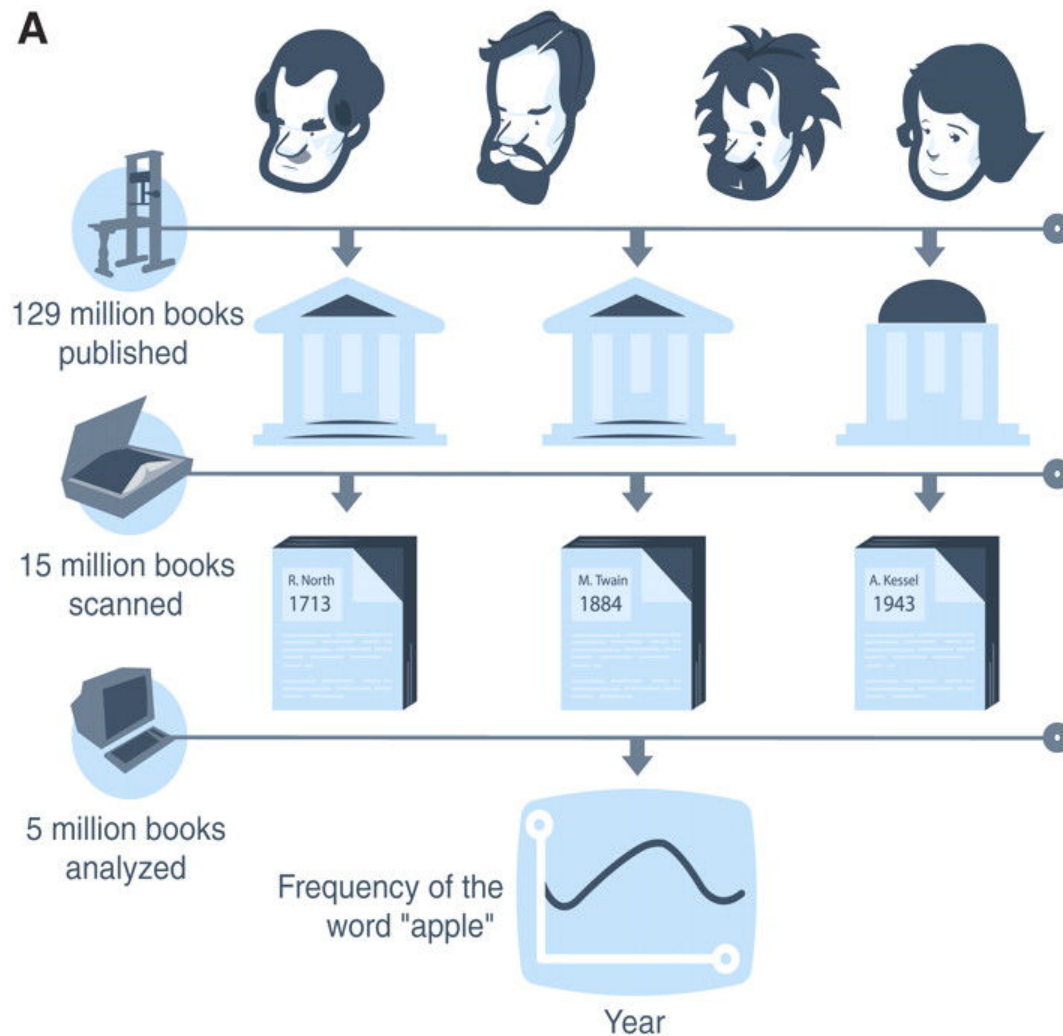
Figure 1 shows the Parallel Topics software interface. The main window displays a network graph with 15 topics (Topic 01 to Topic 15) on the x-axis and 15 topics on the y-axis. The graph shows connections between topics, with some topics having a '1.0' value. To the right, a 'TimeRiver' plot shows the evolution of topics over time from 2006 to 2010, with a color gradient from green to red. Below the network graph, a 'Topics' list shows the top 15 topics with their associated keywords and descriptions. The bottom right corner shows a 'Scatterplot' of Y Maximum vs X Entry.

47



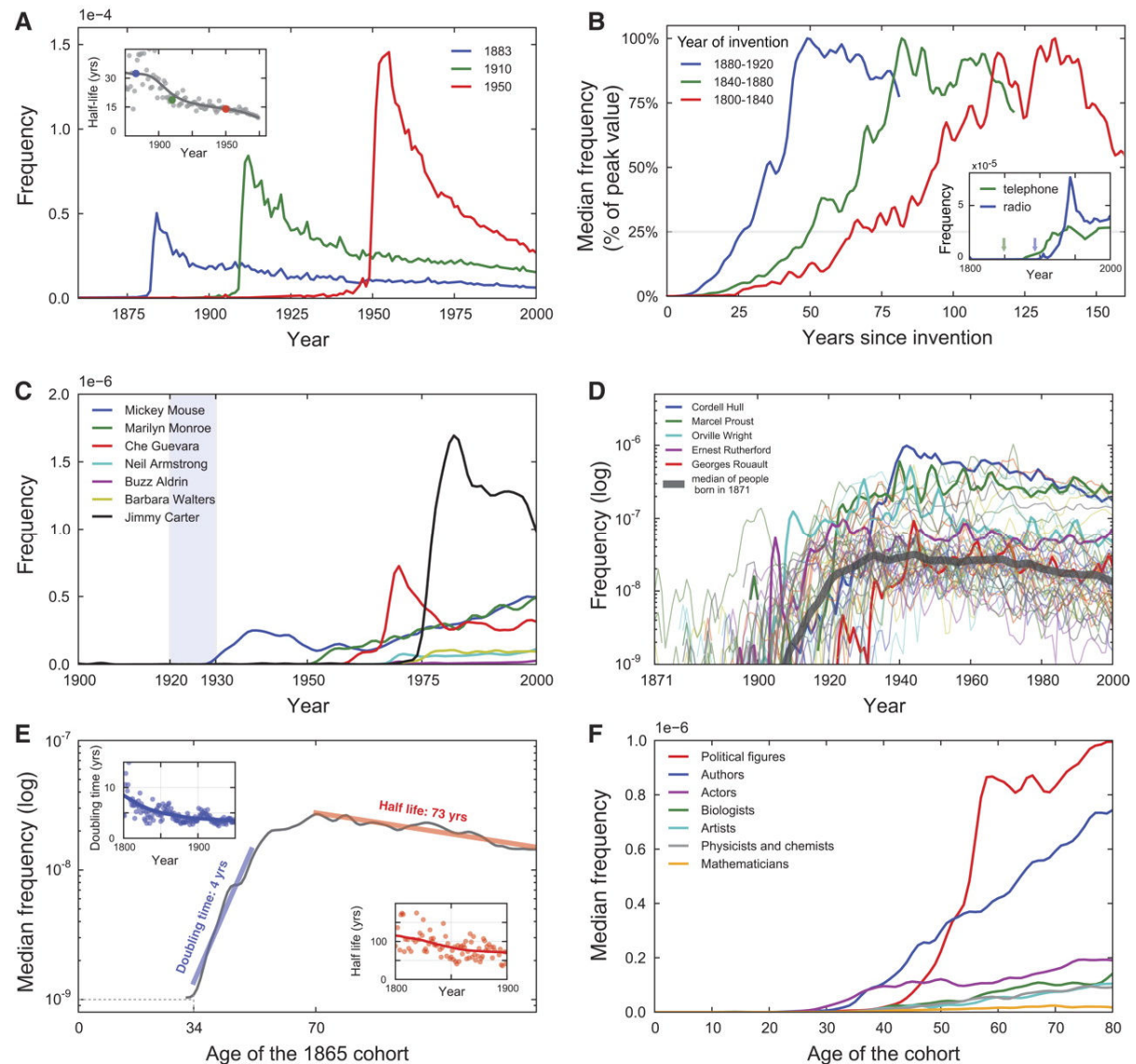
S. Liu, J. Yin, X. Wang, W. Cui, K. Cao and J. Pei, "Online Visual Analytics of Text Streams," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 11, pp. 2451-2466, Nov. 1 2016.

Large Scale Corpus Analysis: Culturomics



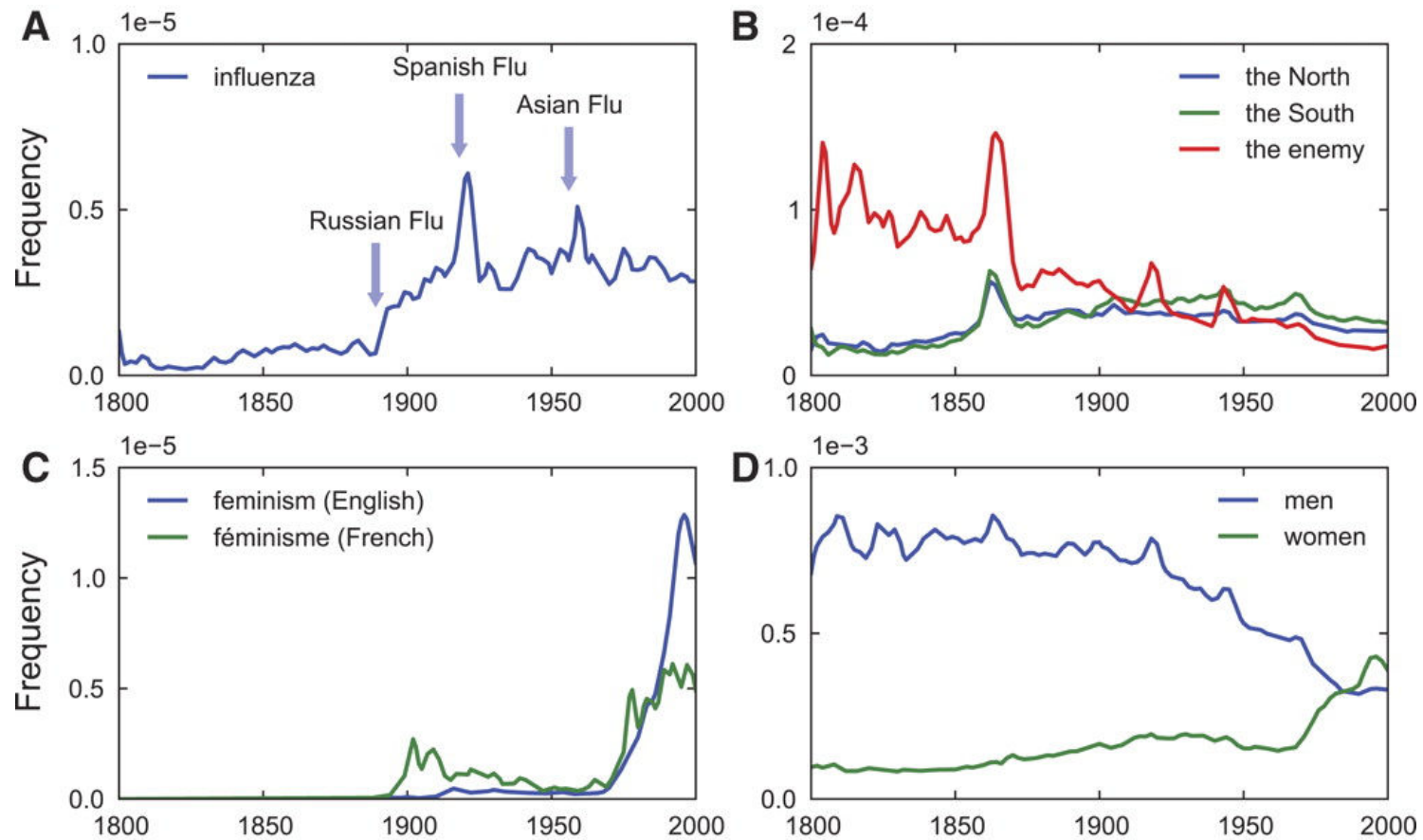
Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331, no. 6014 (January 14, 2011): 176–182.

Culturomics

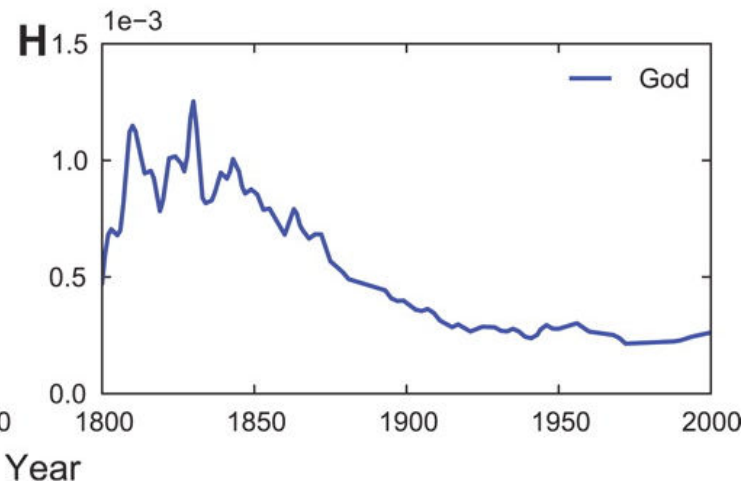
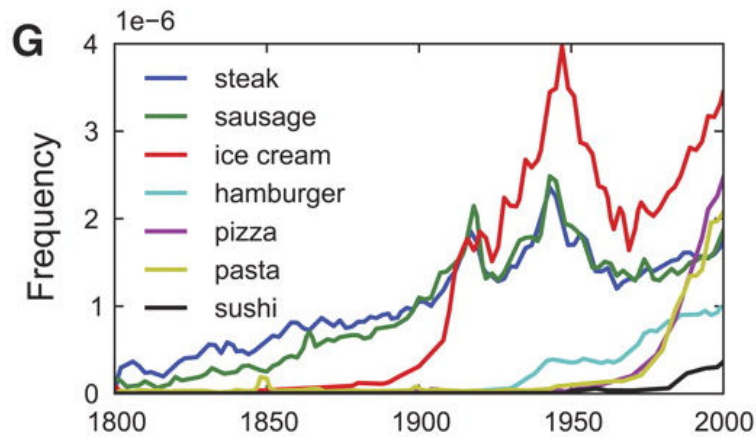
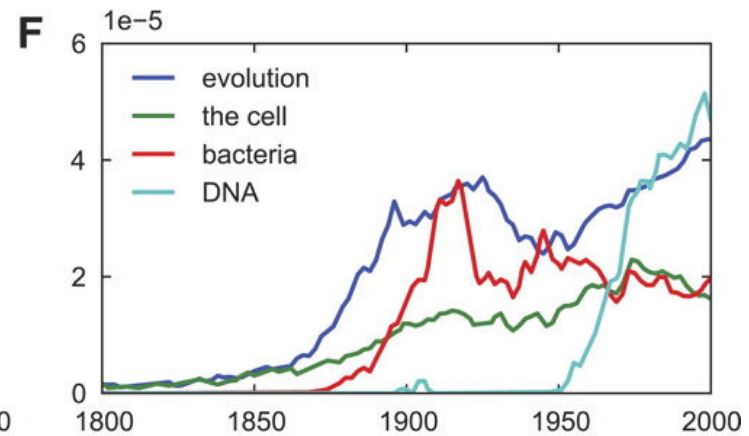
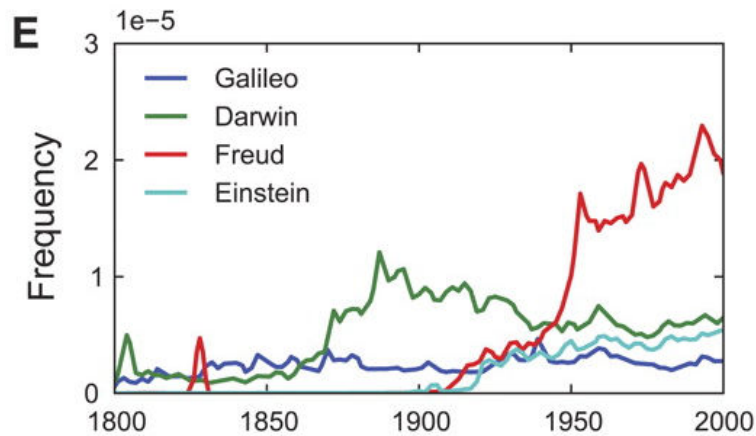


Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331, no. 6014 (January 14, 2011): 176–182.

Culturomics



Culturomics



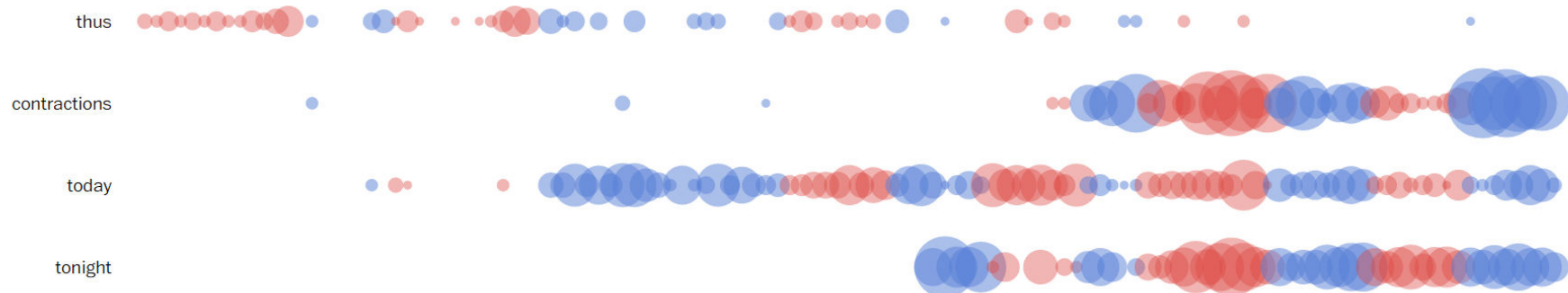
History of the President's Words

Daily Lexicon

With the **rise of contractions** and the **fall of "thus,"** Liberman said this is illustrative of a larger, secular trend. "This obviously interacts with genre, and in particular with formality," he said, stating that "won't" appears much more frequently in magazines than "will not," while less of a gap exists in academic papers.

State of the Union addresses incorporated contractions even later than secular dialogue, which hints at the level of formality presidents wish to project. Liberman suggested that the spoken version of the addresses may be more relaxed and contraction-laden than the much more prescribed written documents let on. "Perhaps Truman, Eisenhower and Kennedy were using contractions in their performances that were not written as such in their texts," he said. The increased use of contractions suggests a growing laxity in the written documents rather than in spoken words.

The **rise of "tonight,"** Fields clarified, was simply the result of the State of the Union addresses occurring in the evening time for the first time with Lyndon B. Johnson's 1965 address.



<https://www.washingtonpost.com/graphics/politics/2016-sotu/language/>

Culturomics: Language Change

EEBO N-GRAM BROWSER

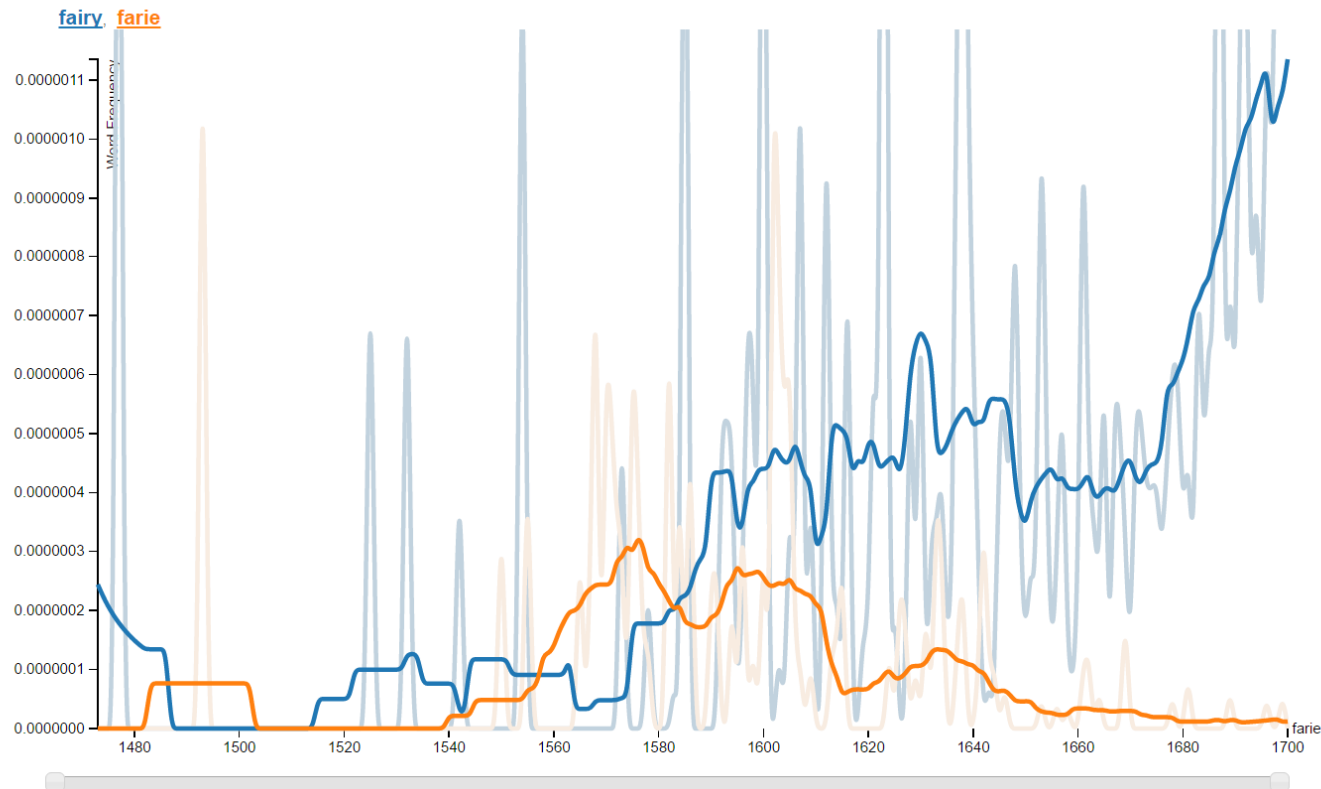
[SHOW INSTRUCTIONS](#)

ngramSize: ☒ unigrams ☐ bigrams ☐ trigrams
spellings: ☒ original ☐ regularized ☐ lemmas

gram 1: search term: pos:

☒ Graph smoothing

Rolling Average:



<http://earlyprint.wustl.edu/>

EEBO N-GRAM BROWSER

[SHOW INSTRUCTIONS](#)

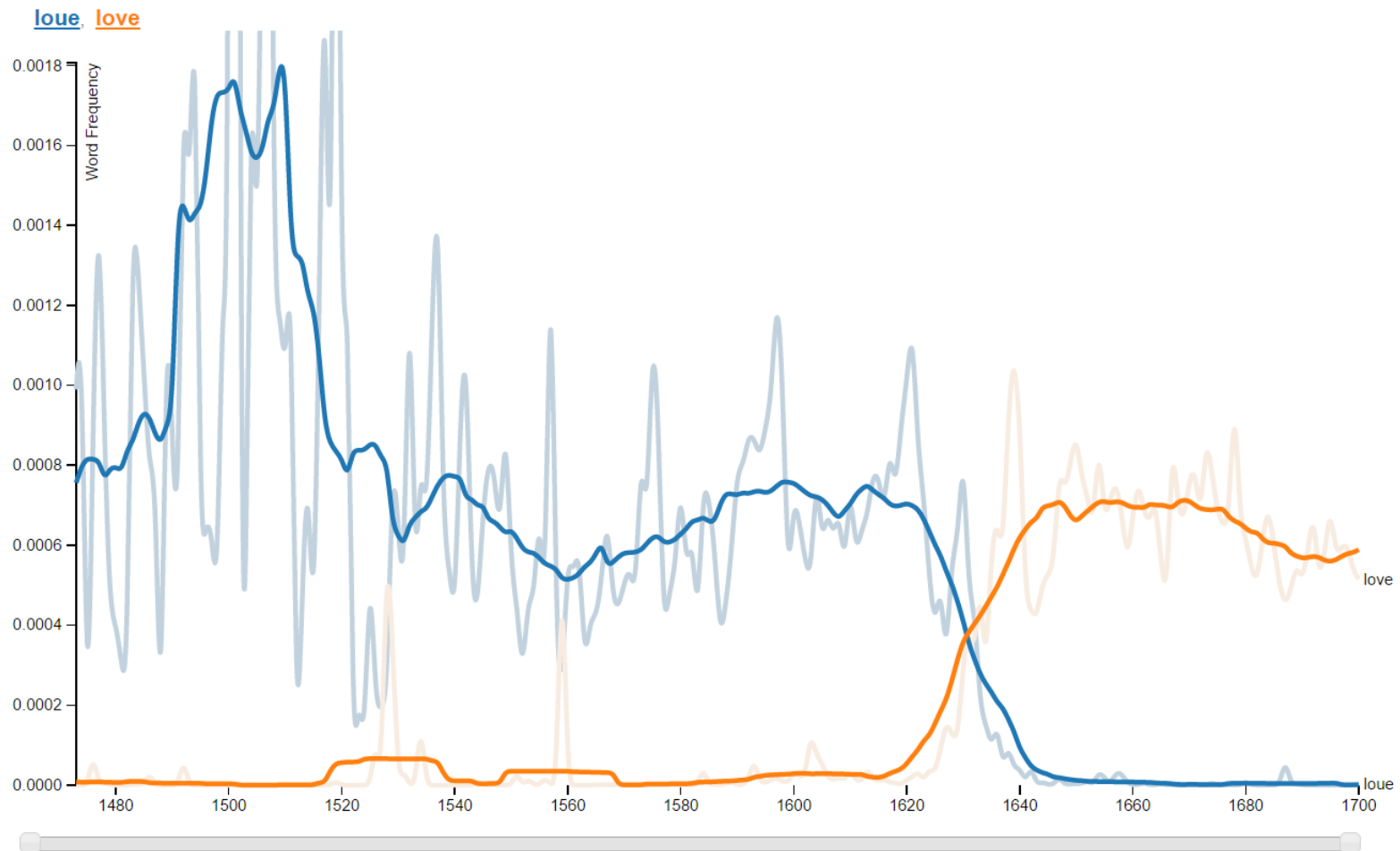
ngramSize: ☒ unigrams ☐ bigrams ☐ trigrams
spellings: ☒ original ☐ regularized ☐ lemmas

gram 1: search term: pos:

☒ Graph smoothing

Rolling Average:

Draw Graph





Metatation


Interdisciplinary Research Case Study

Hrim Mehta, Adam Bradley, Mark Hancock, Christopher Collins



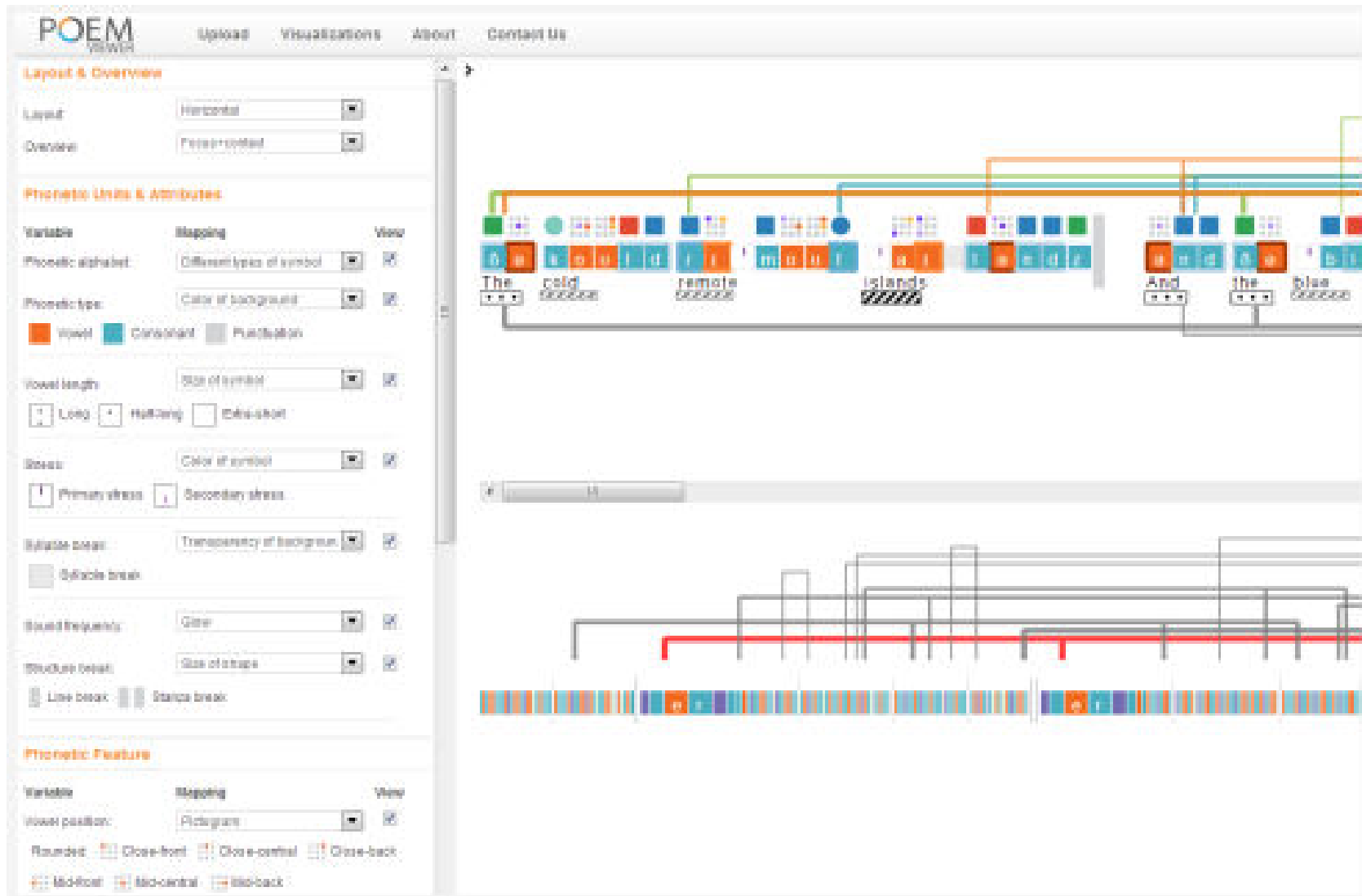
Project Overview

- Team Expertise:
 - Human-Computer Interaction
 - Computational Linguistics
 - English Literature and Poetry
- Methodology:
 - Grounding Study + Analysis
 - Derive Requirements
 - Prototyping & Design
 - Expert Feedback



Free-form annotations as implicit interactions

Existing Tools for Literary Analysis



PoemViewer by Oxford Visual Informatics Lab



Limitations of Existing Literary Analysis Tools

- Disconnected from workflow
- Premature data presentation



Understanding the Context

Observational study of
annotation practice of poetry critics
as they analyse a poem

Observational Study

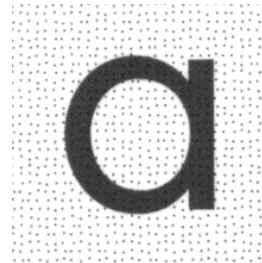
Participants

14 literary critics (3 PhD students & 11 university professors)

Task

Close reading of assigned **pairs of poems** printed on **Anoto** paper with Livescribe Anoto pen for annotations, if required

Analysis session followed by **retrospective think aloud**



Dataset

14 poems in total, **7 pairs of poems** assigned to 7 participants

Categorisation of Annotations

Annotation Form

Ellipses, Underlines, Connectors, Polygons, Brackets, Text, Miscellaneous notations

Cognitive Purpose

CO (Computational Offloading)

EML (Externalizing to reduce Memory Load)

EML + CO

Space on Page

Word space, White space, Margin

Results

- Pen-and-paper over digital tools
- Experiential cognition
- Access of external resources
- Polymorphism of annotations
 - Annotation **form** and annotation **function do not** have a **one-to-one** relation

Results

- Pen-and-paper over digital tools
- Experiential cognition
- Access of external resources
- Polymorphism of annotations
 - Annotation **form** and annotation **function do not** have a **one-to-one** relation
 - **Consistency** in participant activity but the **form, space & time** vary

Results

- Pen-and-paper over digital tools
- Experiential cognition
- Access of external resources
- Polymorphism of annotations
 - Annotation **form** and annotation **function do not** have a **one-to-one** relation
 - **Consistency** in participant activity but the **form, space & time** vary
 - Characteristics of **CO** & **EML**

CO Characteristics

making a short, unhealthy life the shorter.
I kill it, and another instant's added
to the horrifying mortmain of
ephemera: keys, drift, sea-urchin shells,
you packrat off with joy ... a dead fly swept

We stood by a pond that winter day,
And the sun was white, as though chidden of God,
And a few leaves lay on the starving sod;
– They had fallen from an ash, and were gray.

Your eyes on me were as eyes that rove
Over tedious riddles of years ago;
And some words played between us to and fro
On which lost the more by our love.

The smile on your mouth was the deadest thing

CO Characteristics

Had, having, and in quest to have, extreme;
A bliss in proof, and proved, a very woe;
Before, a joy proposed; behind, a dream.

CO Characteristics

making a short, unhealthy life the shorter.
I kill it, and another instant's added
to the horrifying mortmain of
ephemera: keys, drift, sea-urchin shells,
you packrat off with joy ... a dead fly swept

We stood by a pond that winter day,
And the sun was white, as though chidden of God,
And a few leaves lay on the starving sod;
— They had fallen from an ash, and were gray.

Your eyes on me were as eyes that rove
Over tedious riddles of years ago;
And some words played between us to and fro
On which lost the more by our love.

The smile on your mouth was the deadest thing

EML Characteristics

not one a fighter. It is like a plane
dusting apple orchards or Arabs on the screen—
one of the mighty ... one of the helpless. It
bumbles and bumps its brow on this and that,
making a short, unhealthy life the shorter.
I kill it, and another instant's added
to the horrifying mortmain of

frail vs. fatal
another
linking of
opposites.

Days
↓
Day - she judges personal relationship with speaker - he feels her judgment

So Days are feminized
- bring gifts
- ritual, ceremony
- gifts - precious + necessary

The "I" is careless, random, busy

Pay attention
Be grateful for nature's gifts.

10

Data Analysis Tool

A

EML
 CO
 A
 EML + CO

B

Ref: |IC| CO ~ 1

Alliteration in 'b's of 'blueback' and 'th's of 'thumb' and 'thick'

1. |IC| CO ~ 2

Annotated Text:

StartTime: 135.20
EndTime: 139.94

Ref: |IC| CO ~ 2

Alliteration in 'b's of 'bumbles', 'bumps' and 'brows' and 'th's of 'this' and 'that'

1. |IC| CO ~ 1

Annotated Text:

StartTime: 164.78
EndTime: 172.46

Ref: |IC| EML ~ 3

Note stating that the poetic form is "Sonnet" after counting the number of lines in the poem

Ref: |IC| CO ~ 1

C

Harriet - Robert Lowell

A repeating fly, blueback, thumb thick—so gross,
it seems apocalyptic in our house—
whams back and forth across the nursery bed
manned by a madhouse of stuffed animals,
not one a fighter. It is like a plane
dusting apple orchards or Arabs on the screen—
one of the mighty . . . one of the helpless. It
bumbles and bumps its brow on this and that,
making a short, unhealthy life the shorter.
I kill it, and another instant's added
to the horrifying mortmain of
ephemera: keys, drift, sea-urchin shells,
you packrat off with joy . . . a dead fly swept
under the carpet, wrinking to fulfillment.

Handwritten notes:
Sonnet
life vs. death
active vs. passive
rhyme
basic metric
personification
sound
verbs
past simple
verb will
active

D

Harriet - Robert Lowell

A repeating fly, blueback, thumb thick—so gross
it seems apocalyptic in our house—
whams back and forth across the nursery bed
manned by a madhouse of stuffed animals,
not one a fighter. It is like a plane
dusting apple orchards or Arabs on the screen—
one of the mighty . . . one of the helpless. It
bumbles and bumps its brow on this and that,
making a short, unhealthy life the shorter.
I kill it, and another instant's added
to the horrifying mortmain of
ephemera: keys, drift, sea-urchin shells,

E

Harriet - Robert Lowell

A repeating fly, blueback, thumb thick—so gross,
it seems apocalyptic in our house—
whams back and forth across the nursery bed
manned by a madhouse of stuffed animals,
not one a fighter. It is like a plane
dusting apple orchards or Arabs on the screen—
one of the mighty . . . one of the helpless. It
bumbles and bumps its brow on this and that,
making a short, unhealthy life the shorter.
I kill it, and another instant's added
to the horrifying mortmain of
ephemera: keys, drift, sea-urchin shells,
you packrat off with joy . . . a dead fly swept
under the carpet, wrinking to fulfillment.

04:32/51:51

Annotations

Timeline Annotations

272.8104248046875

Data Coding Example

A

PRINTABLE NOTENO 1

Sonnet 129 - William Shakespeare

best

's'

*anaphora
lust is inhouse
body not head*

*Who lays
it?*

spad

Th' expense of spirit in a waste of shame
Is lust in action; and till action, lust
Is perjured, murderous, bloody, full of blame,
Savage, extreme, rude, cruel, not to trust;
Enjoyed no sooner but despised straight;
Past reason hunted; and no sooner had,
Past reason hated, as a swallowed bait,
On purpose laid to make the taker mad;
Mad in pursuit, and in possession so;
Had, having, and in quest to have, extreme;
A bliss in proof, and proved, a very woe;
Before, a joy proposed; behind, a dream.
All this the world well knows; yet none knows well
To shun the heaven that leads men to this hell.

a
b
a
b
c
a
c
d
e
f
e
f
g
g

*series of reversals
in sentence
structure*

*diction
harsh
+ extreme*

*lust had
past had
outcome*

*Very preachy tone
emphasized by
stop/pauses within
lines
- breaks up the regular
iambic cadence*

① Before lust becomes action - it's dangerous
② When acted upon - poisonous consequences
③ Not as good as the dream

After

We all know this, but we act on lust anyway

h sounds = declarative
p/b sounds = negative, harsh
long "a" = almost wailing
's' = sibilant, sneaky

record stop jump backmark 0% jump to position 100% playback speed volume

B

Had, having, and in quest to have, extreme;
A bliss in proof, and proved, a very woe;
Before, a joy proposed; behind, a dream.

C

h sounds = declarative
p/b sounds = negative, harsh
long "a" = almost wailing
's' = sibilant, sneaky



Implications for Design

- Support for free-form annotations
- Permitting experiential cognition
- Minimal interruption to work flow
- Invoking analytic assistance via annotations

Metatation

Desktop-based **application + physical paper**

- Support for free-form annotations

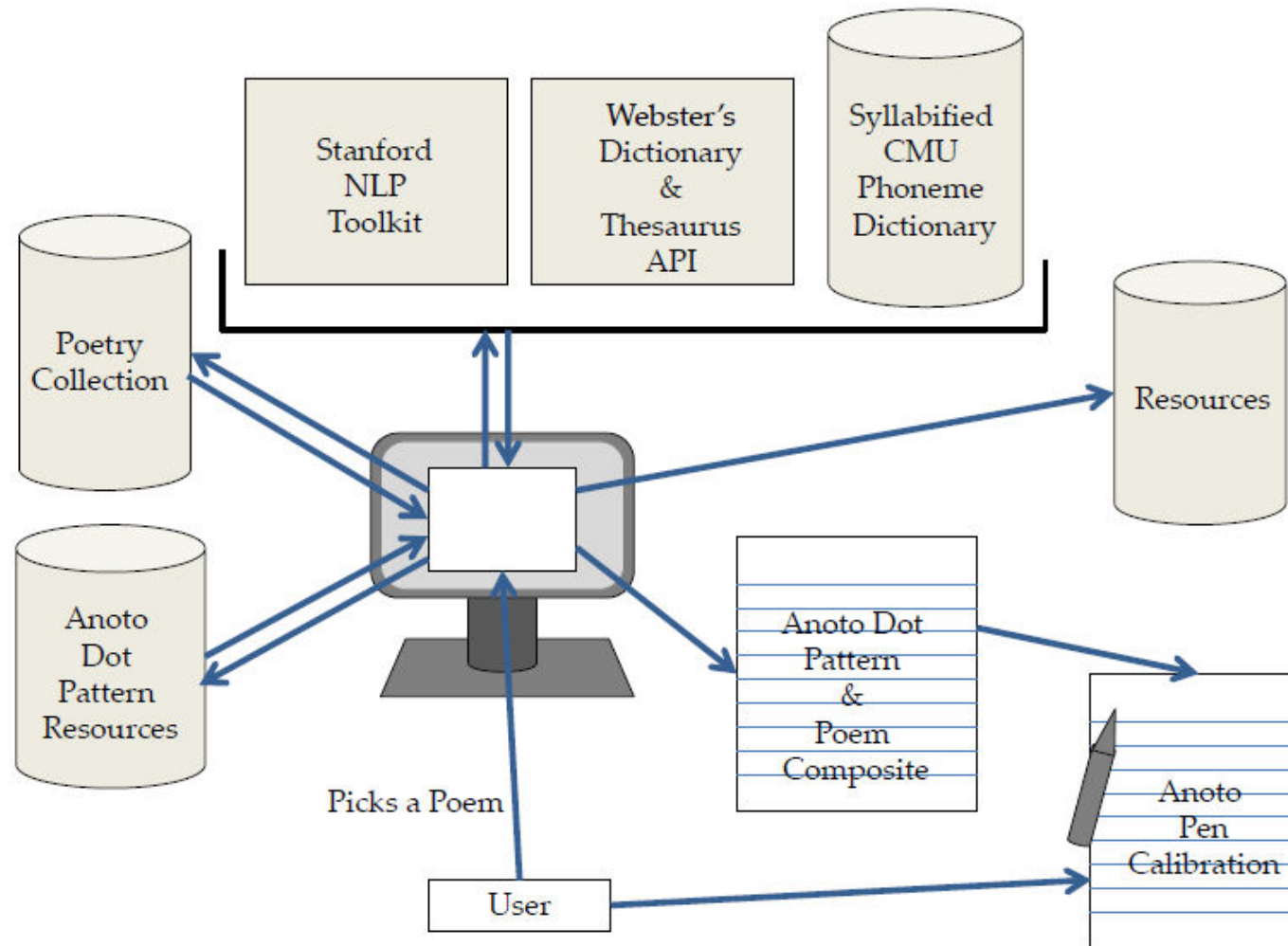
CO annotations for generating **supplementary data**

- Permitting experiential cognition
- Invoking analytic assistance via annotations

Analytic assistance presented on **peripheral display**

- Minimal interruption to work flow

Metatation: Preprocessing



Assonance, Consonance & Alliteration

Assonance

- Repetition of **vowel sounds**
- Rude: (R/**UW1**/D) & Cruel: (K/R/**UW1**/L)

Consonance

- Repetition of **consonant sounds**
- Extreme: (IH0/K)(S/T/**R**/IY1/M) & Rude: (**R**/UW1/D)

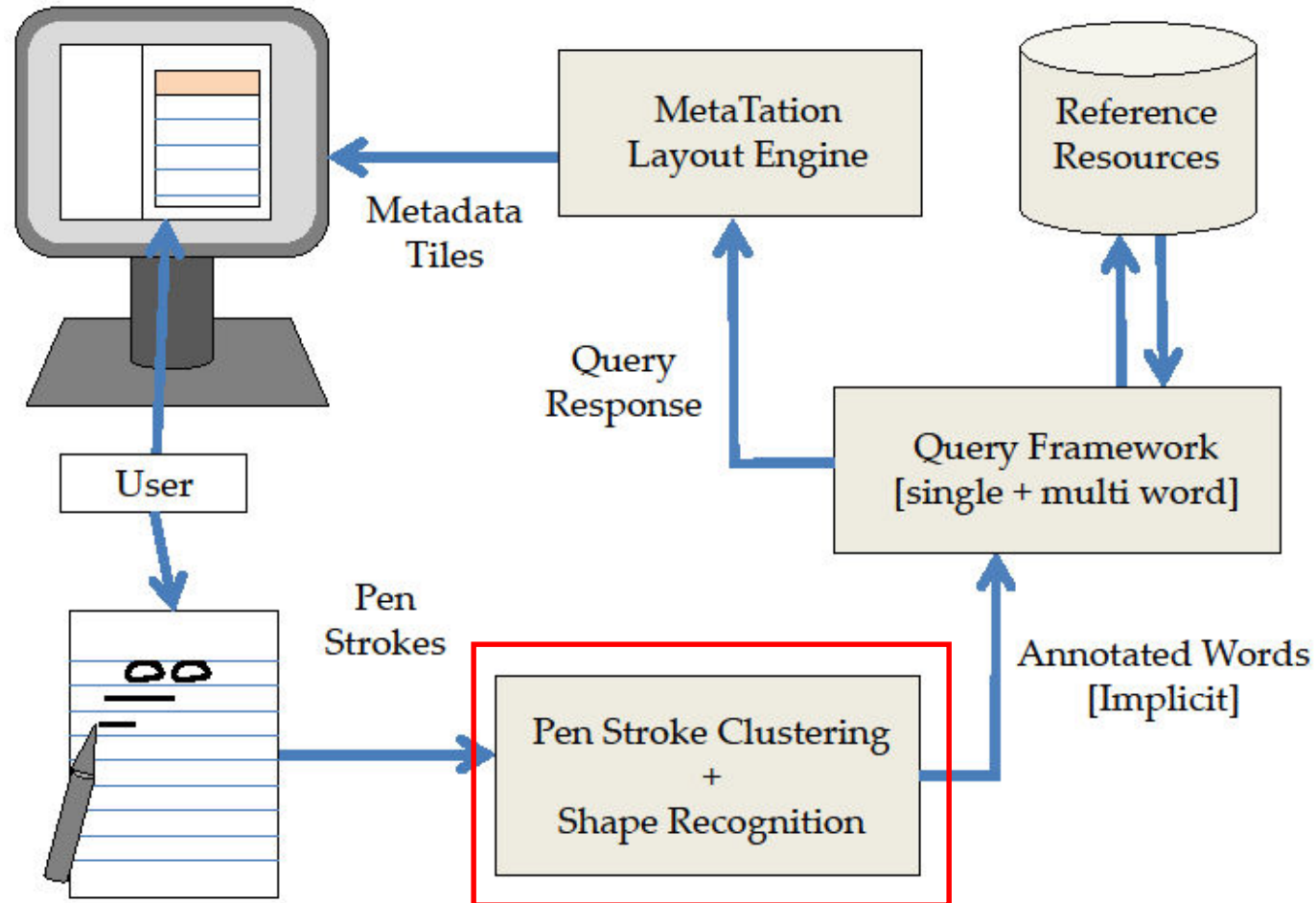
Alliteration

- Repetition of **consonant sounds** in **stressed syllable**
- Proof: (**P**/R/UW1/F), Proved: (**P**/R/UW1/V/D) & Proposed: (P/R/AH0)(**P**/OW1/Z/D)

End Rhyme

- If common vowel sound in stressed syllable, all prior sounds dissimilar & all following sounds similar
- Shame: (SH/**EY1**/M) & Blame: (BL/**EY1**/M)

Metatation: Stroke Clustering & Shape Recognition



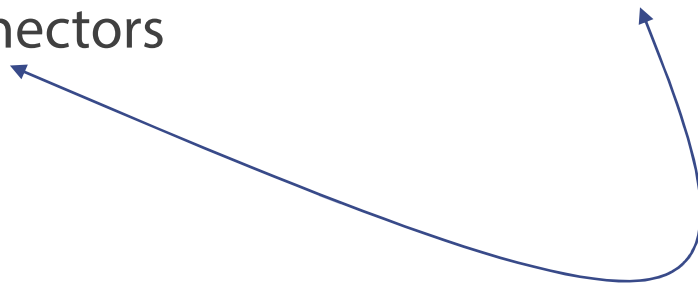
Metatation: Stroke Clustering & Shape Recognition

Stroke clustering

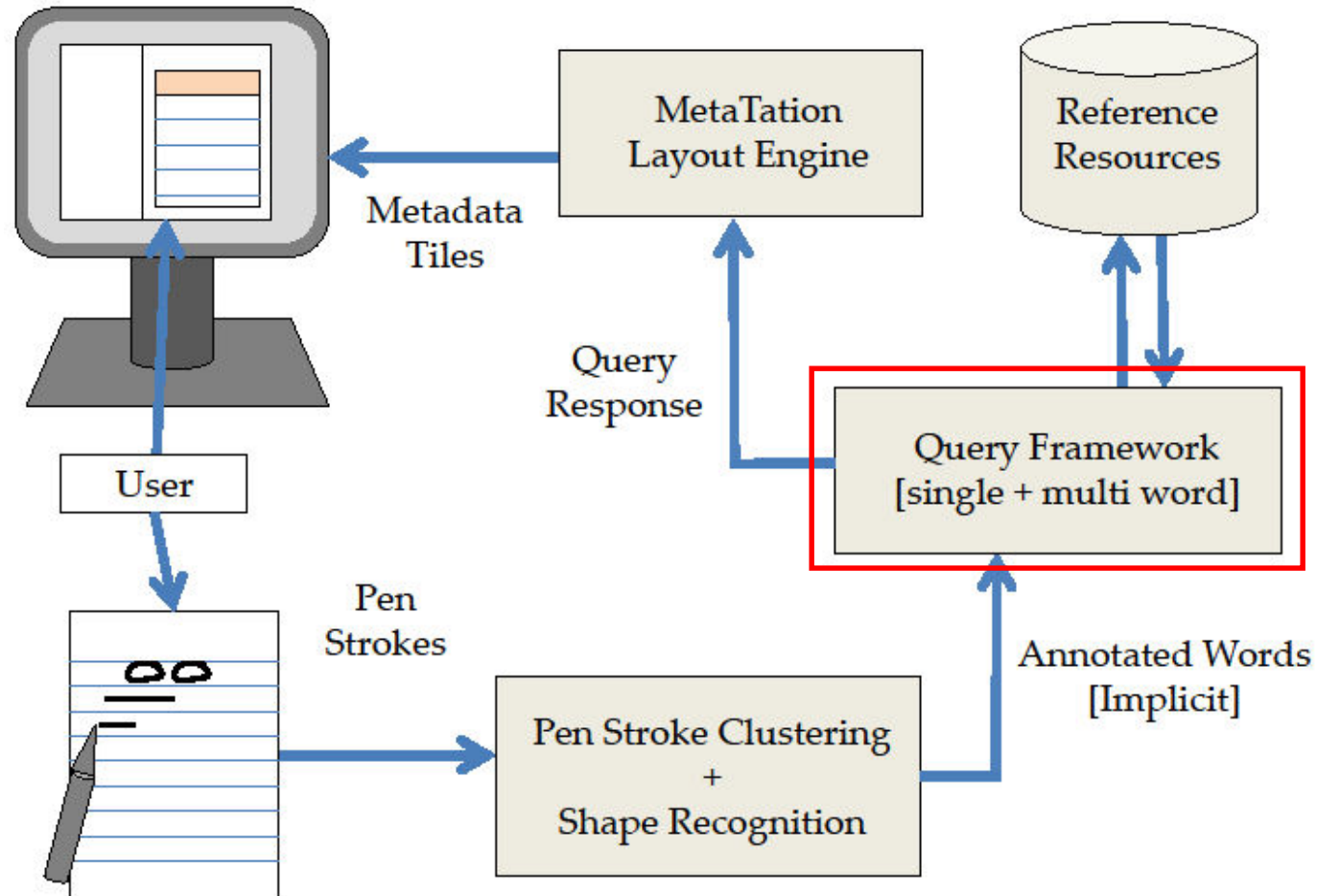
- Hierarchical agglomerative clustering
- **Spatiotemporal distance** between pen strokes as distance metric for the clustering process (CO characteristics)

Shape Recognition

- Geometric recognizer for detecting underlines, ellipses and connectors



Metatation: Query Framework



Metatation: Query Framework

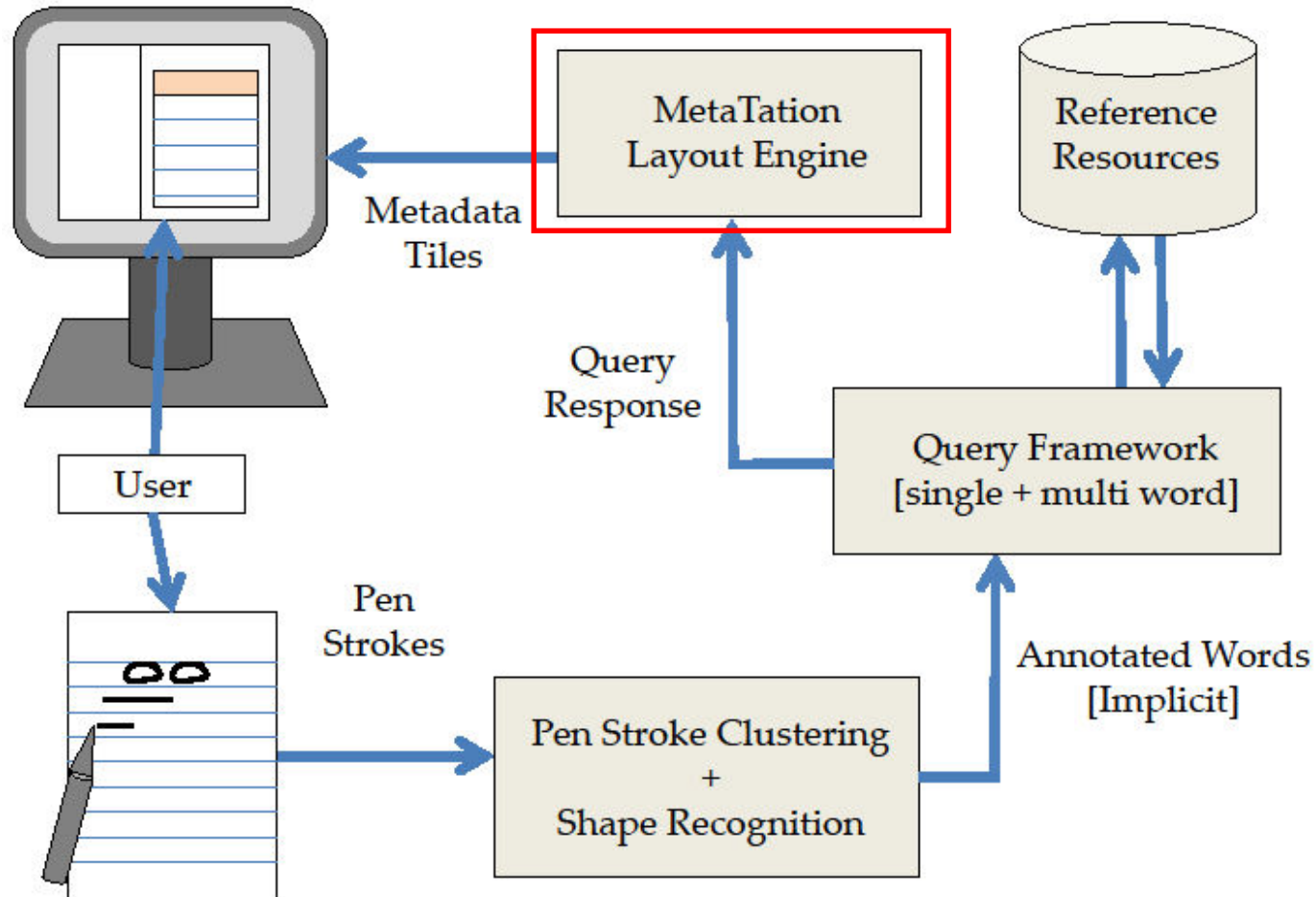
Semantic Relations

- Word details, **Synonyms**, **Antonyms**
- Source: Merriam Webster Dictionary & Thesaurus

Phonetic Relations

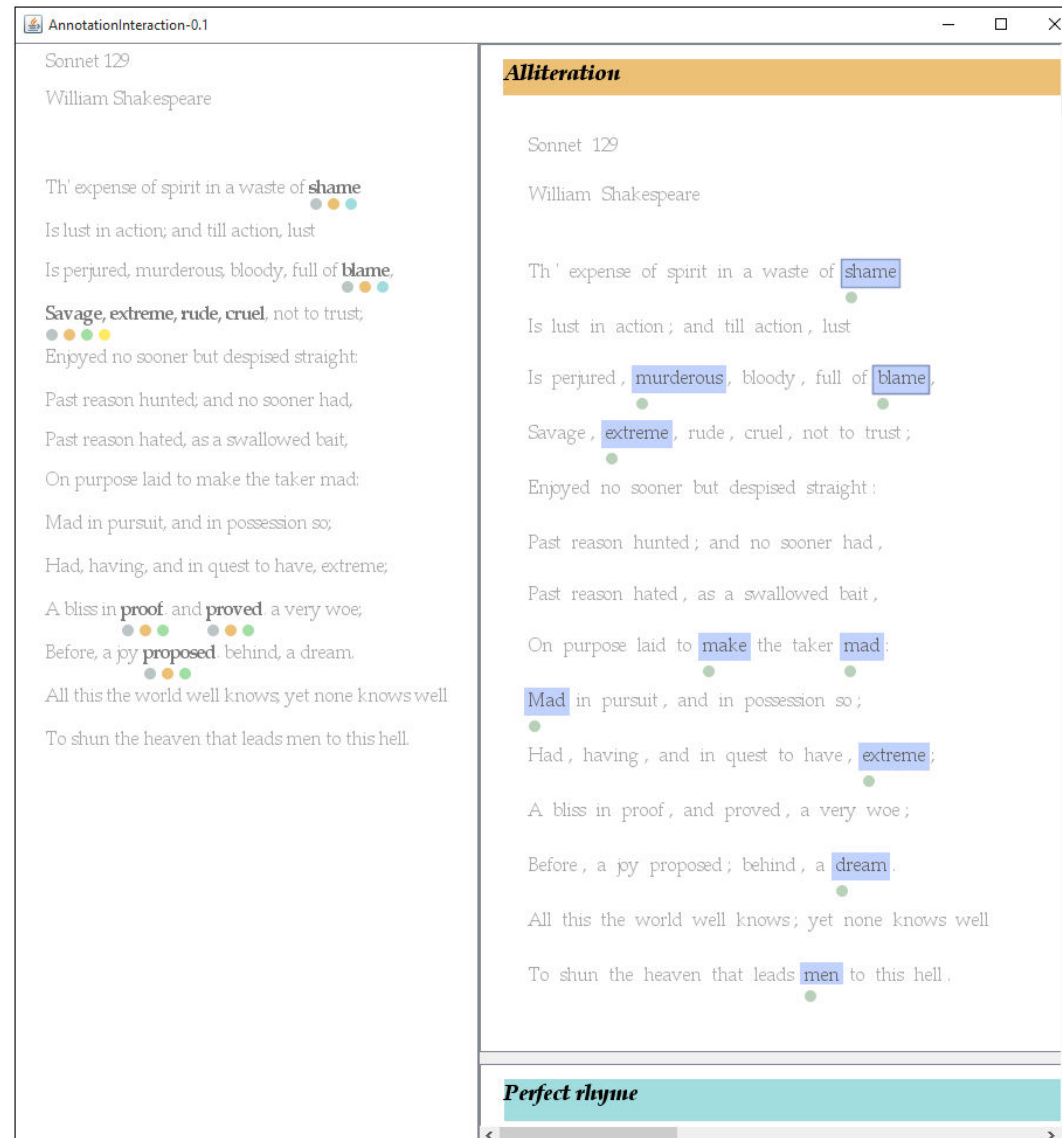
- **Assonance**, **Consonance**, **Alliteration**, **End rhyme**
- Sources:
 - CMU Phoneme Dictionary (in NLTK),
 - Broken into syllables with algorithm in “On the syllabification of phonemes” by Bartlett et al. (Proc. HLT-ACL 2009)

Metatation: Interface



Metatation: Interface

Worksheet
Viewer
Panel



Metadata
Tile Stream
Panel

Assonance, Consonance, Alliteration Metadata Tile

Assonance

Sonnet 129

William Shakespeare

Th' expense of spirit in a waste of shame

Is lust in action; and till action, lust

Is perjured, murderous, bloody, full of blame,

Savage, extreme, rude, cruel, not to trust;

Enjoyed no sooner but despised straight:

Past reason hunted; and no sooner had,

Past reason hated, as a swallowed bait,

On purpose laid to make the taker mad;

Mad in pursuit, and in possession so;

Had, having, and in quest to have, extreme;

A bliss in proof, and proved, a very woe;

Before, a joy proposed; behind, a dream.

All this the world well knows; yet none knows well

To shun the heaven that leads men to this hell.

Assonance

Sonnet 129

William Shakespeare

Th' expense of spirit in a waste of shame

Is lust in action; and till action, lust

Is perjured, murderous, bloody, full of blame,

Savage, extreme, rude, cruel, not to trust;

Enjoyed no sooner but despised straight:

Past reason hunted; and no sooner had,

Past reason hated, as a swallowed bait,

On purpose laid to make the taker mad:

Mad in pursuit, and in possession so;

Had, having, and in quest to have, extreme;

A bliss in proof, and proved, a very woe;

Before, a joy proposed; behind, a dream.

All this the world well knows; yet none knows well

To shun the heaven that leads men to this hell.

Assonance

Sonnet 129

William Shakespeare

Th' expense of spirit in a waste of shame

Is lust in action; and till action, lust

Is perjured, murderous, bloody, full of blame,

Savage, extreme, rude, cruel, not to trust;

Enjoyed no sooner but despised straight:

Past reason hunted; and no sooner had,

Past reason hated, as a swallowed bait,

On purpose laid to make the taker mad;

Mad in pursuit, and in possession so;

Had, having, and in quest to have, extreme;

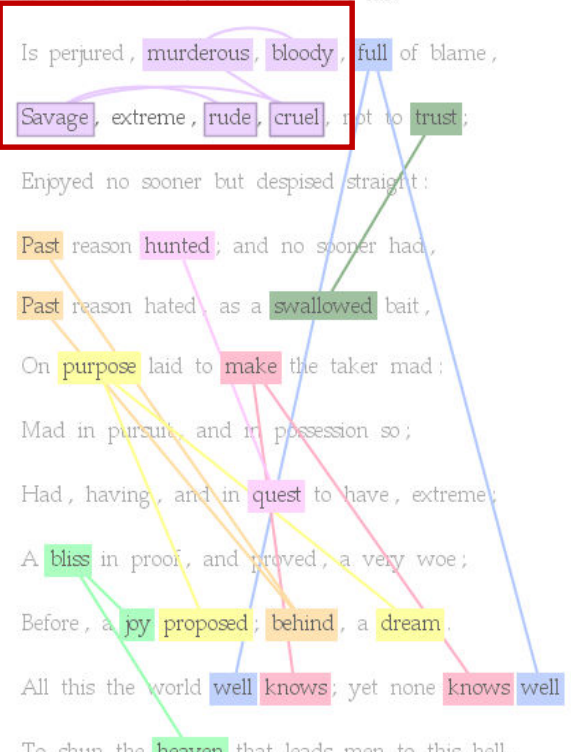
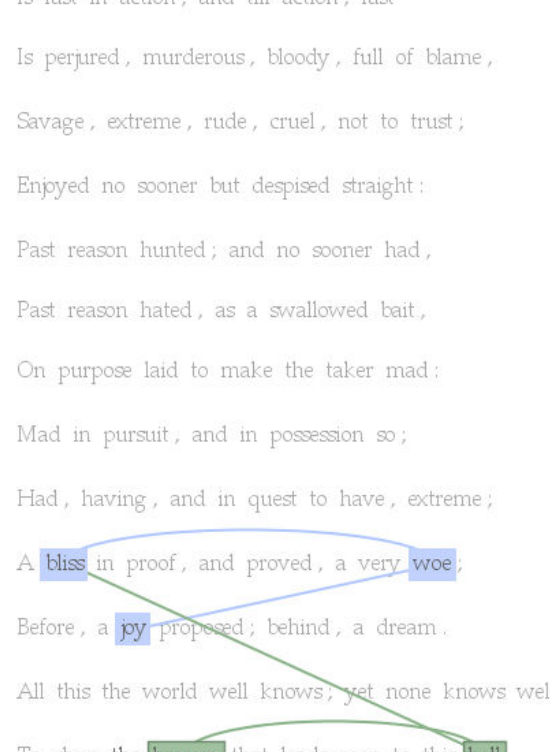
A bliss in proof, and proved, a very woe;

Before, a joy proposed; behind, a dream.

All this the world well knows; yet none knows well

To shun the heaven that leads men to this hell.

Synonyms & Antonyms Metadata Tile

Synonyms	Antonyms
<p>Sonnet 129</p> <p>William Shakespeare</p> <p>Th' expense of spirit in a waste of shame</p> <p>Is lust in action; and till action, lust</p> <p>Is perjured, murderous, bloody, full of blame,</p> <p>Savage, extreme, rude, cruel, not to trust;</p> <p>Enjoyed no sooner but despised straight:</p> <p>Past reason hunted, and no sooner had,</p> <p>Past reason hated, as a swallowed bait,</p> <p>On purpose laid to make the taker mad:</p> <p>Mad in pursuit, and in possession so;</p> <p>Had, having, and in quest to have, extreme;</p> <p>A bliss in proof, and proved, a very woe;</p> <p>Before, a joy proposed; behind, a dream.</p> <p>All this the world well knows; yet none knows well</p> <p>To shun the heaven that leads men to this hell.</p> 	<p>Sonnet 129</p> <p>William Shakespeare</p> <p>Th' expense of spirit in a waste of shame</p> <p>Is lust in action; and till action, lust</p> <p>Is perjured, murderous, bloody, full of blame,</p> <p>Savage, extreme, rude, cruel, not to trust;</p> <p>Enjoyed no sooner but despised straight:</p> <p>Past reason hunted; and no sooner had,</p> <p>Past reason hated, as a swallowed bait,</p> <p>On purpose laid to make the taker mad:</p> <p>Mad in pursuit, and in possession so;</p> <p>Had, having, and in quest to have, extreme;</p> <p>A bliss in proof, and proved, a very woe;</p> <p>Before, a joy proposed; behind, a dream.</p> <p>All this the world well knows; yet none knows well</p> <p>To shun the heaven that leads men to this hell.</p> 

End Rhyme Metadata Tile

Perfect rhyme

Sonnet 129

William Shakespeare

Th' expense of spirit in a waste of shame
Is lust in action; and till action, lust
Is perjured, murderous, bloody, full of blame,
Savage, extreme, rude, cruel, not to trust;
Enjoy'd no sooner but despised straight;
Past reason hunted; and no sooner had,
Past reason hated, as a swallowed bait,
On purpose laid to make the taker mad;
Mad in pursuit, and in possession so;
Had, having, and in quest to have, extreme;
A bliss in proof, and proved, a very woe;
Before, a joy proposed; behind, a dream.
All this the world well knows; yet none knows well
To shun the heaven that leads men to this hell.

```
graph TD
    shame[shame] --- blame[blame]
    lust[lust] --- trust[trust]
    extreme[extreme] --- woe[woe]
    straight[straight] --- bait[bait]
    had[had] --- mad[mad]
    so[so] --- had2[had]
    extreme2[extreme] --- woe2[woe]
    bliss[bliss] --- dream[dream]
    well1[well] --- hell[hell]
    shun[shun] --- hell2[hell]
```

Word Details Metadata Tile

shame

Origin

Middle English, from Old English [scamu;] akin to Old High German [scama] shame

Pronuciations

(SHEY1.M)

Noun

First recorded use

before 12th century

Senses

Sense 1

a painful emotion caused by consciousness of guilt, shortcoming, or impropriety

the susceptibility to such emotion

Usage examples: have you no [shame]?

Sense 2

a condition of humiliating disgrace or disrepute :[ignominy]

Usage examples: the [shame] of being arrested

Sense 3

something that brings censure or reproach also something to be regretted :[pity]

Usage examples: it's a [shame] you can't go

a cause of feeling shame

Verb

First recorded use

13th century

Senses

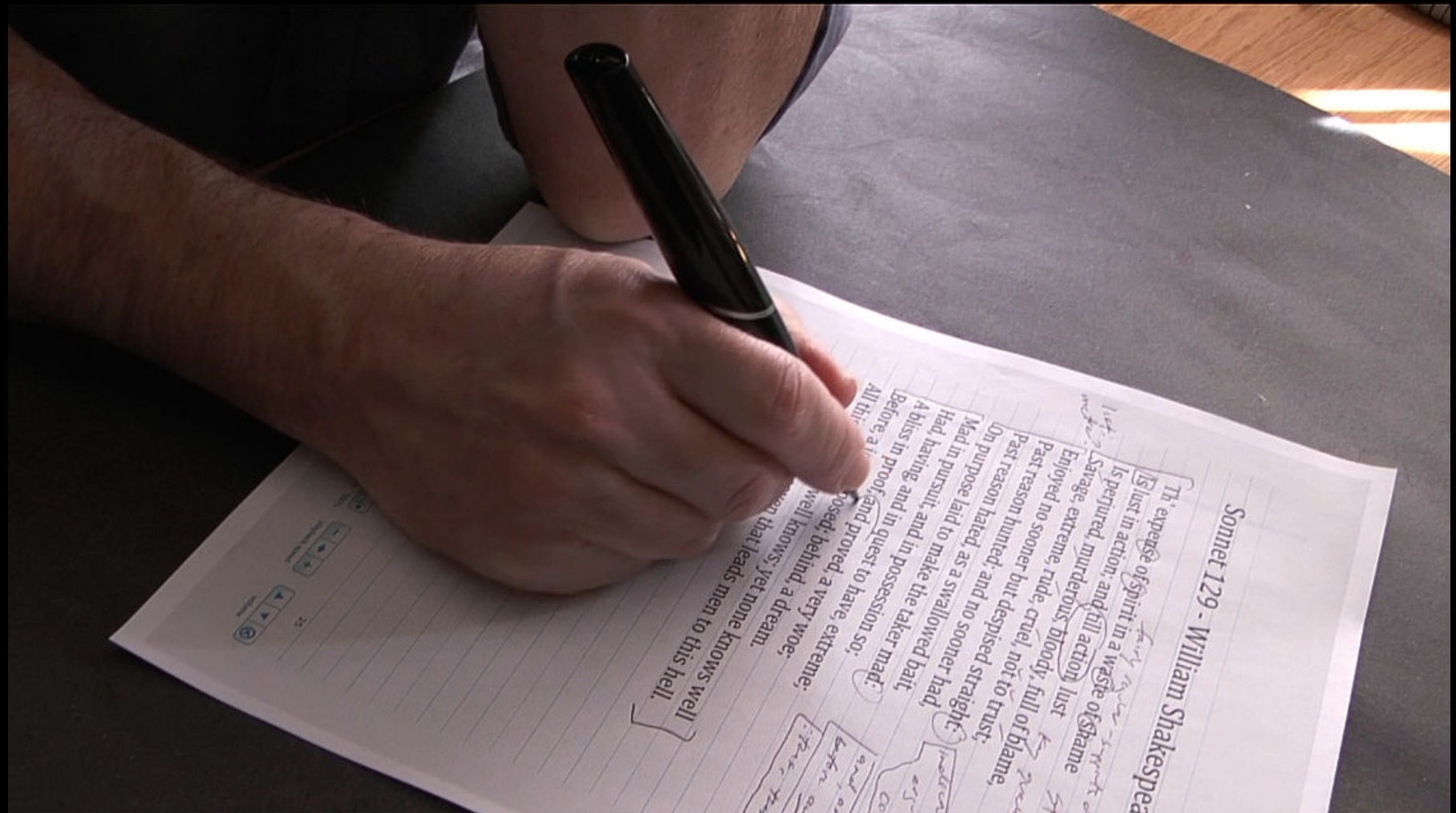
Sense 1

to bring shame to :[disgrace]

Usage examples: [shamed] the family name

Sonnet 129 - William Shakespeare

Th' expense of spirit in a waste of shame
Is lust in action; and fill action lust
Is perjured, murderous, bloody, lust
Savage, extreme, rude, cruel, full of blame,
Employed no sooner but despised straight;
Past reason hunted; and no sooner had,
Past reason hated, as a swallowed bait,
On purpose laid to make the taker mad;
Mad in pursuit, and in possession so;
Mad, having, and in quest to have;
A bliss in proof, and proved, a very woe;
Before, a joy proposed, behind, a dream.
All this the world well knows; yet none knows well
How that leads men to this hell.



Preliminary Evaluation

Evaluation Design

- 2 of the domain experts from the previous study explored the tool as they performed a reading of an assigned poem

Qualitative Results

- 2 different modes of use of MetaTation based on when the reader chooses to interact with the system
- MetaTation for teaching vs. research based analysis

Contributions

- Results of an **observational study** of poetry critics
- **Design guidelines** for the development of digital tools for supporting linguistic inquiry
- Design and implementation of **Metatation**
- Results of **preliminary expert review** of our tool



VisArgue

Interdisciplinary Research Case Study

Visual Analytics
(Daniel Keim)

Linguistics
(Miriam Butt)

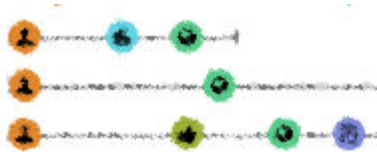
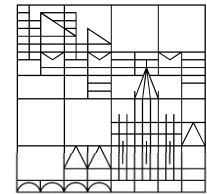
Political science
(Katharina Holzinger)

A word cloud where the letters of the word 'visAArgue' are formed by smaller words. The letters are primarily blue and orange. The word 'visAArgue' is written in a stylized font with a mix of uppercase and lowercase letters.

Visualizing the linguistic structure
of deliberative communication



Universität
Konstanz



Katharina Holzinger **Valentin Gold**

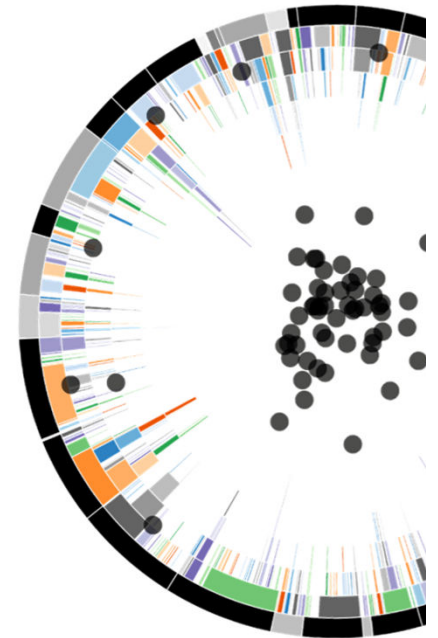


Miriam Butt

Annette Hautli-Janisz

Tina Bögel

Maike Müller



Daniel Keim

Menna El-Assady

Wolfgang Jentner

Rita Sevastjanova

Carmela Acevedo



Visual Analytics

Linguistics

Political science

Goal: To develop an instrument
for the automatic measurement
of the deliberative quality of communication

“ Deliberation is a **communicative process** that aims at **taking a decision** (or recommendation) on collectively **binding rules** or public projects. The substantive goal is to **achieve the common good** and universality of rules. ”



Visual Analytics

Linguistics

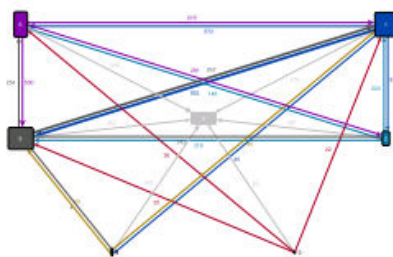
Political science

Goals:

- Deep automatic **syntactic parsing** of the utterances to identify deliberative patterns.
- Define negotiation vs. argumentation through the analysis of the **speaker stands** and **attitude**.
- Identification of different roles of utterances by the usage of **connectors** and **function words**.

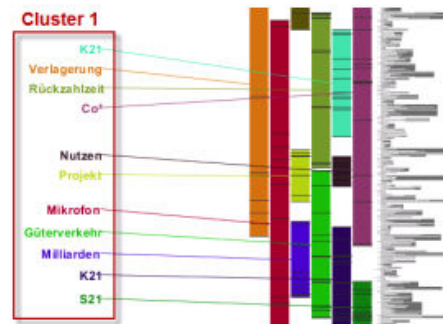
Visual Analytics

Participation



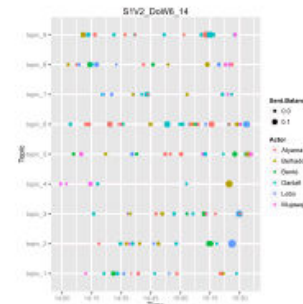
Linguistics

Content

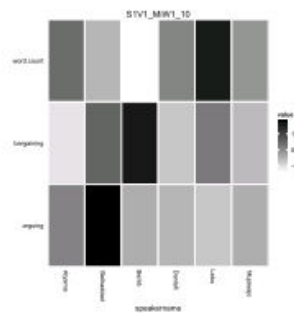


Political science

Sentiments



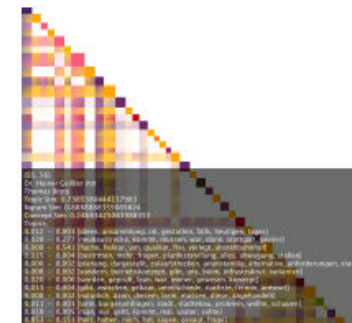
Intentions



Argumentativeness



Interactions



(... and some more!)



Visual Analytics

Linguistics

Political science

Research questions

- What factors make deliberative argumentation successful?
- Can we detect these factors both by shallow statistical and deep linguistic approaches?
- How can visual analytics support the analysis of deliberation in large corpora?



Content Overview using Lexical Episode Plots

First Debate



Second Debate

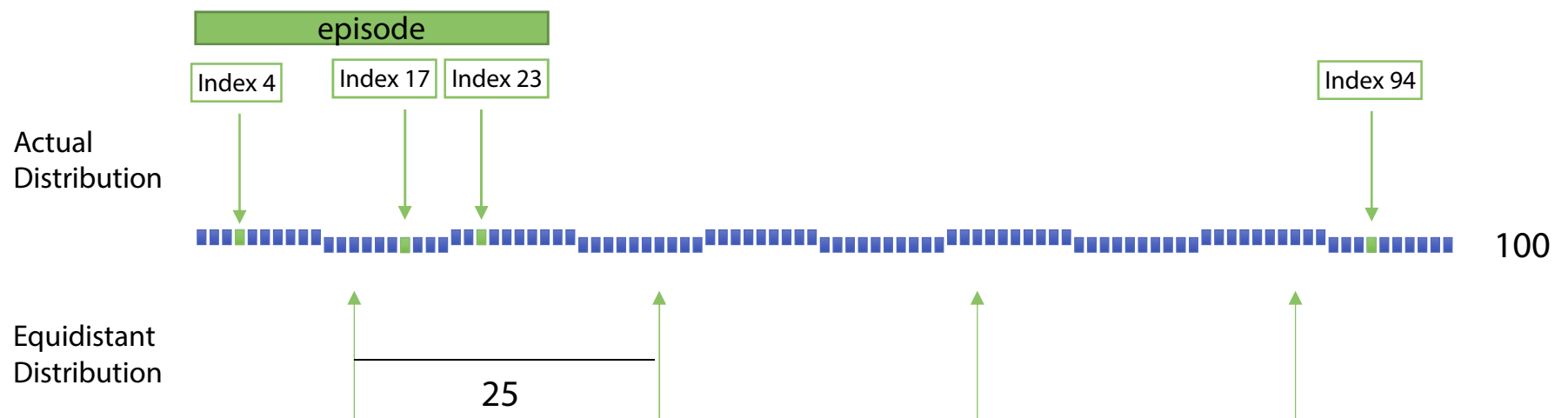


Third Debate



Lexical Episodes

= portion within the word sequence of a corpus where a certain word appears more densely than expected from its frequency in the whole text.

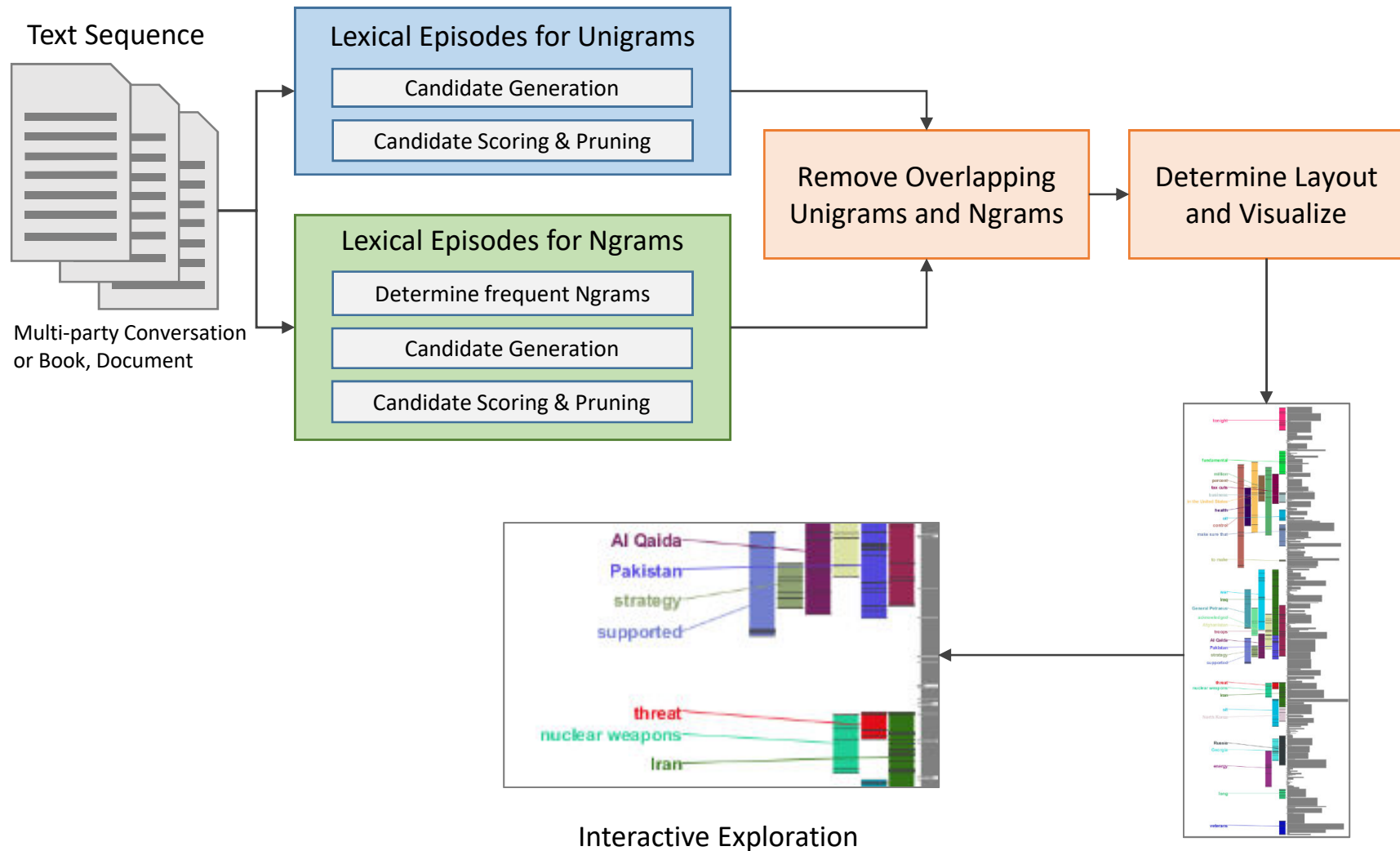


Lexical Episodes Plots

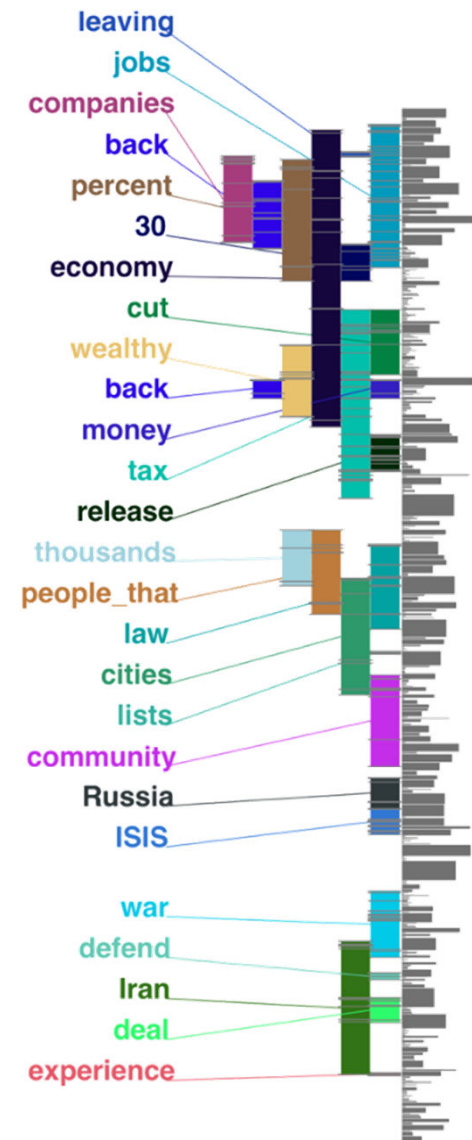


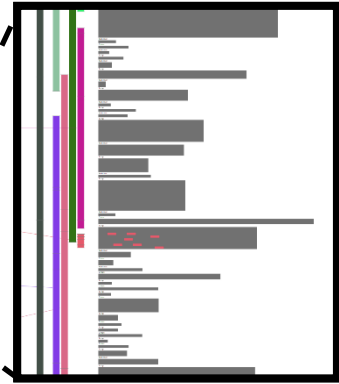
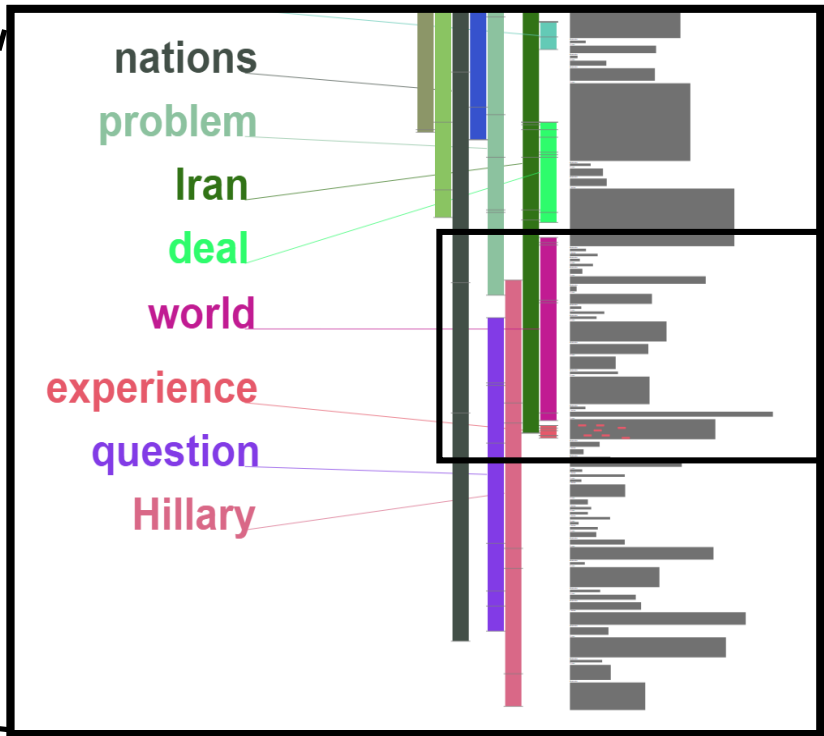
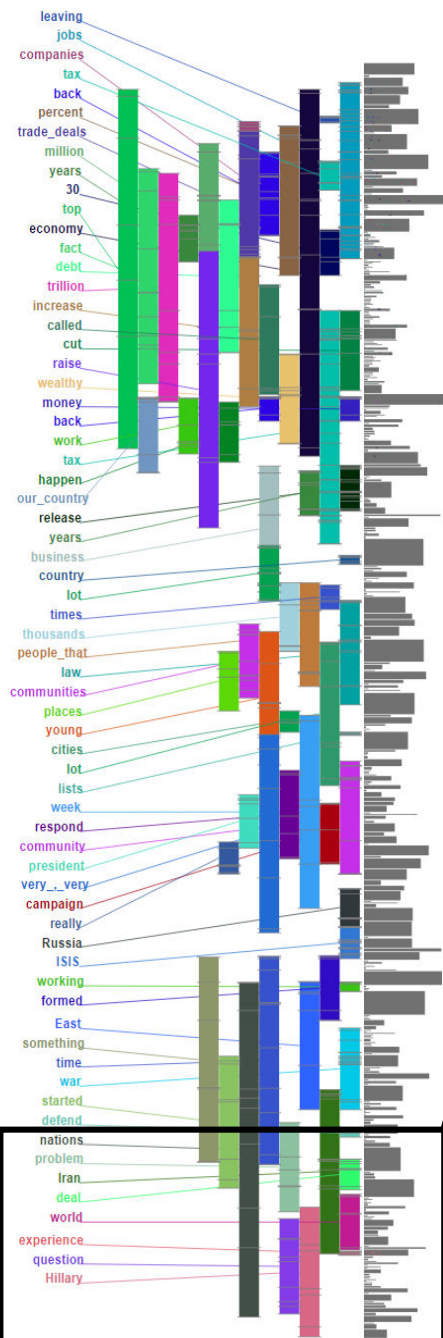
Close Reading

Lexical Episodes - Processing Pipeline



First Debate







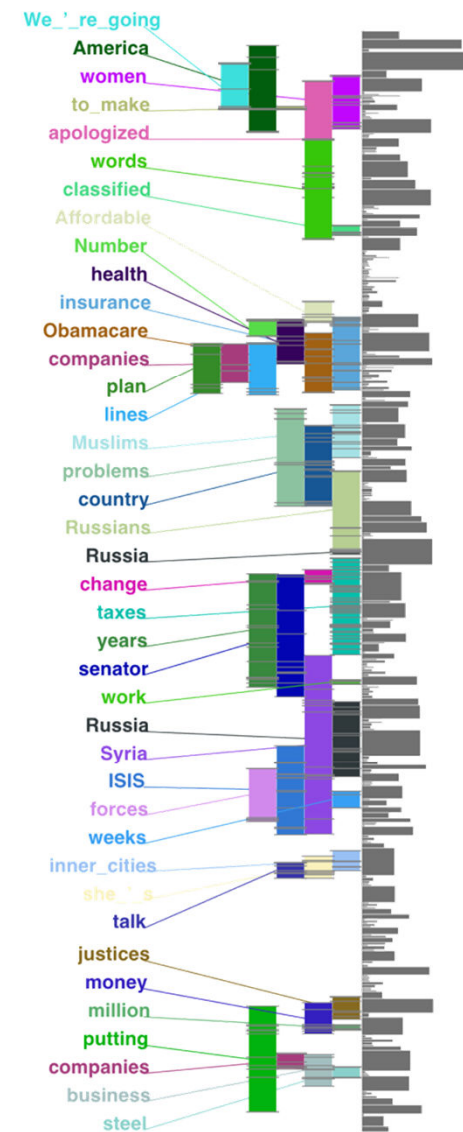
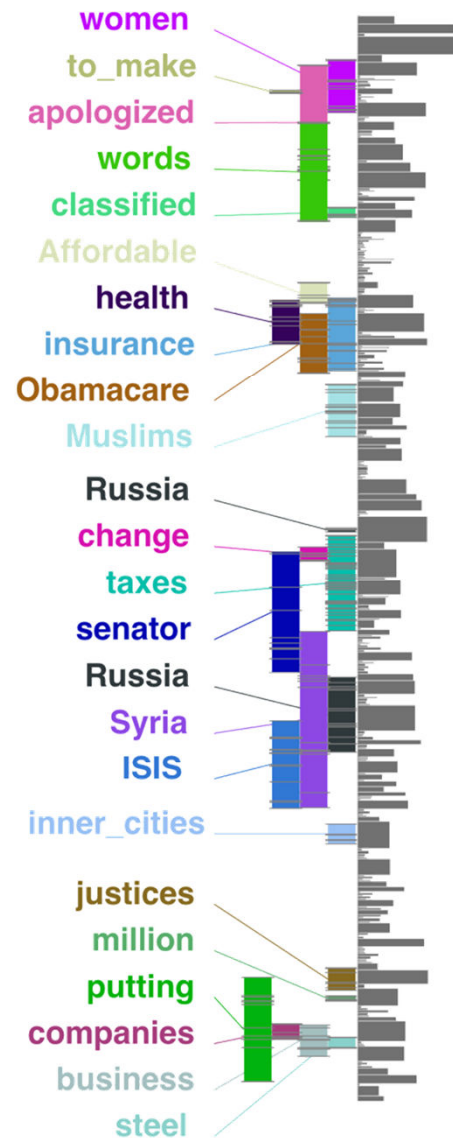
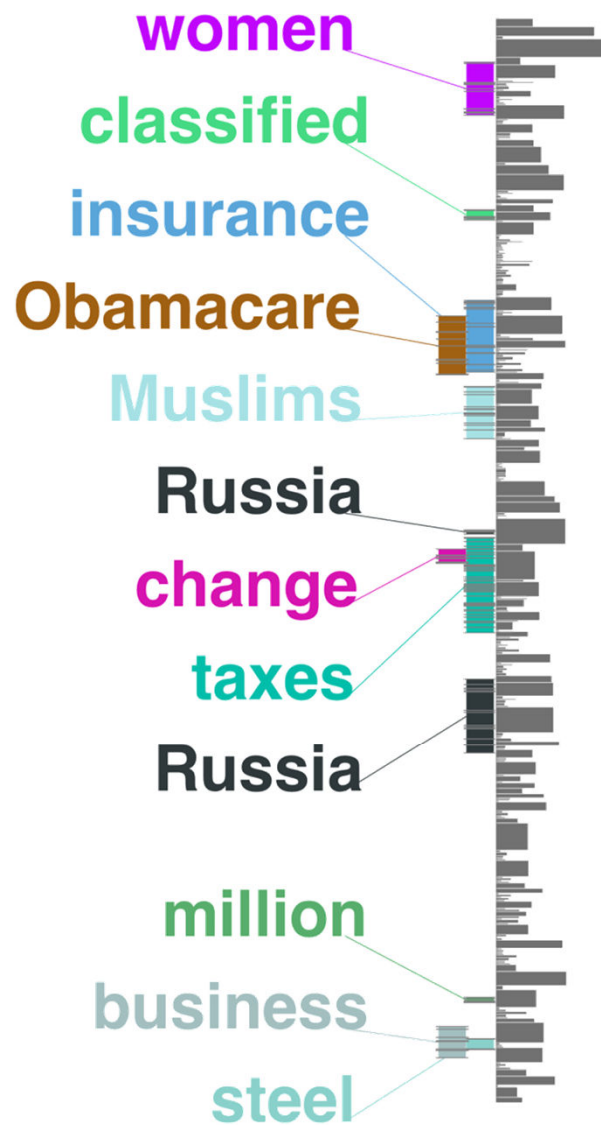
Moderator:
Let's let her respond .

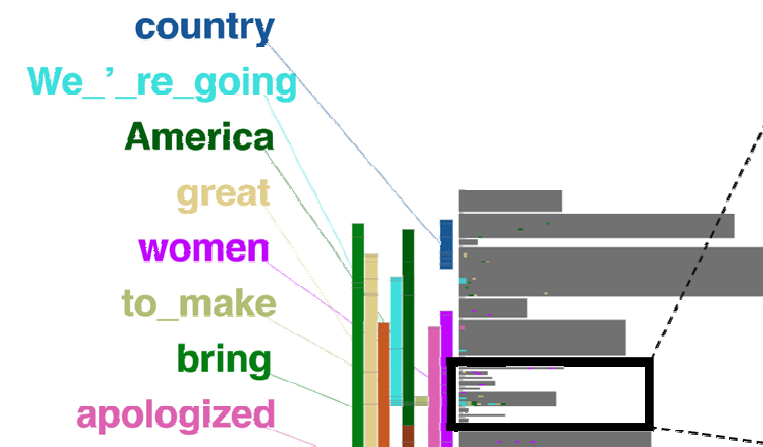
Clinton:
Well , as soon as he travels to 112 countries and negotiates a peace deal , a cease-fire , a release of dissidents , an opening of new opportu
(APPLAUSE)

Trump:
The world — let me tell you .
Let me tell you .
Hillary has **experience** , but it is bad **experience** .
We have made so many bad deals during the last — so she is got **experience** , that I agree .
(APPLAUSE) But it is bad , bad **experience** .
Whether it is the Iran deal that you're so in love with , where we gave them \$150 billion back , whether it is the Iran deal , whether it is anythi
I agree . she is got **experience** , but it is bad **experience** .
And this country can not afford to have another four years of that kind of **experience** .



Second Debate





Moderator:
Just for the record , though , are you saying that what you said on that bus 11 years ago that you did not actually kiss **women** without consent or grope **women** without consent ?

Trump:
I have **great** respect for **women** .
Nobody has more respect for **women** than I do .

Moderator:
So , for the record , you 're saying you never did that ?

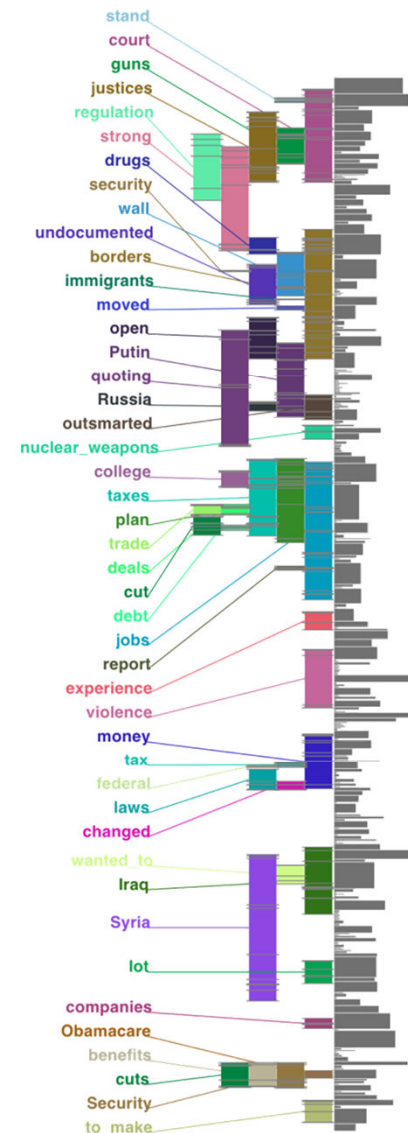
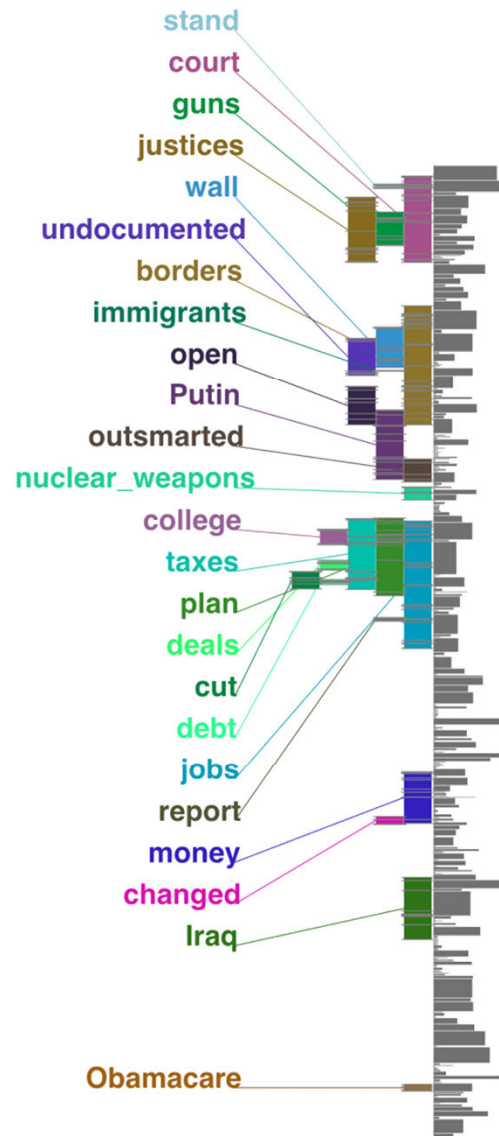
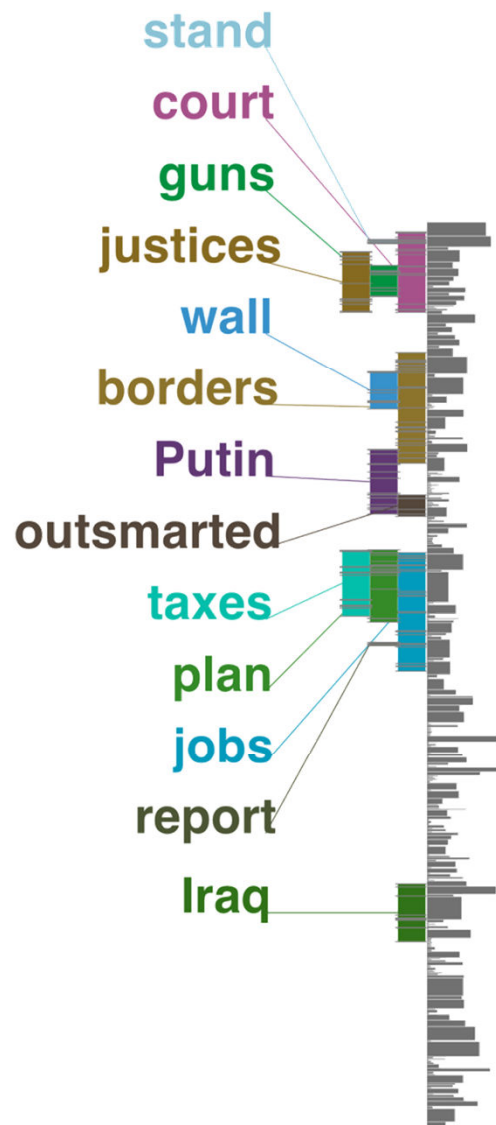
Trump:
I 've said things that , frankly , you hear these things I said .
And I was embarrassed by it .
But I have tremendous respect for **women** .

Moderator:
Have you ever done those things ?

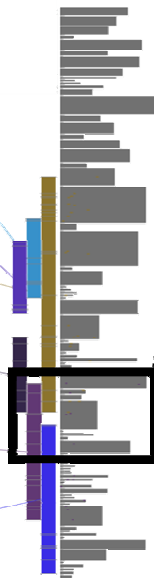
Trump:
And **women** have respect for me .
And I will tell you : No , I have not .
And I will tell you that I 'm going **to make** our country safe .
We 're going to have borders in our country , which we don 't have now .
People are pouring into our country , and they 're coming in from the Middle East and other places .
We 're going to make **America** safe again .
We 're going to make **America** great again , but **we 're going to make** **America** safe again .
And **we 're going to make** **America** wealthy again , because if you don 't do that , it just — it sounds harsh to say , but we have to build up the wealth of our nation .

Moderator:

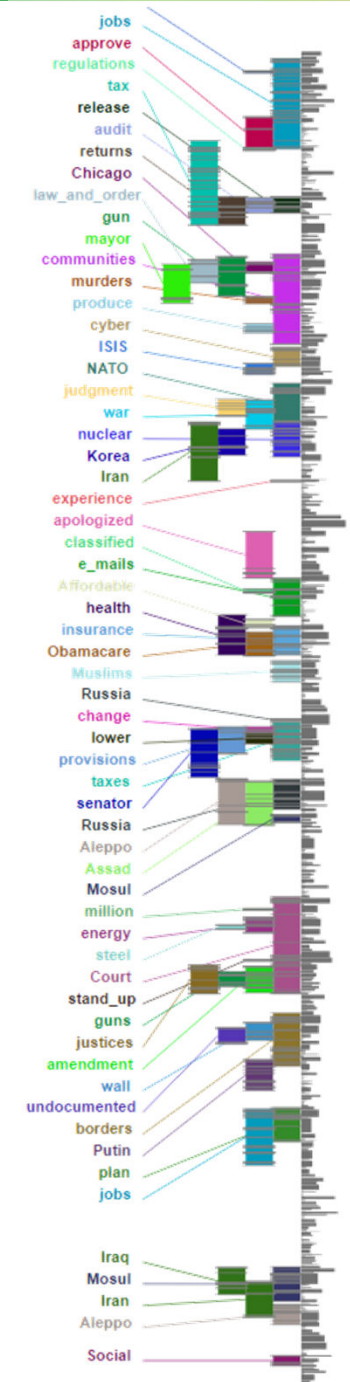
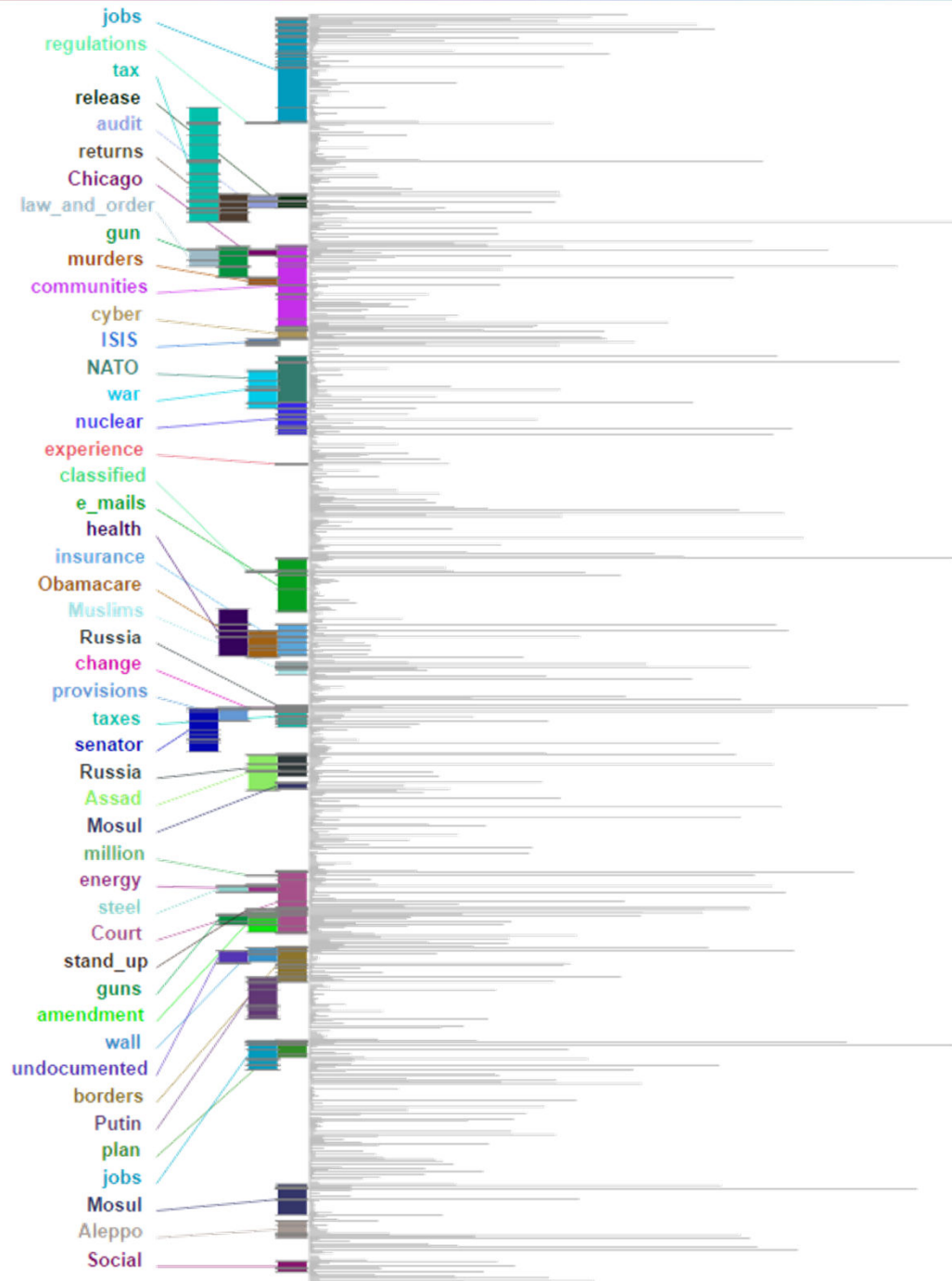
Third Debate



wall
undocumented
borders
open
Putin
nuclear

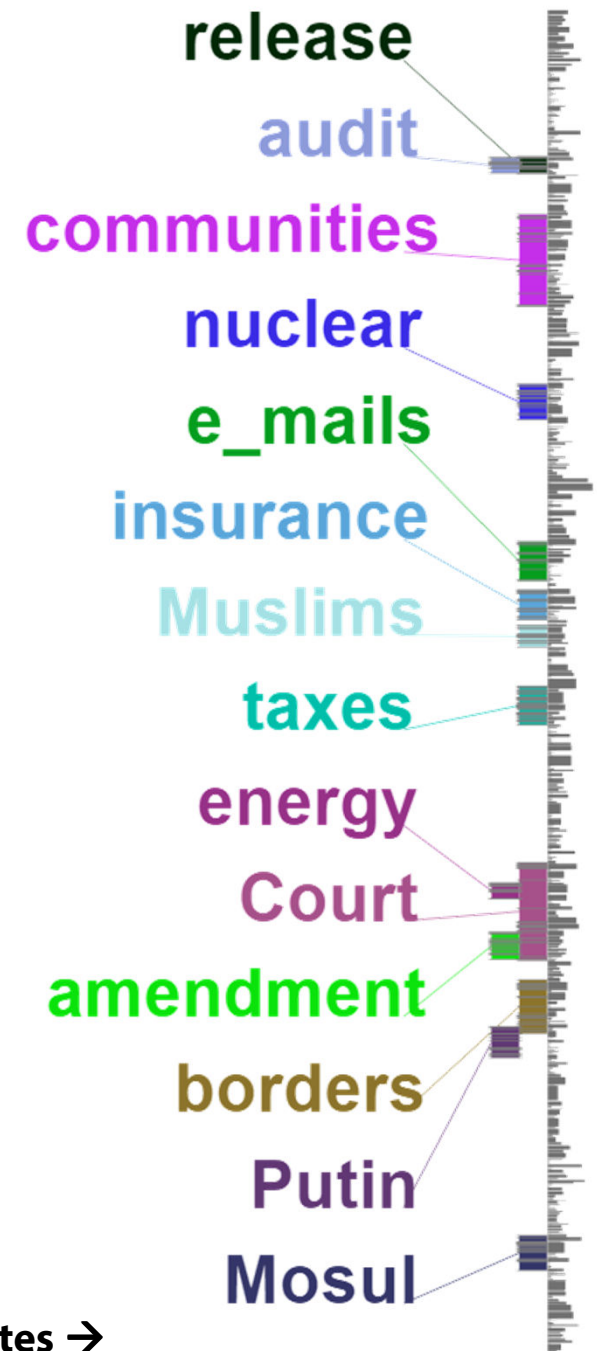


Clinton:
If you went on to read the rest of the sentence , I was talking about energy .
We trade more energy with our neighbors than we trade with the rest of the world combined .
And I do want us to have an electric grid , an energy system that crosses borders .
I think that would be a great benefit to us .
But you are very clearly quoting from WikiLeaks .
What is really important about WikiLeaks is that the Russian government has engaged in espionage against Americans .
They have hacked American websites , American accounts of private people , of institutions .
Then they have given that information to WikiLeaks for the purpose of putting it on the internet .
This has come from the highest levels of the Russian government .
Clearly from [redacted] himself in an effort , as 17 of our intelligence agencies have confirmed , to influence our election .
So I actually think the most important question of this evening , Chris , is finally , will Donald Trump admit and condemn that the Russians are doing this , and make it clear that he will not have the help of [redacted] in this election .
That he rejects Russian espionage against Americans , which he actually encouraged in the past .
Those are the questions we need answered .
We've never had anything like this happen in any of our elections before .
Trump:
That was a great pivot off the fact that she wants open borders .
Okay ?
How did we get on to [redacted] ?
Moderator:
Hold on , folks .
Because this is going to end up getting out of control .
Let's try to keep it quiet .
For the candidates and for the American people .
Trump:
Just to finish on the [redacted] , she wants open borders .
People are going to pour into our country .
People are going to come in from Syria .
She wants 50% more people than Barack Obama .
And he has thousands and thousands of people .
They have no idea where they come from .
And you see , we are going to stop radical Islamic terrorism in this country .
She won't even mention the words nor neither will President Obama .
So I just want to tell you .
She wants open borders .
Now we can talk about [redacted] .
I don't know [redacted] .
He said nice things about me .
If we got along well , that would be good .
If Russia and the United States got along well and went after ISIS , that would be good .
He has no respect for her .
He has no respect for our president .
And I'll tell you what .
We're in very serious trouble .
Because we have a country with tremendous numbers of nuclear warheads , 1,800 , by the way .
Where they expanded and we didn't , 1,800 nuclear warheads .
And she is playing chicken .
Look .
Clinton:

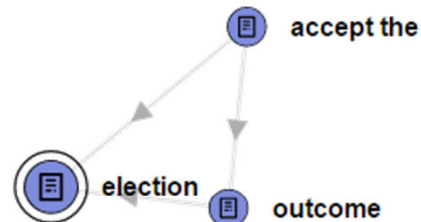


Lexical Episodes

- Generic approach for generating text overviews using lexical chaining
- Text-type and language independent
- Distant Reading vs. Close Reading
- Interactivity through Zooming and Highlighting
- Different levels of detail by steering the significance level



Named-Entity Recognition



Moderator

Good evening from Hofstra University in Hempstead, New York.

I am Lester Holt, anchor of "NBC Nightly News."

I want to welcome you to the first presidential debate.

The participants tonight are Donald Trump and Hillary Clinton.

This debate is sponsored by the Commission on Presidential Debates, a nonpartisan, nonprofit organization.

The commission drafted tonight's format, and the rules have been agreed to by the campaigns.

The 90 minute debate is divided into six segments, each 15 minutes long.

We'll explore three topic areas tonight: Achieving prosperity, America's direction, and securing America.

At the start of each segment, I will ask the same lead-off question to both candidates, and they will each have up to two minutes to respond.

From that point until the end of the segment, we'll have an open discussion.

The questions are mine and have not been shared with the commission or the campaigns.

The audience here in the room has agreed to remain silent so that we can focus on what the candidates are saying.

I will invite you to applaud, however, at this moment, as we welcome the candidates: Democratic nominee for president of the United States Hillary Clinton, and Republican nominee for president of the United States Donald Trump.

Trump

APPLAUSE

Clinton

How are you, Donald?

APPLAUSE

Moderator

Good luck to you.

APPLAUSE Well, I do not expect us to cover all the issues of this campaign tonight, but I remind everyone, there are two more presidential debates scheduled.

We are going to focus on many of the issues that voters tell us are most important, and we are going to press for specifics.

I am honored to have this role, but this evening belongs to the candidates and, just as important, to the American people.

Candidates, we look forward to hearing you articulate your policies and your positions, as well as your visions and your values.

So, let's begin. We are calling this opening segment "Achieving Prosperity."

And central to that is jobs.

There are two economic realities in America today.

There's been a record six straight years of job growth, and new census numbers show incomes have increased at a record rate after years of stagnation.

However, income inequality remains significant, and nearly half of Americans are living paycheck to paycheck.

Beginning with you, Secretary Clinton, why are you a better choice than your opponent to create the kinds of jobs that will put more money into the pockets of American workers?

Clinton

Well, thank you, Lester, and thanks to Hofstra for hosting us.

The central question in this election is really what kind of country we want to be and what kind of future we'll build together.

Today is my granddaughter's second birthday, so I think about this a lot.

First, we have to build an economy that works for everyone, not just those at the top.

That means we need new jobs, good jobs, with rising incomes.

I want us to invest in you.

I want us to invest in your future.

That means jobs in infrastructure, in advanced manufacturing, innovation and technology, clean, renewable energy, and small business, because most of the new jobs will come from small business.

We also have to make the economy fairer.

That starts with raising the national minimum wage and also guaranteeing, finally, equal pay for women's work.

Clinton

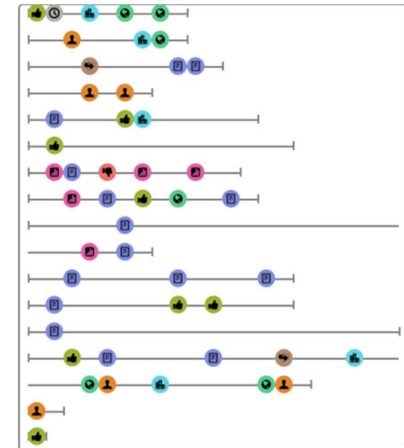
I also want to see more companies do profit-sharing.

If you help create the profits, you should be able to share in them, not just the executives at the top.

And I want us to do more to support people who are struggling to balance family and work.

I've heard from so many of you about the difficult choices you face and the stresses that you're under.

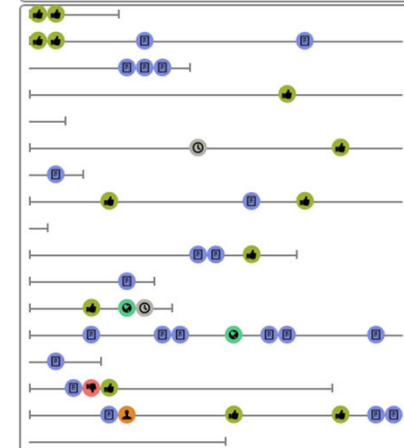
Moderator



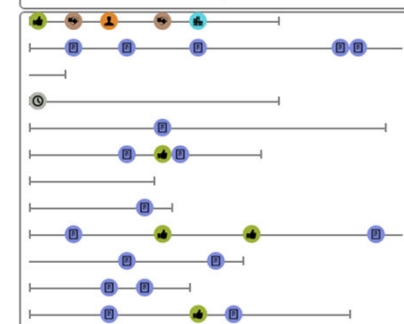
Clinton



Moderator



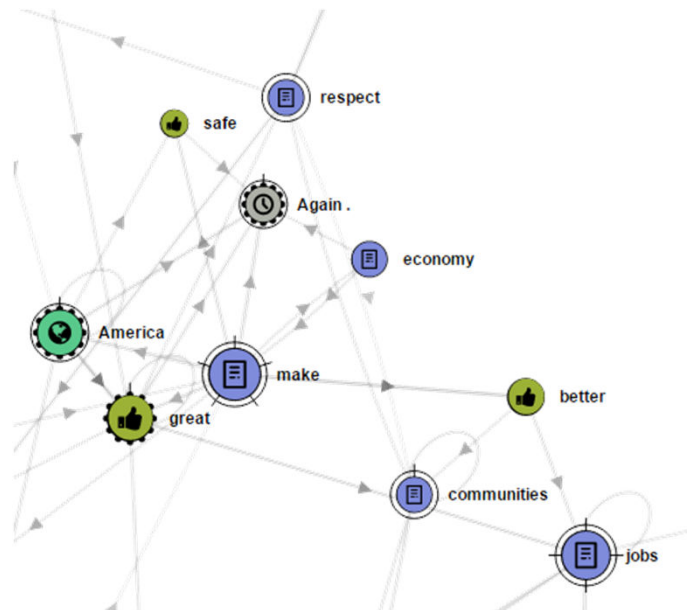
Clinton













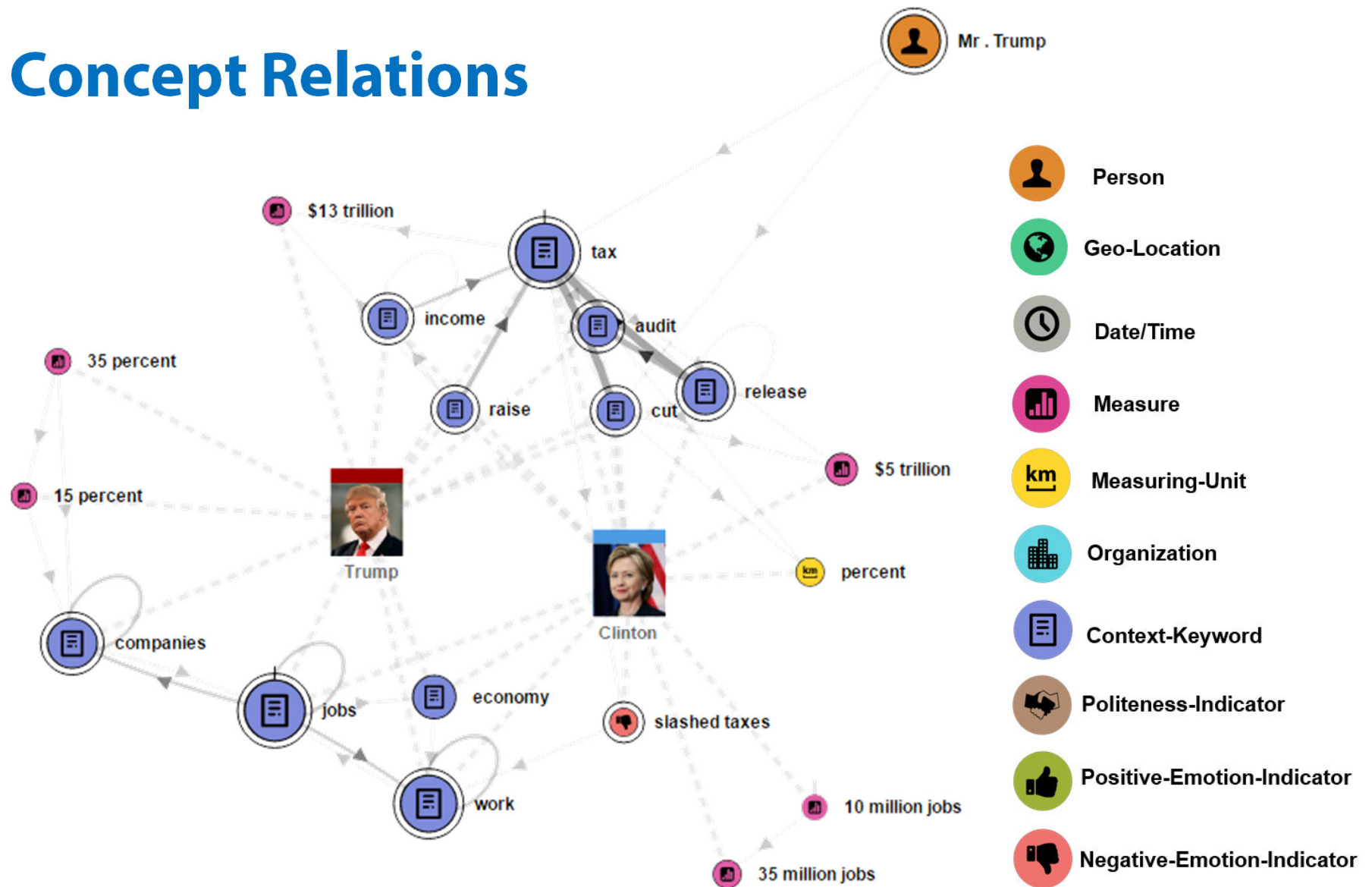
Concept Relations

“I believe that we can cut taxes by an additional \$5 trillion .“

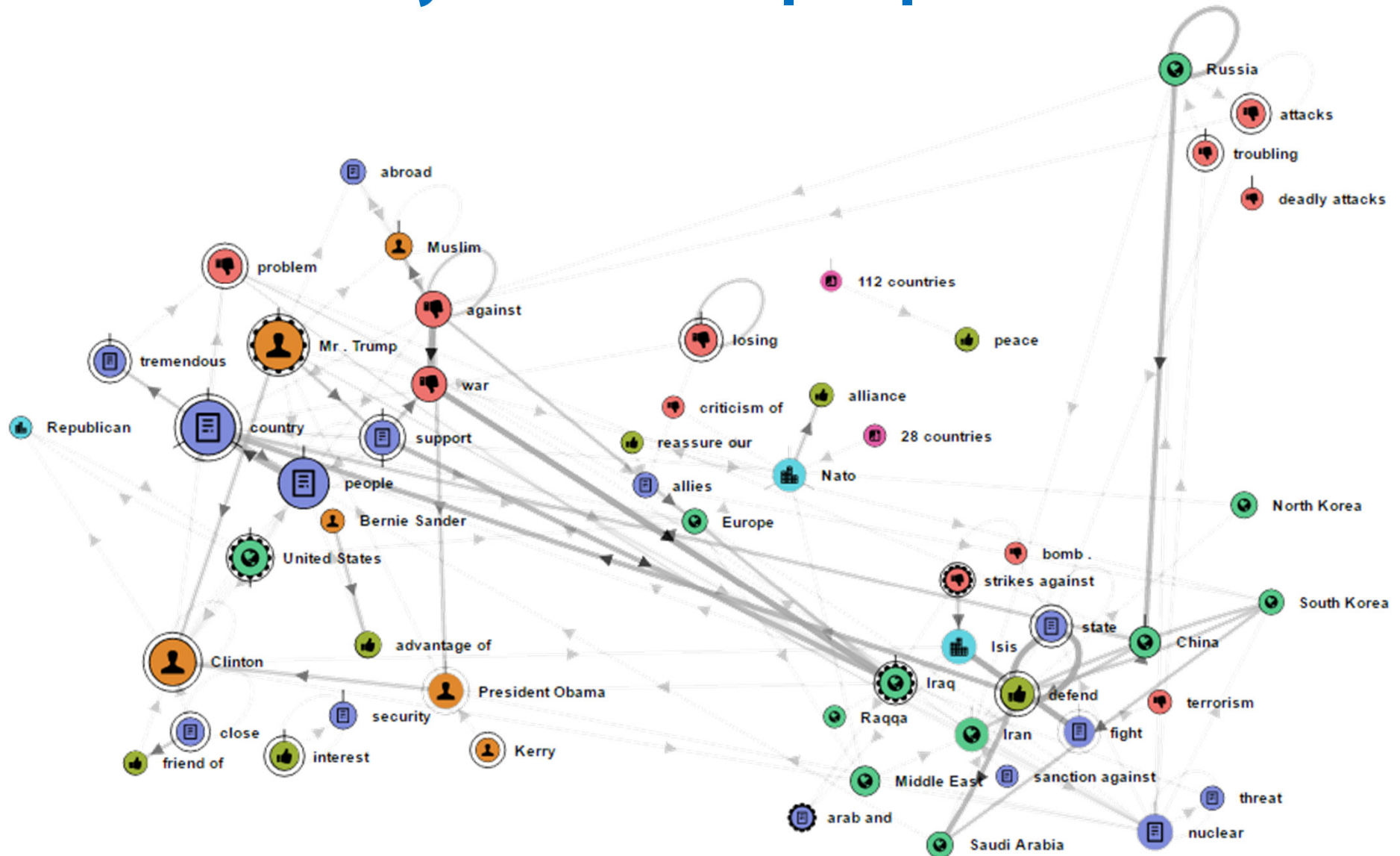
Entity-Pairs:



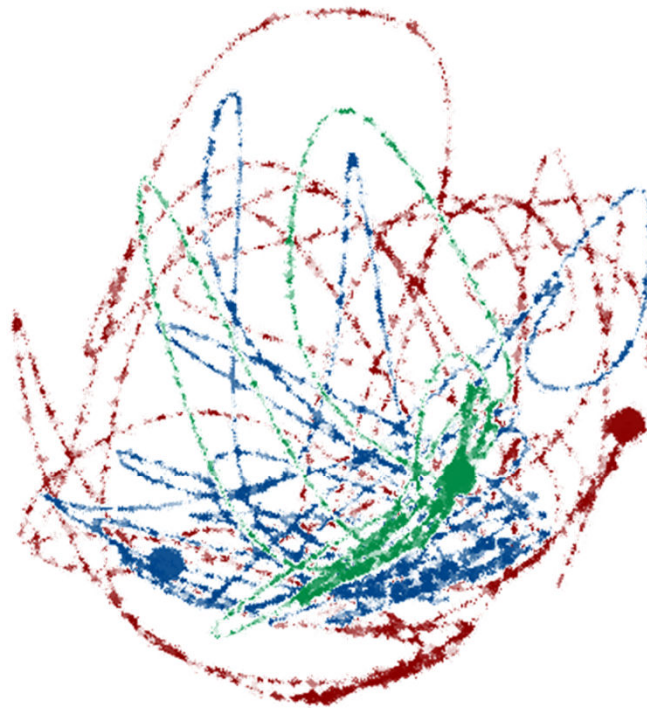
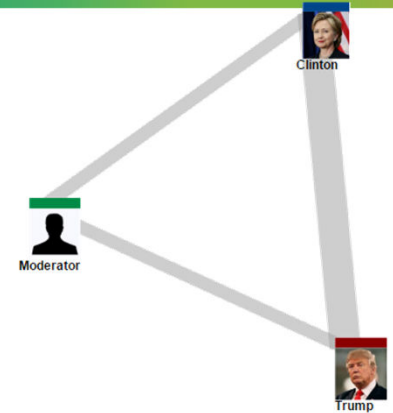
-  Person
-  Geo-Location
-  Date/Time
-  Measure
-  Measuring-Unit
-  Organization
-  Context-Keyword
-  Politeness-Indicator
-  Positive-Emotion-Indicator
-  Negative-Emotion-Indicator



Named-Entity Relationship Exploration



Speaker Interactions

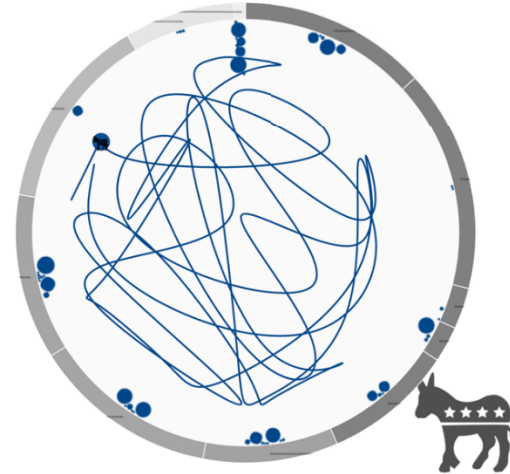
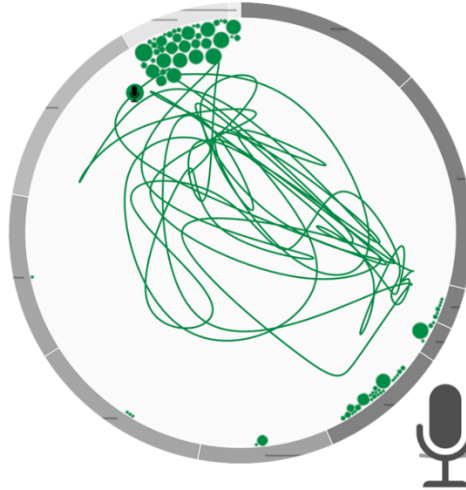
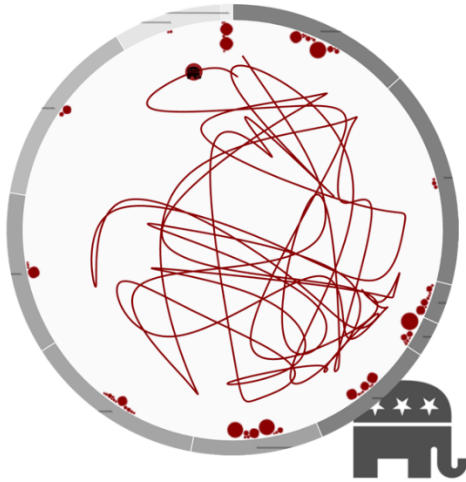


Donald Trump

Hillary Rodham Clinton

How were the speakers moving in the topic space?

Second Debate

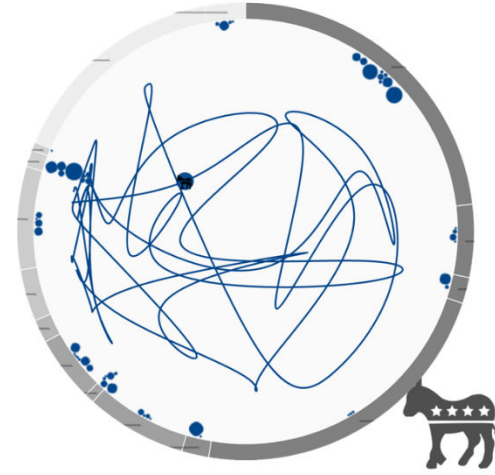
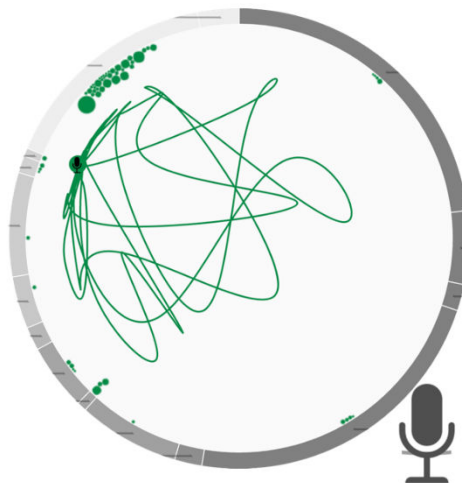
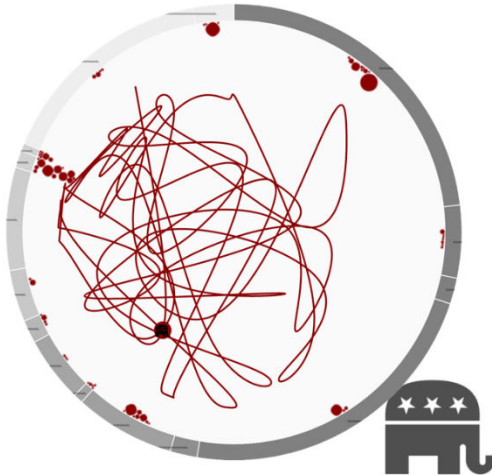


Moderator

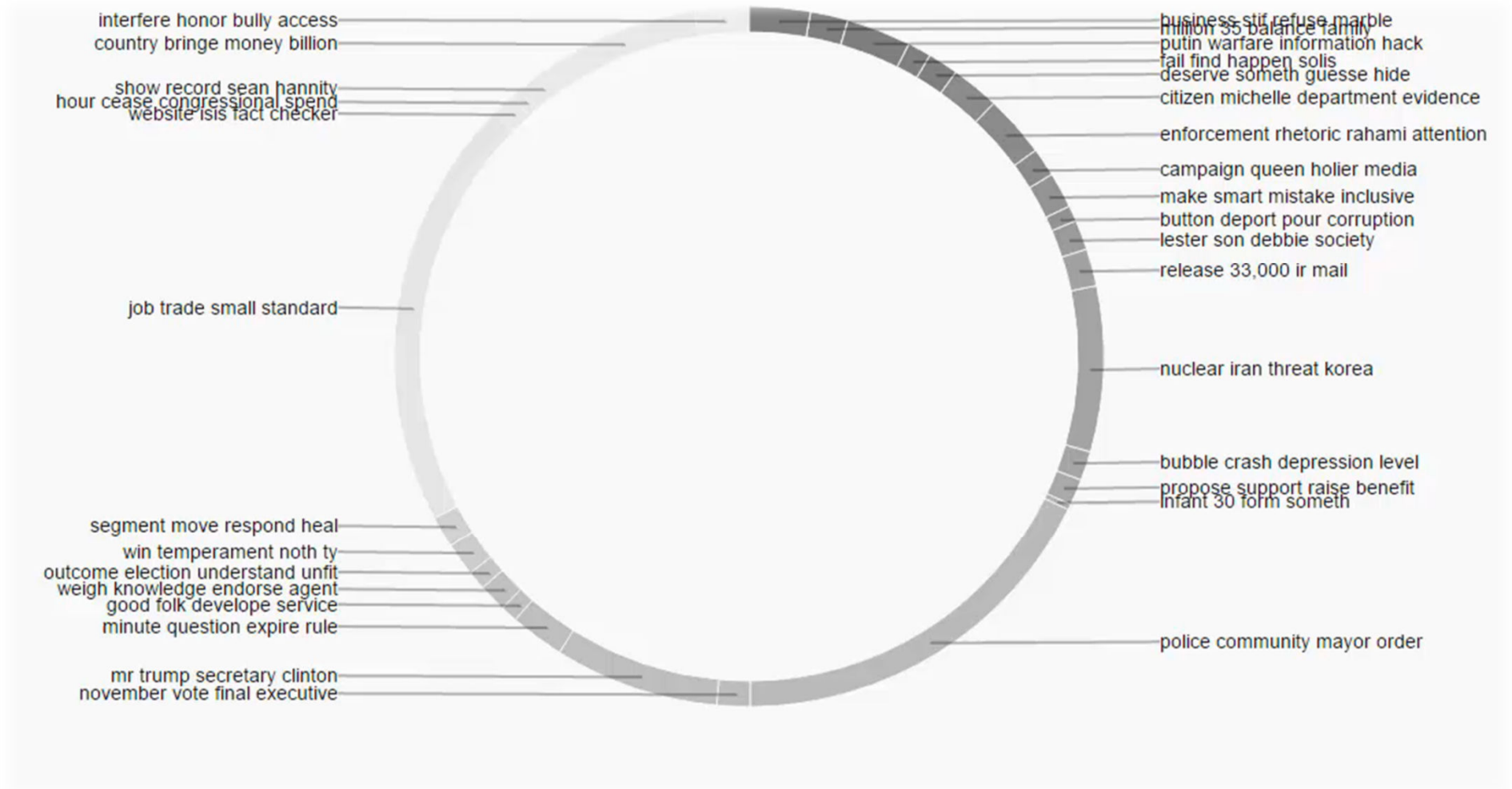
Clinton

Trump

Third Debate



How were the speakers moving in the topic space?





Summary

- Many types of visualization for text data
- Types of NLP applied differ, from counting words to topic modelling and machine learning
- User tasks differ
- Interdisciplinary teams