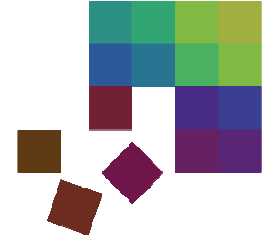


University of Konstanz  
Data Analysis and Visualization Group



# NLP and Visualization for Digital Humanities

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



# Outline

- What is Digital Humanities?
- Problems and Challenges for DH Research
- Using Natural Language Processing to Solve DH Problems
- Exploratory Analysis of Alice in Wonderland with Demos



# Outline

- **What is Digital Humanities?**
- Problems and Challenges for DH Research
- Using Natural Language Processing to Solve DH Problems
- Exploratory Analysis of Alice in Wonderland with Demos

---

# What is Digital Humanities?







# What are the humanities?

The humanities can be described as the study of how people process and document the human experience.

Since humans have been able, we have used philosophy, literature, religion, art, music, history and language to understand and record our world.

These modes of expression have become some of the subjects that traditionally fall under the humanities umbrella.

Knowledge of these records of human experience gives us the opportunity to feel a sense of connection to those who have come before us, as well as to our contemporaries.

<http://shc.stanford.edu/what-are-the-humanities>



# Goals of the Humanities

- To think about the **past** in the **present** in order to take action in the **future**...
- To examine the past critically, thinking about what you wish to emulate, what you wish to avoid and what you wish to do that may be new...
- Establishing shared spaces beyond/ complementing our tribal/ ethnic/ religious identities...

-- Gregory Crane

# What is Digital Humanities?



WIKIPEDIA  
The Free Encyclopedia

Digital humanities (DH) is an area of scholarly activity at the **intersection of computing and the disciplines of the humanities**.

The nature of this activity ranges broadly, from the **practical**, such as digitizing historical texts, to the **philosophical**, such as reflection on the nature of representation itself.

This spectrum of activities is reflected in definitions of the field that range from it being a **collection of methods** to being a **distinct epistemology** and a **kind of science**.

Within this variation, a distinctive feature of digital humanities is its **cultivation of a two-way relationship between the humanities and the digital**: the field both employs technology in the pursuit of humanities research and subjects technology to humanistic questioning and interrogation, often simultaneously.

# What is Digital Humanities?

“Interdisciplinary, collaborative, design-aware, iteration-positive, oriented toward accessibility (in many senses).

Applies **digital technologies alongside more conventional ones** to cultural and social questions & has the potential to enable new questions through the use of those technologies.”

-- Margarita Nafpaktitis

“Work that engages with the **development of digital tools** or uses these same tools within the traditional scope of Humanities.”

-- Henriette Roued-Cunliffe

<http://whatisdigitalhumanities.com/>



# What is Digital Humanities?

You might be a digital humanist if:

- You use emerging technologies to rethink what "the humanities" is all about.
- You collaborate with people across disciplines and backgrounds to explore innovations in academia
- You are willing to experiment with new pedagogies and research approaches made possible by new technologies
- You are not afraid to work closely and collaboratively with your friendly neighborhood technologist
- You are willing to take a do it yourself approach to creating the tools and technologies you need for your teaching and/research
- You use technology in a way that disregards traditional boundaries between disciplines and hierarchies

-- Patrick Murray-John

# Collaborations between CS and...

- Disciplines
  - Classics
  - History
  - Linguistics
  - Literatures, Cultures & Languages
  - Music
  - Philosophy
  - Religious Studies
  - Theater & Performance Studies
- Associated
  - Journalism
  - Political Science
  - Legal Studies

philosophy | religion | communication | new  
history | civility | language | sappho | writing  
literature | sociology | rhetoric | the good

## HUMANITIES

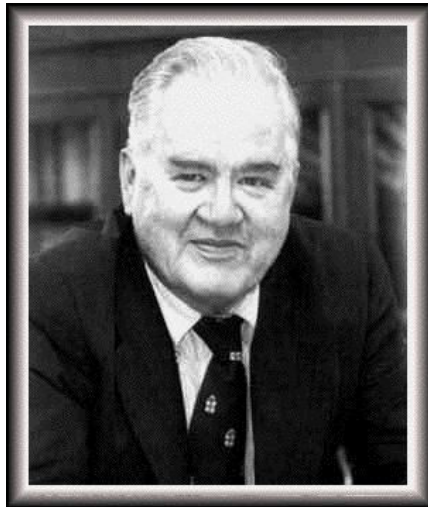
culture | ethics | environment | myth | beauty  
sustainability | education | linguistics | arts  
media | lucretius | sciences | imagination

# Digital Humanities Collaborations

- Interdisciplinary research is very challenging
- It is crucial to have an open communication
  - ... about the used concepts and terminology
  - ... and the process of tackling the overall research question
- Divide and conquer: Smaller self-contained tasks are easier to solve
- The research question has to be:
  - clear (allows only one interpretation)
  - focused (concentrates on specific aspects)
  - concise (based on rigorous conceptualization)
  - complex (includes multiple layers)
  - arguable (addresses a research gap)



# Digital Humanities Collaborations



JOHN W. TUKEY\*

We often forget how science and engineering function. Ideas come from previous exploration more often than from lightning strokes. Important questions can demand the most careful planning for confirmatory analysis. Broad general inquiries are also important. Finding the question is often more important than finding the answer. Exploratory data analysis is an attitude, a flexibility, and a reliance on display, NOT a bundle of techniques, and should be so taught. Confirmatory data analysis, by contrast, is easier to teach and easier to computerize. We need to teach both; to think about science and engineering more broadly; to be prepared to randomize and avoid multiplicity.





# Outline

- What is Digital Humanities?
- **Problems and Challenges for DH Research**
- Using Natural Language Processing to Solve DH Problems
- Exploratory Analysis of Alice in Wonderland with Demos

# Unify Close and Distant Reading

The screenshot displays the VarifocalReader interface, which integrates distant reading (table of contents) and close reading (text view).

**Table of Contents (Left Panel):**

- Chapter 1: Gleichmut Selbstigkeit Achill Besinnung** (Page 121)
  - Zwerchfell Helle Licht Meer Achill
  - Briseis Eros Hunde Seelisches Unruhe
  - Vorrecht Nacht Menelaos Reichtum
  - Spittaler Bildlichkeit Hera Lautmalerei
- Chapter 2: Handlungen Lessing Endzweck** (Page 122)
  - Nacheinander Nebeneinander
  - Beschreibung Bräutungen Flammen
  - Huft Laokoon Lykurgos Malerei
  - Myrmidonen Schrein Stänze Sterblicher
- Chapter 3: Ilias Kasus Schwerpunkt Haus** (Page 123)
  - Aias Organismus Addition Birnen
  - Pfeilen Geschlossenheit Graun
  - Ruhen Stift Gleichnis Odyssee Homer
  - Teile von Hexameter Erz Fliegen R
- Chapter 4: Schrift Zeus Ilias Orakel** (Page 124)
  - Menschen Funktionen Gesamtplan
  - Geschicke Gesinnung Gutes Milde
  - Seher Verpflichtung Weitblick
  - Homer Menelaos Troia Jährlin Hektor
  - Heiders Jahre Odysseus Moers Tapferkeit
- Chapter 5: Homer Europas Technik** (Page 125)
  - Xenophanes Epos Fuchs Moral
  - von Volk Spittaler Christenrum Dachs
  - Gattungsgesetz Geschichten Hesiod
  - Idyll Idylle Inferno Mythen Ochsen
  - Paradies Stiftung Vortrags Epischen
- Chapter 6: Pathos Rampe Redner** (Page 126)
  - Gewalt Bestehende
  - Rome Schmerzes

**Text View (Right Panel):**

The text view shows a passage from a German text, with annotations 'g' and 'h' marking specific sentences.

**Annotation 'g':** Unter dem »Todes« haben wir den Anfang, die Mitte, das Ende, Gestirge und römische Verse des Epos verstanden. Ihre Selbstlosigkeit ist aber nur möglich und sinnvoll, wenn auch die Teile des dargestellten Lebens selbstständig sind. Gerade darin zeigt sich nun die einzigartige Kraft Homers.

**Annotation 'h':** In diesem Versen scheint sich jedoch ein alterer Mythos erhalten zu haben, die Spur einer ungeheuren Welt.

Koch et al., VarifocalReader — In-Depth Visual Analysis of Large Text Documents

# “Prosthetic Reading” vs. “Slow Analysis”

Sonnet 129  
William Shakespeare

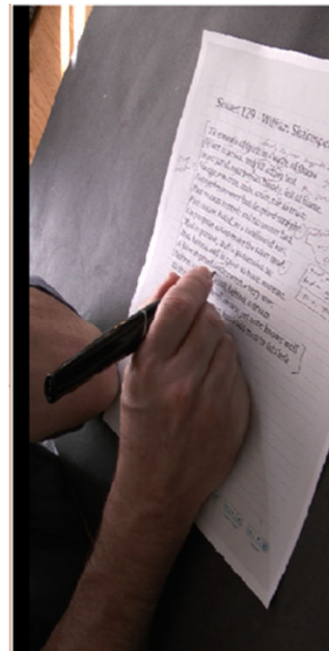
Th' expense of spirit in a waste of shame  
Is lust in action; and till action, lust  
Is perjured, murderous, bloody, full of blame,  
Savage, extreme, rude, cruel, not to trust;  
Enjoyed no sooner but despised straight;  
Past reason hunted; and no sooner had,  
Past reason hated, as a swallowed bait,  
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.

**Alliteration**

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**



who chose his epitaph?—himself alone  
Could thus have dared the grave to agitate,  
And claim, among the dead, this awful crown;  
Nor doubt that He marked also for his own  
Close to these cloistral steps a burial-place,  
That every foot might fall with heavier tread.

**A** 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.

**B** 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.

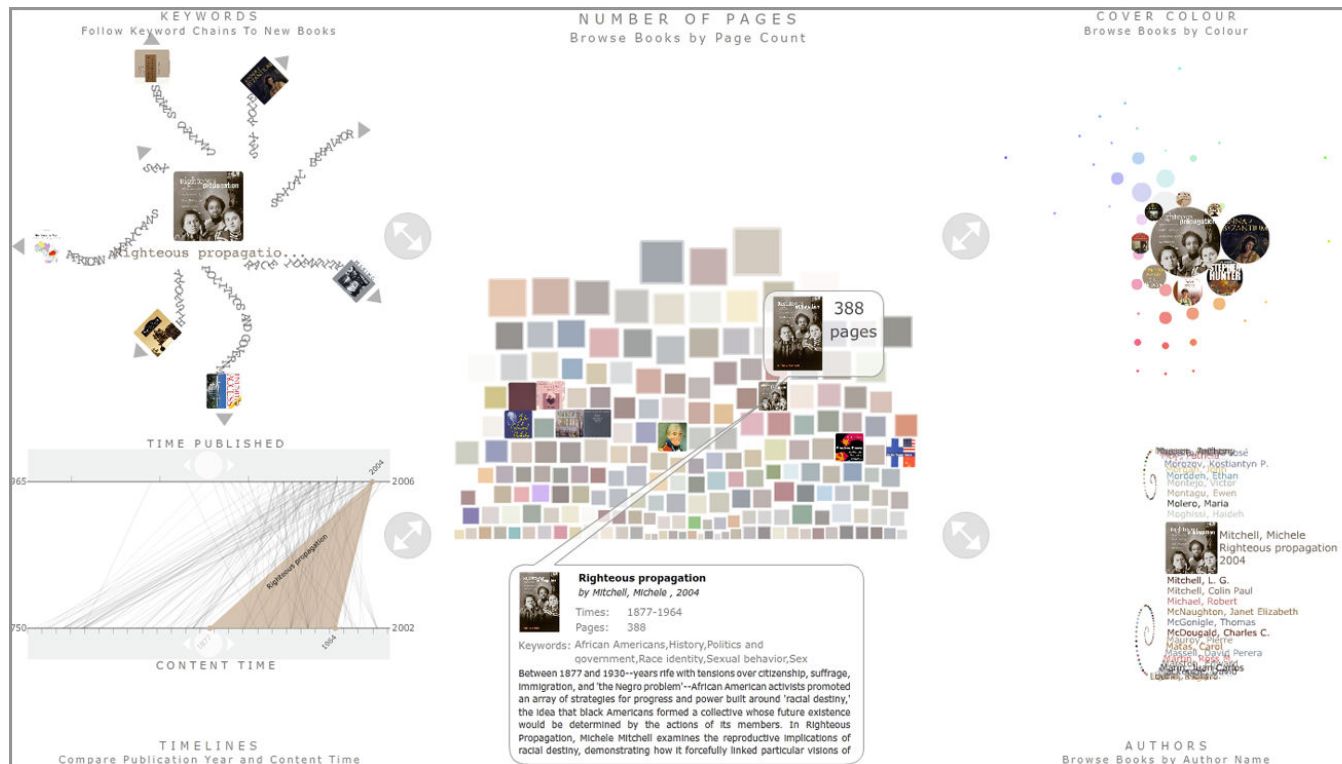
**C** not one a fighter; It is like a plane  
dusting apple orchards of Arabs or  
one of the mighty... one of the hell

**D** Perhaps you consider yourself an oracle,  
Mouthpiece of the dead, or of some god or  
Thirty years now I have labored  
To dredge the silt from your throat:  
I am none the wiser.  
Scaling little ladders with gluepots and pa  
I crawl like an ant in mourning  
Over the weedy acres of your brow  
To mend the immense skull plates and cle  
The bald, white tumult of your eyes.  
A blue sky out of the (Crestia)  
Arches above us, O father, all by yourself  
You are with us, and I am not with you

**E** On thy withered lips and dry,  
Which like barren furrows lie,  
Brooding kisses I will pour  
Shall thy youthful (heat) restore  
(Such kind showers in autumn fall,  
And a second spring recall)  
Nor from thee will ever part,  
Ancient person of my heart.  
Thy nobler part, which but to name  
In our sex would be counted shame  
By age's frozen grasp possessed,  
From [his] face shall be released,  
And soothed by my reviving hand,  
In former warmth and vigor stand.  
All a lover's tribute due to thee

Mehta et al., MetaTation: Augmenting Free-Form Annotations to Bridge Close and Distant Reading of Poetry

## Supporting Serendipitous and Guided Discovery



Thudt et al., The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries through Information Visualization

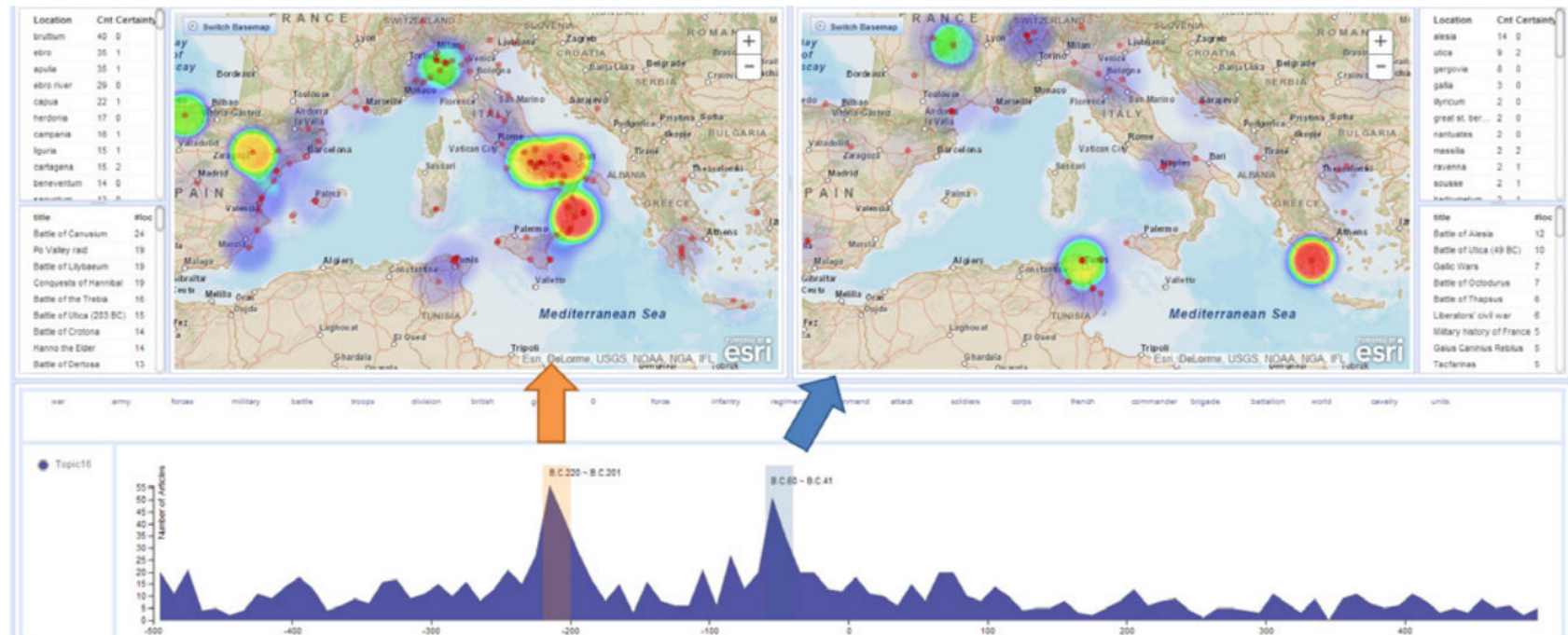


# Supporting Serendipitous and Guided Discovery



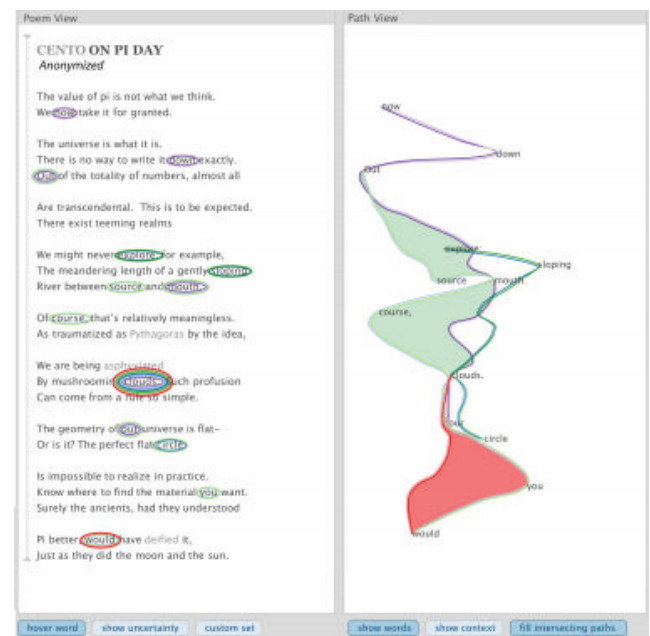
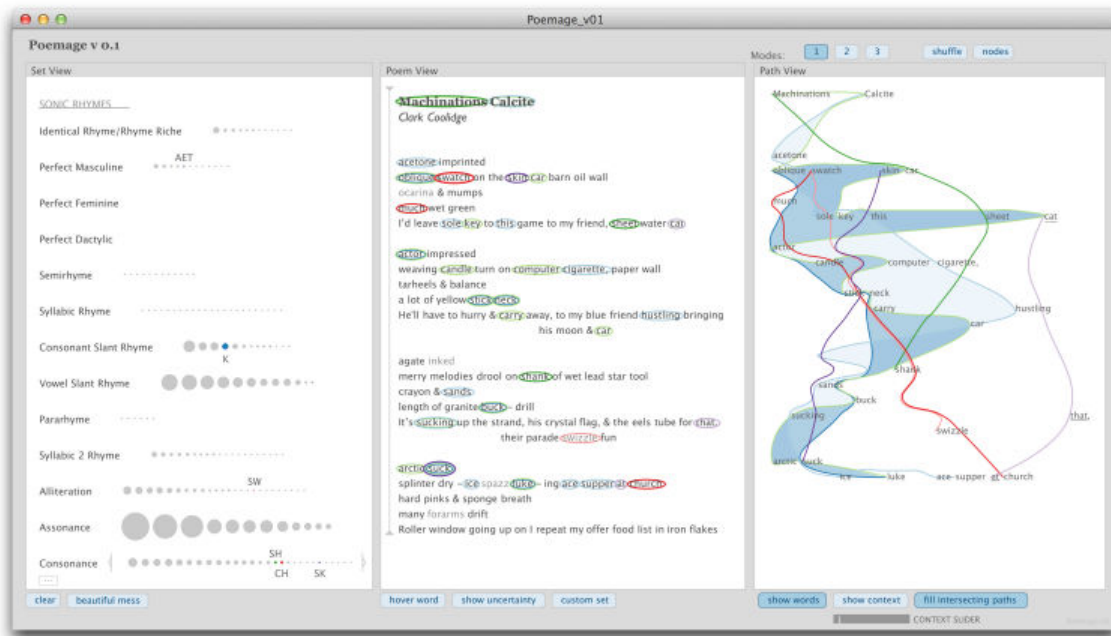
Alexander et al., Serendip: Topic Model-Driven Visual Exploration of Text Corpora

# Linking Outside Resources into the Analysis



Cho et al., VAIroma: A Visual Analytics System for Making Sense of Places, Times, and Events in Roman History

# Aesthetic Experiences



McCurdy et al., Poemage: Visualizing the Sonic Topology of a Poem

# Annotation of Resources

CHAPTER 2. I OBSERVE

Audio Narration of Chapter 2

David seems to remember alot from his childhood stay at various places

Seems to be thinking about a fantasy from past. May be has fear from animals (dogs, crow etc)??

Pew at church - from book

CHAPTER 2. I OBSERVE

The first objects that assume a distinct presence before me, as I look far back, into the blank of my infancy, are my mother with her pretty hair and youthful shape, and Peggoty with no shape at all, and eyes so dark that they seemed to darken their whole neighbourhood in her face, and cheeks and arms so hard and red that I wondered the birds didn't peck her in preference to apples.

I believe I can remember these two at a little distance apart, dwarfed by my sight by stooping down or kneeling on the floor, and I going unsteadily from the one to the other. I have an impression on my mind which I cannot distinguish from actual remembrance, of the touch of Peggoty's forefinger as she used to hold it out to me, and of its being roughened by needlework, like a pocket nutmeg grater.

This may be fancy, though I think the **memory** of most of us can go farther back into such times than many of us suppose; just as I believe the power of **observation** in numbers of very young children to be quite wonderful for its closeness and accuracy. Indeed, I think that most grown men who are remarkable in this respect, may with greater propriety be said not to have lost the faculty, than to have acquired it; the rather, as I generally observe such men to retain a certain freshness, and gentleness, and capacity of being pleased, which are also an inheritance they have preserved from their **childhood**.

I might have a misgiving that I am 'meandering' in stopping to say this, but that it brings me to remark that I build these conclusions, in part upon my own experience of myself; and if it should appear from anything I may set down in this narrative that I was a child of close observation, or that as a man I have a strong memory of my childhood, I undoubtedly lay claim to both of these characteristics.

Looking back, as I was saying, into the blank of my infancy, the first objects I can **remember** as standing out by themselves from a confusion of things, are my mother and Peggoty. What else do I remember? Let me see.

There comes out of this cloud, our house - just new to me; but quite familiar; in its earliest remembrance. On the ground-floor is Peggoty's kitchen, opening into a back yard; with a pigeon-house on a pole; in the centre, without any pigeons in it; a great dog-kennel in a corner; without any dog; and a quantity of fowls that look terribly tall to me; walking about; in a menacing and ferocious manner. There is one cock who gets upon a post to crow, and seems to take particular notice of me as I look at him through the kitchen window, who makes me shiver, he is so fierce. Of the geese outside the side-gate who come waddling after me with their long necks stretched out when I go that way, I dream at night; as a man exulting in wild beasts might dream of lions.

Here is a long passage - what an enormous perspective I make of it! - leading from Peggoty's kitchen to the front door. A dark store-room opens out of it, and that is a place to be run past at night; for I don't know what may be among those tubs and jars and old tea-chests, when there is nobody in there with a dimly-burning light, letting a mouldy air come out of the door. In which there is the smell of soap, pickles, pepper, candles, and coffee, all at one whiff. Then there are the two parlours: the parlour in which we sit of an evening, my mother and I and Peggoty - for Peggoty is quite our companion, when her work is done and we are alone - and the best parlour where we sit on a Sunday; grandly, but not so comfortably. There is something of a diabolical air about that room to me, for Peggoty has told me - I don't know when, but apparently ages ago - about my father's funeral, and the company having their black cloaks put on. One Sunday night my mother reads to Peggoty and me in there, how Lazarus was raised up from the dead. And I am so frightened that they are afterwards obliged to take me out of bed, and show me the quiet churchyard out of the bedroom window, with the dead all lying in their graves at rest, below the solemn moon.

There is nothing half so green that I know anywhere, as the grass of that churchyard; nothing half so shady as its trees; nothing half so quiet as its tombs ones. The sheep are feeding there, when I kneel up, early in the morning, in my little bed in a closet within my mother's room, to look out at it, and I see the red light shining on the sun-dial, and think within myself, 'Is the sun-dial glad, I wonder, that it can tell the time again?'

**Here is our pew in the church.** What a high-backed pew! With a window near it, out of which our house can be seen, and is seen many times during the morning's service, by Peggoty, who likes to make herself as sure as she can that it's not being robbed, or is not in flames. But though Peggoty's eyes wander, she is much offended if mine does, and grows to me, as I stand upon the seat, that I am to look at the clergyman. But I can't always look at him - I know him without that white thing on, and I am afraid of his wondering why I stare so, and perhaps stopping the service to inquire - and what am I to do? It's a dreadful thing to gaze, but I must do something. I look at my mother, but she pretends not to see me. I look at a boy in the aisle, and he makes faces at me. I look at the sunlight coming in at the open door through the porch, and there I see a stray sheep - I don't mean a sinner, but mutism - half making up his mind to come into the church. I feel that if I looked at him any longer, I might be tempted to say something out loud; and what would become of me then! I look up at the monumental tablets on the wall, and try to think of Mr. Bodgers late of this parish, and what the feelings of Mrs. Bodgers must have been, when affliction sore, long time Mr. Bodgers here, and physicians were in vain. I wonder whether they called in Mr. Chillis, and he was in vain; and if so, how he likes to be reminded of it once a week. I look from Mr. Chillis, in his Sunday neckties, to the pulpit; and think what a good place it would be to play in, and what a castle it would make, with another boy coming up the stairs to attack it, and having the velvet cushion with the tassels thrown down on his head. In time my eyes gradually shut up; and, from seeming to hear the clergyman singing a drowsy song in the heat, I hear nothing, until I fall off the seat with a crash, and am taken out, more dead than alive, by Peggoty.

Are memory, observe and remember used in same context by the author? Let's analyze with TagPies!

Chapter 2

David childhood (Chapter 1-10)

Whole book

Western Roman Empire

Eastern Roman Empire

Roma

Peggoty seems to be gentle and caring, opening herself and her family to David whenever he is in need

Seems to be thinking about a fantasy from past. May be has fear from animals (dogs, crow etc)??

In the second century of the **Christian Era**, the **empire of Rome** comprehended the fairest part of the earth, and the most civilized portion of mankind. The frontiers of that extensive monarchy were guarded by ancient renown and disciplined valor. The gentle but powerful influence of laws and manners had gradually cemented the union of the provinces. Their peaceful inhabitants enjoyed and abused the advantages of wealth and luxury. The image of a free constitution was preserved with decent reverence: the Roman senate appeared to possess the sovereign authority, and devolved on the emperors all the executive powers of government. During a happy period of more than fourscore years, the public administration was conducted by the virtue and abilities of **Nerva, Trajan, Hadrian**, and the two Antonines. It is the design

How Christian Era overlays with other religions

Faisal Cheema et al., AnnotateVis: Combining Traditional Close Reading with Visual Text Analysis





# Outline

- What is Digital Humanities?
- Problems and Challenges for DH Research
- Using Natural Language Processing to Solve DH Problems
- Exploratory Analysis of Alice in Wonderland with Demos



# NLP for DH

- Word counting
- POS tagging
- Stop word removal
- Keyword extraction
- Named entity recognition
- Lexical distribution
- Clustering documents in a corpus
- Comparing documents / document with corpus
- Word sense disambiguation
- Parsing
- Topic modelling
- Edit distance (variation)
- Pronunciation / sonic patterns
- Extracting sentiment and other features
- ...

# NLP for DH

- Word counting
- POS tagging
- Stop word removal
- Keyword extraction
- Named entity recognition
- Lexical distribution
- Clustering documents in a corpus
- Comparing documents / document with corpus
- Word sense disambiguation
- Parsing
- Topic modelling
- Edit distance (variation)
- Pronunciation / sonic patterns
- Extracting sentiment and other features
- ...



# Outline

- What is Digital Humanities?
- Problems and Challenges for DH Research
- Using Natural Language Processing to Solve DH Problems
- Exploratory Analysis of Alice in Wonderland with Demos

# Exploratory Analysis of Alice in Wonderland



Project Gutenberg offers 53,494 free ebooks to download.

[Search](#) [Latest](#) [Terms of Use](#) [Donate?](#) [Mobile](#)

## All Books (sorted by popularity)

A

[Sort Alphabetically](#)



[Sort by Release Date](#)

Displaying results 1-25 | [Next](#)



[Pride and Prejudice](#)

Jane Austen  
20330 downloads



[Alice's Adventures in Wonderland](#)

Lewis Carroll  
14055 downloads



[Frankenstein; Or, The Modern Prometheus](#)

Mary Wollstonecraft Shelley  
12601 downloads



[The Adventures of Sherlock Holmes](#)

Arthur Conan Doyle  
10807 downloads



[Il Principe, English](#)

Niccolò Machiavelli  
10317 downloads



[Dracula](#)

Bram Stoker  
9776 downloads



[Adventures of Huckleberry Finn](#)

Mark Twain  
9379 downloads



[Metamorphosis](#)

Franz Kafka  
9148 downloads



[A Doll's House : a play](#)

Henrik Ibsen  
8959 downloads



[The Importance of Being Earnest: A Trivial Comedy for Serious People](#)

Oscar Wilde  
8577 downloads



[A Tale of Two Cities](#)

Charles Dickens  
8021 downloads



[Narrative of the Life of Frederick Douglass, an American Slave](#)

Frederick Douglass  
8011 downloads



[The Yellow Wallpaper](#)

Charlotte Perkins Gilman  
7917 downloads



[Great Expectations](#)

Charles Dickens  
7452 downloads



[Leviathan](#)

Thomas Hobbes  
7215 downloads



[Ulysses](#)

James Joyce  
7003 downloads

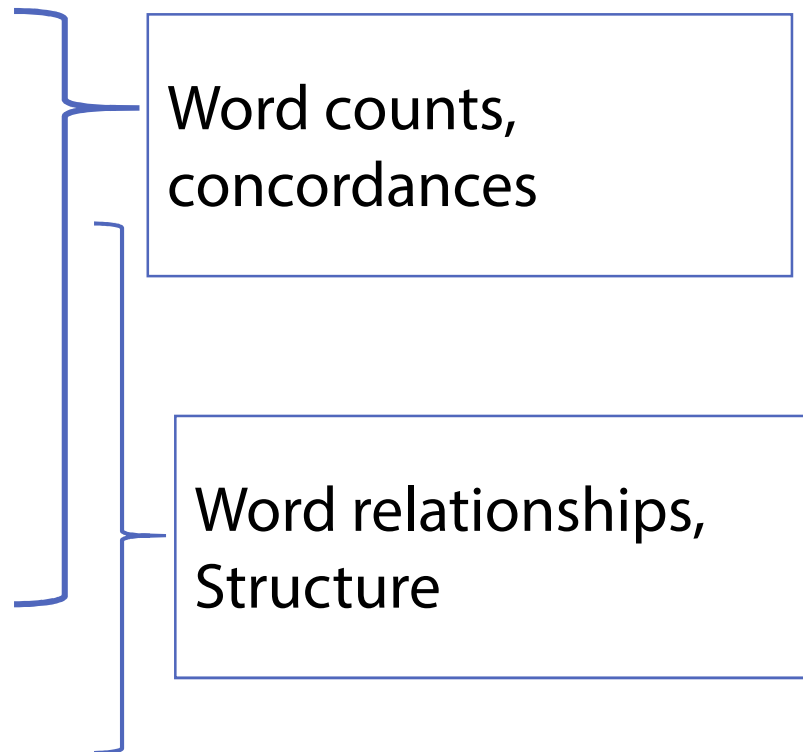


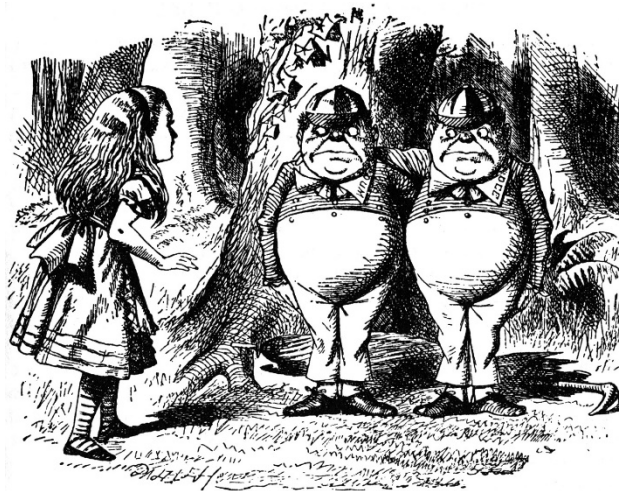
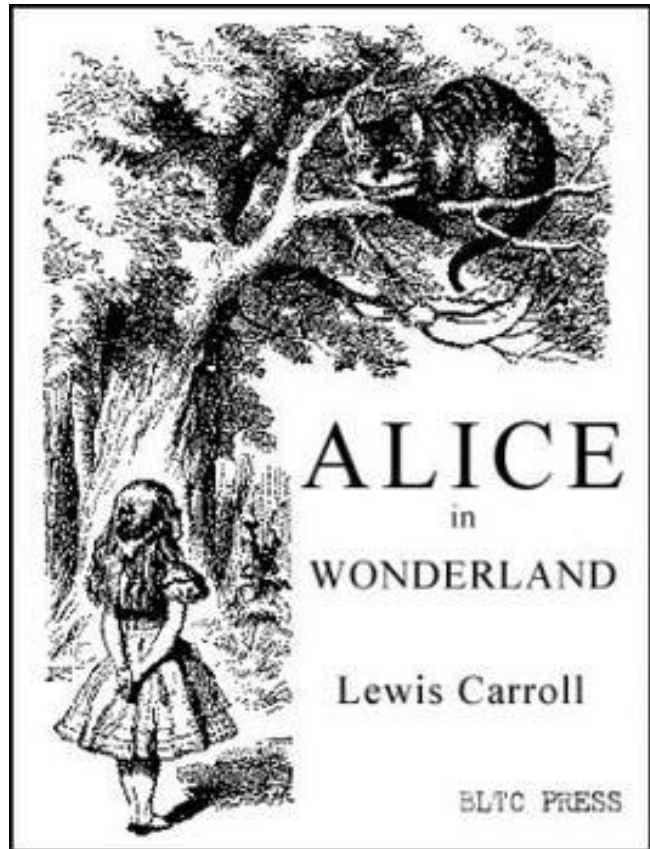
[The Romance of Lust: A Victorian erotic novel](#)

Anonymous  
6947 downloads


## Tools – Lexical Data

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









Since Carroll was a mathematician at Christ Church, it has been suggested that there are many references and mathematical concepts in both this story and also in *Through the Looking-Glass*; examples include:

- In chapter 1, "Down the Rabbit-Hole", in the midst of shrinking, Alice waxes philosophic concerning what final size she will end up as, perhaps "going out altogether, like a candle."; this pondering reflects the concept of a limit.
- In chapter 2, "The Pool of Tears", Alice tries to perform multiplication but produces some odd results: "Let me see: four times five is twelve, and four times six is thirteen, and four times seven is—oh dear! I shall never get to twenty at that rate!" This explores the representation of numbers using different bases and positional numeral systems:  $4 \times 5 = 12$  in base 18 notation,  $4 \times 6 = 13$  in base 21 notation, and  $4 \times 7$  could be 14 in base 24 notation. Continuing this sequence, going up three bases each time, the result will continue to be less than 20 in the corresponding base notation. (After 19



# Word Counts

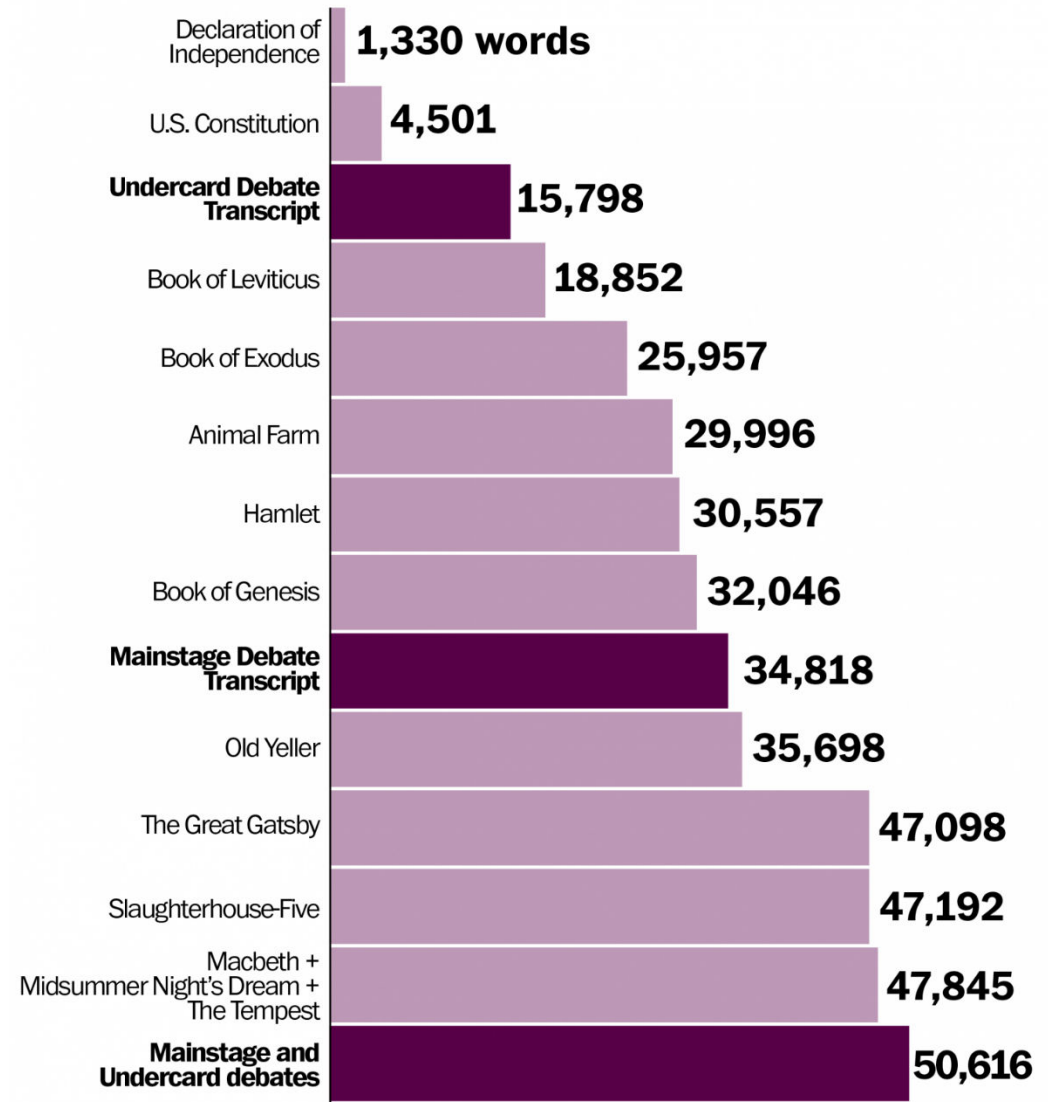
- How many times does each (stemmed?) word appear?



# Word Counts

## A lengthy debate

Word counts



<https://www.washingtonpost.com/news/wonk/wp/2015/09/17/last-nights-primetime-debate-was-longer-than-the-book-of-genesis/>

WAPO.ST/**WONKBLOG**

Source: CNN, National Archives, NaNoWriMo, overviewbible.com, Open Source Shakespeare

# Concordance

- Lines taken from a corpus (one book, multiple books) showing a **target word** in **context**
- Target word is always in the **middle** allowing skimming of contexts

1	Is Oxfam	interested	in disarmament?
2	But she was not	interested	in education for its own sake.
3	Parents are obviously	interested	in the various skills and ...
4	.. to change as more women become	interested	in outdoor pursuits.
5	In other words, he was not	interested,	as Singer was, in proving ...
6	.. last month said a range of other	interested	parties had emerged, including ..

# Concordance

- Lines taken from a corpus (one book, multiple books) showing a **target word** in **context**
- Target word is always in the **middle** allowing skimming of contexts

1	Is Oxfam	interested	in disarmament?
2	But she was not	interested	in education for its own sake.
3	Parents are obviously	interested	in the various skills and ...
4	.. to change as more women become	interested	in outdoor pursuits.
5	In other words, he was not	interested,	as Singer was, in proving ...
6	.. last month said a range of other	interested	parties had emerged, including ..

# Concordance

- Lines taken from a corpus (one book, multiple books) showing a **target word** in **context**
- Target word is always in the **middle** allowing skimming of contexts

1	Is Oxfam	interested	in disarmament?
2	But she was not	interested	in education for its own sake.
3	Parents are obviously	interested	in the various skills and ...
4	.. to change as more women become	interested	in outdoor pursuits.
5	In other words, he was not	interested,	as Singer was, in proving ...
6	.. last month said a range of other	interested	parties had emerged, including ..

# KWIC – Keyword in Context

- A form of permuted index

Titel	Catalognummer
Office 2003 Timesaving Techniques for Dummies	100344
Excel for Scientists and Engineers	2704
How To Do Everything With Microsoft Office Excel 2003	904
How To Do Everything With Microsoft Office PowerPoint 2003	900
Data Analysis Using SQL and Excel	12377
Automated Data Analysis Using Excel	3434
How To Do Everything With Microsoft Office Excel 2003	904
Accessing and Analyzing Data with Excel 2003	5
Excel 2003 Formulas	708
Microsoft Office Excel 2003 Inside Out	18765
Mastering Excel 2003 Programming with VBA	7655
Excel 2003 VBA Programmer's Reference	3099
Creating Spreadsheets and Charts in Microsoft Office Excel 2007	203
Excel for Scientists and Engineers	2704
Beginning Excel: What-If Data Analysis Tools	8467
Fixing Powerpoint Annoyances	5575
Excel 2003 Formulas	708
Microsoft Visio 2003 User's Guide	32001
How To Do Everything With Microsoft Office Excel 2003	904
How To Do Everything With Microsoft Office PowerPoint	900
Microsoft Office Excel 2003 Inside Out	18765
Mastering Excel 2003 Programming with VBA	7655
How To Do Everything With Microsoft Office Excel 2003	904

[https://en.wikipedia.org/wiki/Key\\_Word\\_in\\_Context](https://en.wikipedia.org/wiki/Key_Word_in_Context)

# AntConc

[Home](#)[Resume](#)[Publications ▼](#)[Software ▼](#)[Classes](#)[Photo Albums ▼](#)[Links](#)[Contact](#)

## AntConc Homepage

### Latest Release



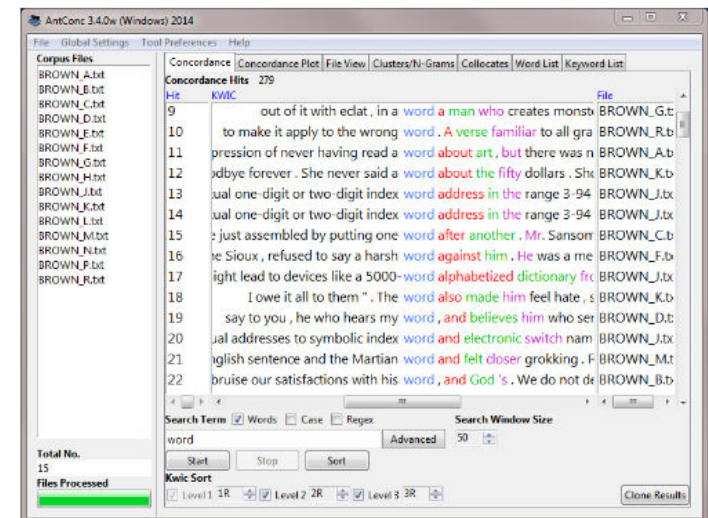
## AntConc

A freeware corpus analysis toolkit for concordancing and text analysis.

[\[AntConc Homepage\]](#) [\[Screenshots\]](#) [\[Help\]](#)

### Downloads:

- [Windows \(3.4.4\)](#)
- [Macintosh OS X 10.7-10.12 \(3.4.4\)](#)
- [Macintosh OS X 10.6 \(3.4.1\)](#)
- [Linux \(3.4.3\)](#)
- [Older versions](#)





AntConc 3.4.4w (Windows) 2014

FileGlobal SettingsTool PreferencesHelp

Corpus Files

Alice in Wonderland.tx

ConcordanceConcordance PlotFile ViewClusters/N-GramsCollocatesWord ListKeyword List

Concordance Hits50

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, witho	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

Advanced

Search Window Size50

StartStopSort

Kwic Sort☒ Level 11R☒ Level 22R☒ Level 33R

Clone Results

Total No.1

Files Processed



# Keyword Plots

Queen



No. of Hits = 75

File Length (in chars) = 163817

Alice



No. of Hits = 403

File Length (in chars) = 163817

Cat

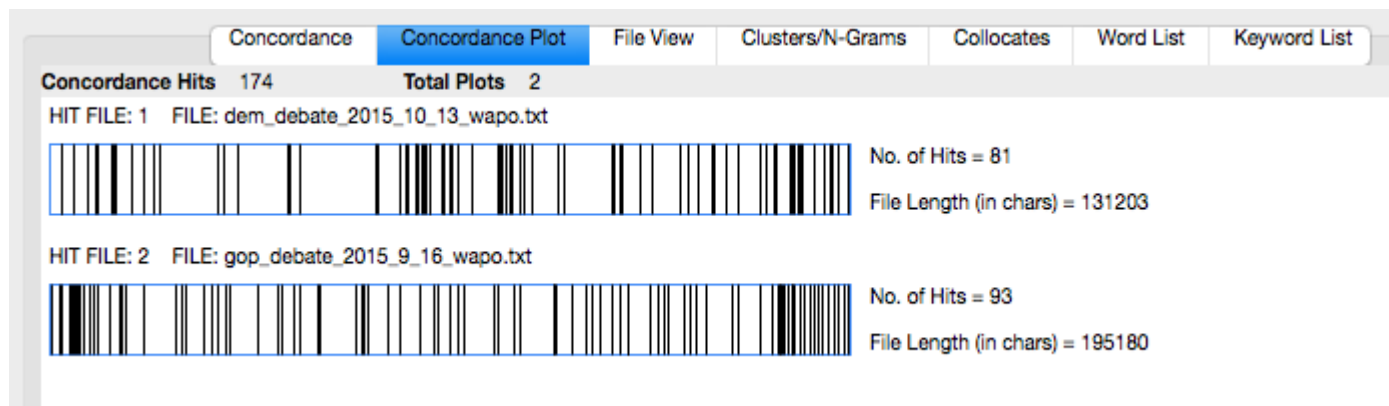


No. of Hits = 37

File Length (in chars) = 163817

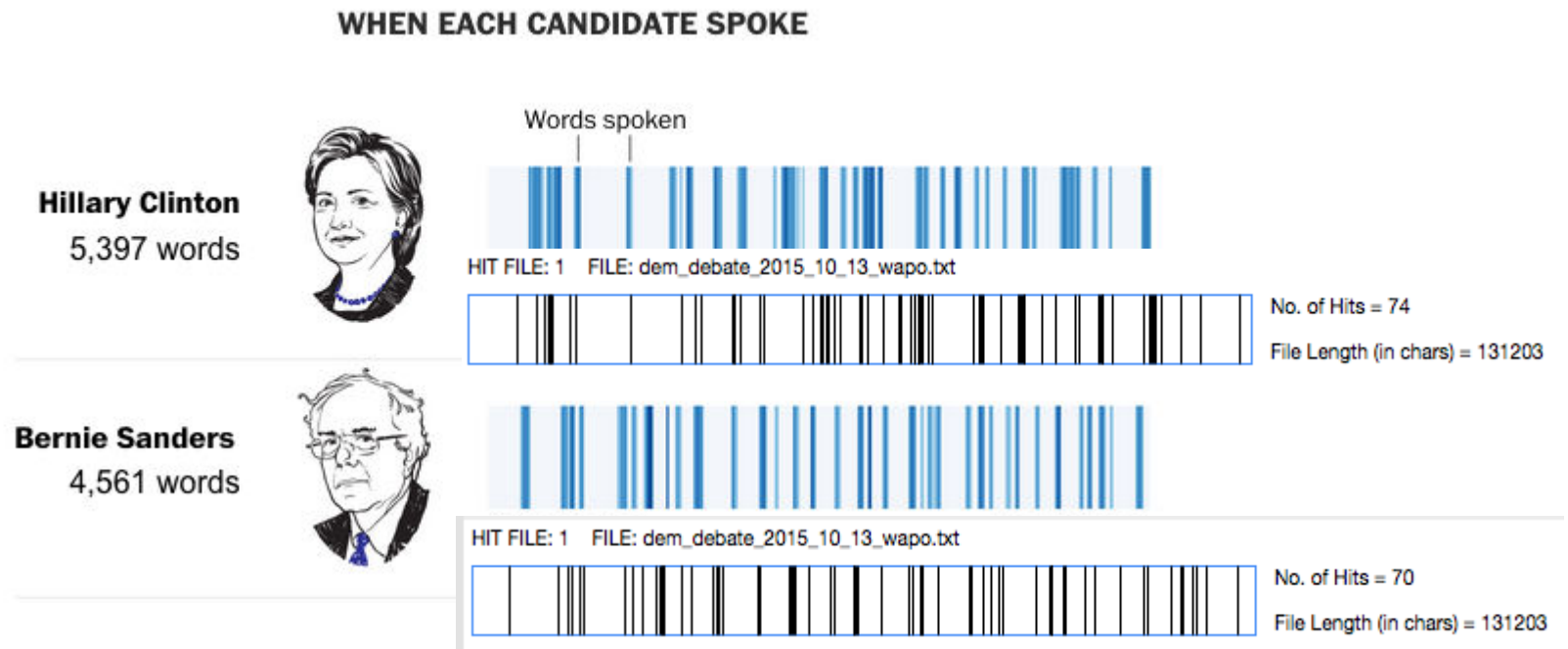
# Keyword Plots

- Using keyword plots, at what points in the last Republican/Democratic debates did <applause> occur?



Example courtesy Lynn Cherny:  
<http://ghostweather.slides.com/lynncherny/text-data-analysis-without-programming#/4/3>

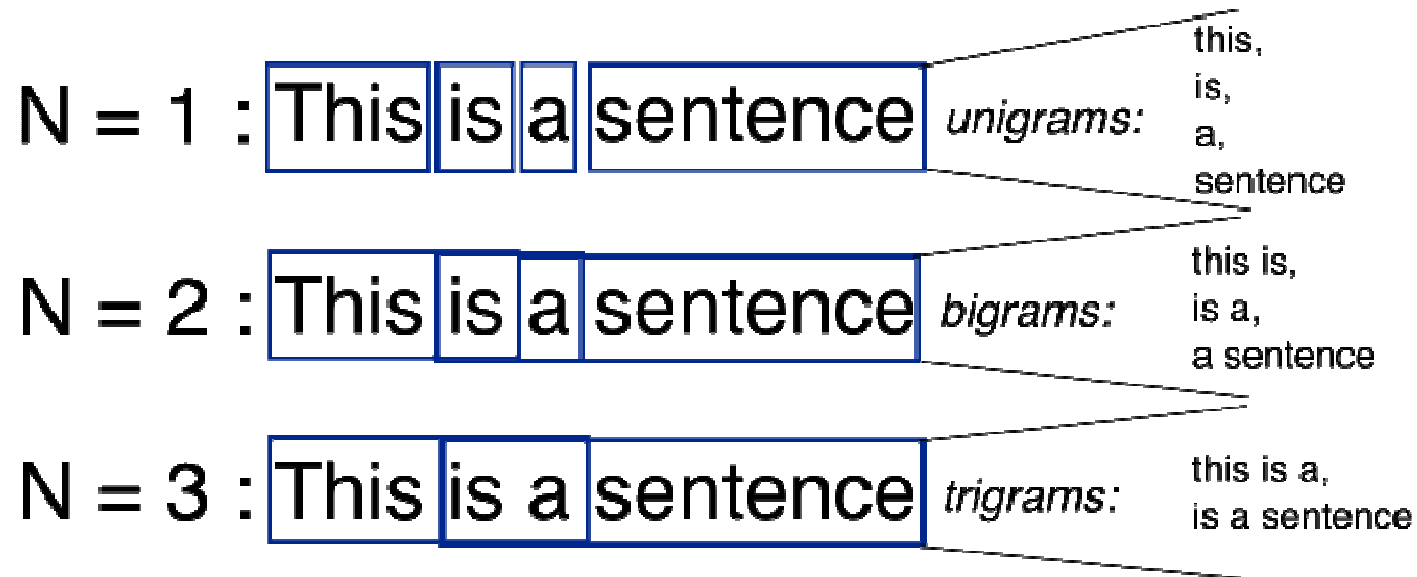
# Look Familiar?



As published in the Washington Post

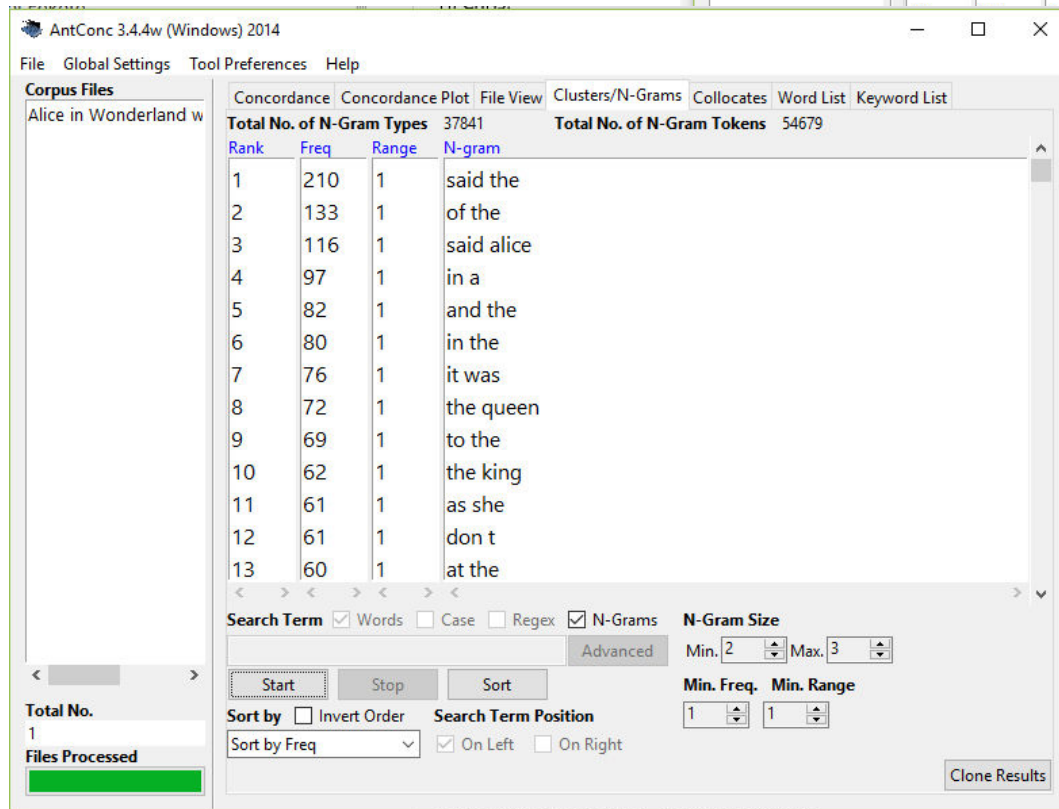
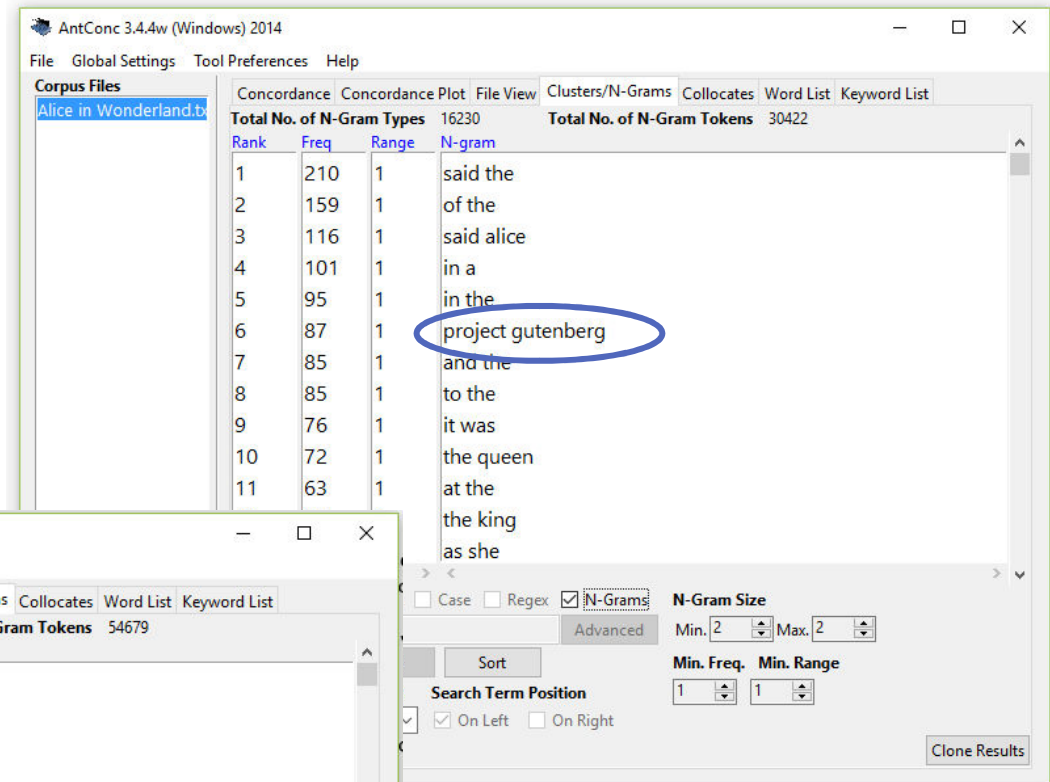
# N-Grams

- An “n-gram” is a sequence of words, POS, or letters of length n
- Each sequence is counted
- Used in document analysis, but also spelling and grammar checking, translation, and more



# AntConc n-grams

- View top n-grams
- Search 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 I!' 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



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 44**

Hit	KWIC	File
1	a bat, and that's very like a mouse, you know. But do cats ea	Alice in Won
2	oon made out that it was only a mouse that had slipped in like h	Alice in Won
3	' thought Alice, 'to speak to this mouse? Everything is so out-of-	Alice in Won
4	arm in trying.' So she began: 'O Mouse, do you know the way ou	Alice in Won
5	red of swimming about here, O Mouse!' (Alice thought this mus	Alice in Won
6	e the right way of speaking to a mouse: she had never done suc	Alice in Won
7	her brother's Latin Grammar, 'A mouse--of a mouse--to a mouse	Alice in Won
8	Latin Grammar, 'A mouse--of a mouse--to a mouse--a mouse--	Alice in Won
9	ar, 'A mouse--of a mouse--to a mouse--a mouse--O mouse!') T	Alice in Won
10	ouse--of a mouse--to a mouse--a mouse--O mouse!') The Mouse	Alice in Won
11	ouse--to a mouse--a mouse--O mouse!') The Mouse looked at h	Alice in Won
12	ouse--a mouse--O mouse!') The Mouse looked at her rather inqu	Alice in Won
13	ght Alice; 'I daresay it's a French mouse, come over with William	Alice in Won
14	in her French lesson-book. The Mouse gave a sudden leap out c	Alice in Won

Search Term ☒ Words ☐ Case ☐ Regex  Advanced Search Window Size 50

Start Stop Sort

**Kwic Sort**  
☒ Level 1 1R ☒ Level 2 2R ☒ Level 3 3R

Total No. 1  
 Files Processed

Clone Results

# Collocates

- “Collocations of a given word are statements of the habitual or customary places of that word.”
- Presence of one is highly indicative of the other (high *mutual information*)
- Semantics of collocates have a lesser degree of compositionality:
  - “strong tea” vs \*”forceful tea”
  - “broad daylight” vs \*”narrow darkness”
- In the extreme, can be used to detect idioms:
  - “heard it through the grapevine”
  - “kick the bucket”



# Collocates

- Generally:
  - Weakly or non-compositional meaning
    - White wine vs white hair
    - “kick the bucket”
  - Non substitutive
    - “White wine” vs. ?? “yellow wine”
    - “kick the bucket” vs. “kick the pail”
  - Non-modifiable
    - “a frog in your throat” vs ?? “an ugly frog in your throat”

# Words

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

**Word Types:** 4210 **Word Tokens:** 61053 **Search Hits:** 0

Rank	Freq	Word	Lemma	Word Form(s)
1	3419	the		
2	1847	and		
3	1546	to		
4	1451	a		
5	1276	it		
6	1205	i		
7	1118	of		
8	1097	she		
9	1097	you		
10	935	said		
11	853	alice		
12	835	in		
13	715	was		
14	689	that		

Search Term ☒ Words ☐ Case ☐ Regex

over Advanced

Start Stop Sort

Hit Location Search Only 0

Sort by ☐ Invert Order

Sort by Freq

Lemma List ☐ Loaded

Clone Results

Total No.  
2

Files Processed

# Stop Word Removal

RANKS NL >> Demo Plans & Pricing More ▾

Sign in

## Stopwords

7 |   

Share this page

[Home](#) > [Resources](#) > Stopwords

## Stopword Lists

Collection of stopword lists in many languages. Find the English stopwords below and/or follow the links to view our other language stop word lists.

## English Stopwords

### 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
.	..

## Stopwords in other languages

[Arabic](#)

[Armenian](#)

[Basque](#)

[Bengali](#)

[Brazilian](#)

[Bulgarian](#)

[Catalan](#)

[Chinese](#)

[Czech](#)

[Danish](#)

[Dutch](#)

[Finnish](#)

[French](#)

[Galician](#)

[German](#)

[Greek](#)

[Hindi](#)

Tool Preferences

**Category**

- Concordance
- Clusters/N-Grams
- Collocates
- Word List**
- Keyword List

**Word List Preferences**

**Display Options**

☒ Rank ☒ Frequency ☒ Word ☒ Lemma Word Form(s)

**Other Options**

☒ Treat all data as lowercase

☐ Treat case in sort

**Lemma List**

☐ Loaded

☐ Treat Word List Range as Lemma List Range

**Word List Range**

☐ Use all words ☐ Use specific words below ☒ Use a stoplist below

Add Word

Add Words From File

about  
above  
after  
again

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

**Word Types:** 3651 **Word Tokens:** 25944 **Search Hits:** 0

Rank	Freq	Word	Lemma Word Form(s)
1	935	said	
2	853	alice	
3	260	queen	
4	259	one	
5	245	little	
6	213	know	
7	210	like	
8	171	went	
9	160	thought	
10	153	see	
11	141	now	
12	141	time	
13	129	king	
14	126	just	

**Search Term** ☒ Words ☐ Case ☐ Regex

**Advanced**

**Hit Location** **Search Only** 0

**Lemma List** ☐ Loaded

**Total No.**  
2

**Files Processed**

**Sort by** ☐ Invert Order  
Sort by Freq

Start Stop Sort

Clone Results

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 Advanced

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

# Corpus n-grams vs Collocates

The screenshot displays the AntConc 3.4.4w (Windows) 2014 interface. The main window is titled "AntConc 3.4.4w (Windows) 2014" and has a menu bar with "File", "Global Settings", "Tool Preferences", and "Help". The "Clusters/N-Grams" tab is active, showing a table of cluster types and tokens. The table has columns for Rank, Freq, Range, and Cluster. The total number of cluster types is 436 and the total number of cluster tokens is 756. The search term is "white". The search options are: Words (checked), Case (unchecked), Regex (unchecked), N-Grams (unchecked). The cluster size is set to 2. The search term position is set to "On Left". The search term is "white". The search options are: Words (checked), Case (unchecked), Regex (unchecked), N-Grams (unchecked). The cluster size is set to 2. The search term position is set to "On Left". The search term is "white".

Rank	Freq	Range	Cluster
1	95	35	white house
2	33	8	white sox
3	19	14	white and
4	17	1	white snake
5	15	10	white man
6	9	6	white people
7	7	6	white males
8	6	2	white beans
9	5	1	white : i
10	5	1	white : it
11	5	5	white of
12	5	2	white river
13	5	5	white shirt

The "Collocates" tab is also visible, showing a table of collocates. The table has columns for Freq(R), Stat, and Collocate. The total number of collocate tokens is 9430. The search term is "white". The search options are: Words (checked), Case (unchecked), Regex (unchecked), N-Grams (unchecked). The window span is set to 5L. The search term position is set to "On Left". The search term is "white".

Freq(R)	Stat	Collocate
0	11.23712	treepie
3	11.23712	throated
1	11.23712	sunbird
1	11.23712	nuthatch
1	11.23712	macque
1	11.23712	darter
1	11.23712	browed
2	11.23712	babbler
1	10.23712	zigzagging
0	10.23712	yuk
0	10.23712	wma
0	10.23712	wingshooting
1	10.23712	whitish
0	10.23712	whitewinn





# **Finding Keywords**

# Dunning Log Likelihood ( $G^2$ )

$$E1 = c \frac{(a + b)}{(c + d)}$$

$$E2 = d \frac{(a + b)}{(c + d)}$$

	Analysis Text	Reference Text	Total
Count of word	a	b	a + b
Count of other words	c - a	d - b	c + d - a - b
Total	c	d	c + d

$$G^2 = 2((a \cdot \ln(a / E1)) + (b \cdot \ln(b / E2)))$$

Translates approximately to significance, e.g.  $6.63 = p < 0.01$



**The Collected Works of  
William Shakespeare**

## Word Counts

begotten

hast

love

thou



Macbeth

## Word Counts

begotten

hast

love

thou



Macbeth

Log Likelihood

wrath

blood

lady

witch

# 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

Sort by ☐ Invert Order Sort by Keyness

Reference Corpus ☒ Loaded

Total No.  
2  
Files Processed

# Negative 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: 3651 Search Hits: 0

Rank	Freq	Keyness	Keyword
3309	3	0.000	reading
3310	1	0.000	shared
3311	2	0.000	train
3312	2	0.000	willing
1	7	37.320	says
2	4	36.358	years
3	5	31.744	year
4	4	28.988	also
5	3	19.350	around
6	17	18.659	new
7	1	16.923	need
8	1	15.564	family
9	1	14.956	today
10	2	14.456	woman

Search Term ☒ Words ☐ Case ☐ Regex

Hit Location Search Only 0

Reference Corpus ☒ Loaded

Clone Results

Total No.  
2

Files Processed

Sort by ☐ Invert Order  
Sort by Keyness

Start Stop Sort

Advanced



# wordandphrase.info

## WORD AND PHRASE . INFO

FREQUENCY LISTS - [ANALYZE TEXTS](#) | [ALL GENRES](#) - ACADEMIC

ENTER TEXT BELOW -SAMPLES- [MY TEXTS](#)

SEARCH

CLEAR

HELP

☒ WORD

☐ PHRASE

Select individual words in the text to see "word sketches"

You can enter any text that you would like in the form at the top, and then see useful information about words and phrases in that text.

First, it will **highlight** all of the medium and lower-frequency words in the text, and it can allow you to see "what the text is about" (i.e. text analysis).

Second, you can click through the **words** in the text to see a "word sketch" (i.e. meaning into the meaning and usage of the word), re-sortable by frequency, etc.

Finally, you can do powerful searches on selected **phrases** in the text, likely in different styles of English. For example, if you click on a phrase, it will show you the frequency of those phrases in COCA -- over time, etc.

By Mark Davies

# wordandphrase.info

Less frequent in reference corpus

SEE LISTS

FREQ RANGE

1-500

501-3000

> 3000

2193 WORDS

67 %

10 %

23 %

## CHAPTER I. Down the Rabbit-Hole

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

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

There was nothing so **VERY remarkable** in that; **nor** did Alice think it **so** **VERY** much out of the way to hear the Rabbit say to **itself**, 'Oh **dear** !

Alice in Wonderland has a pretty high percentage of rare words for a children's book.

Word frequency highlighting

# 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<sub>j</sub> (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 Ozzfest 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



## WordWanderer

*take your text for a walk*

We are experimenting with visual ways in which we can enhance people's engagement with language. By fusing the information we can obtain from corpus searches, concordance outputs and word clouds we are aiming to enable and encourage people to notice and wander through the words they read, write and speak.

To try an early prototype of the WordWanderer just select a text below or paste your own and take it for a walk.

- ☐ Hansel and Gretel (Brothers Grimm)
- ☒ The Yellow Wallpaper (Gillman)
- ☐ In the Penal Colony (Kafka)
- ☐ 2 B R 0 2 B (Vonnegut)
- ☐ The Library of Babel (Borges)
- ☐ Jekyll & Hyde (Stevenson)
- ☐ Your own text: paste in box →

remembering her own child-life, and  
the happy summer days.

THE END

☒ turn text to lower case

wander

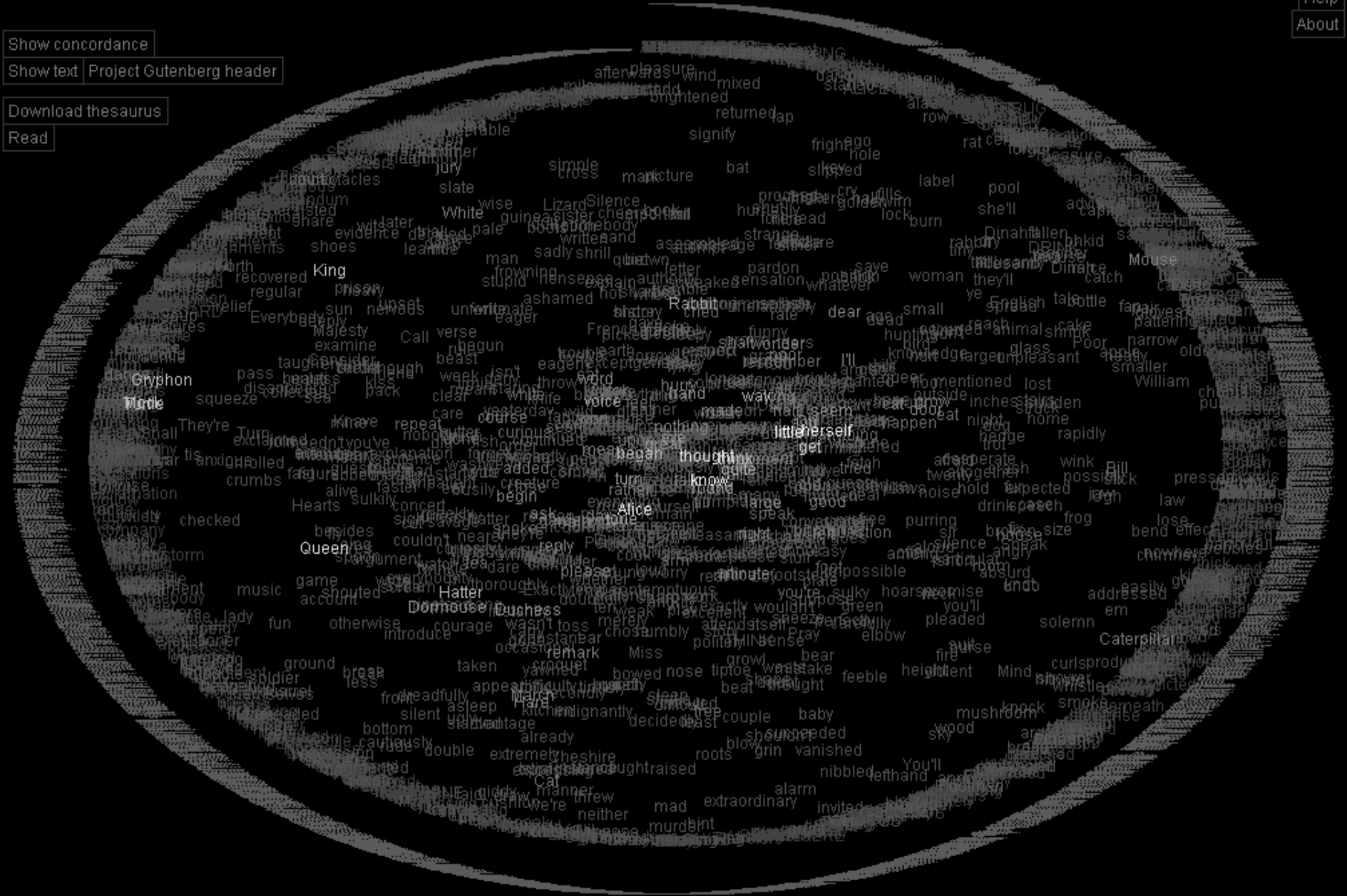


queen

alice

[illegible]

# Alice's Adventures In Wonderland

[Help](#)[About](#)[Show concordance](#)[Show text](#) [Project Gutenberg header](#)[Download thesaurus](#)[Read](#)

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



# Alice's Adventures In Wonderland

They very soon came upon a Gryphon, lying fast asleep in the

Show concordance

Show text Project Gutenberg header

Download thesaurus

Read

Help

About

Gryphon

Mock

Queen

Hatter

Domouse

Duchess

March

Cheshire

Cat

Caterpillar

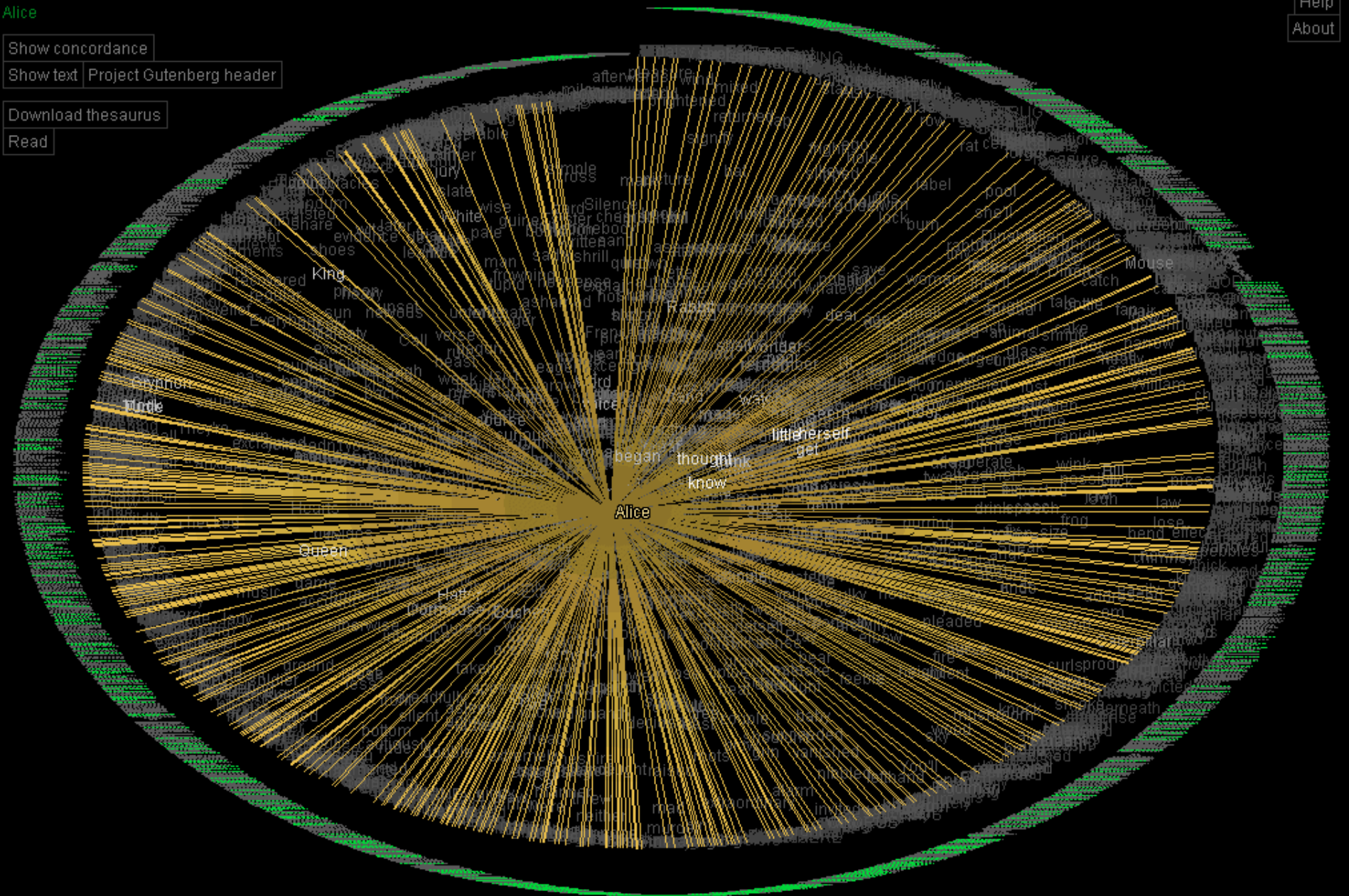
Help

Show concordance

Show text | Project Gutenberg header

[Download thesaurus](#)

Read



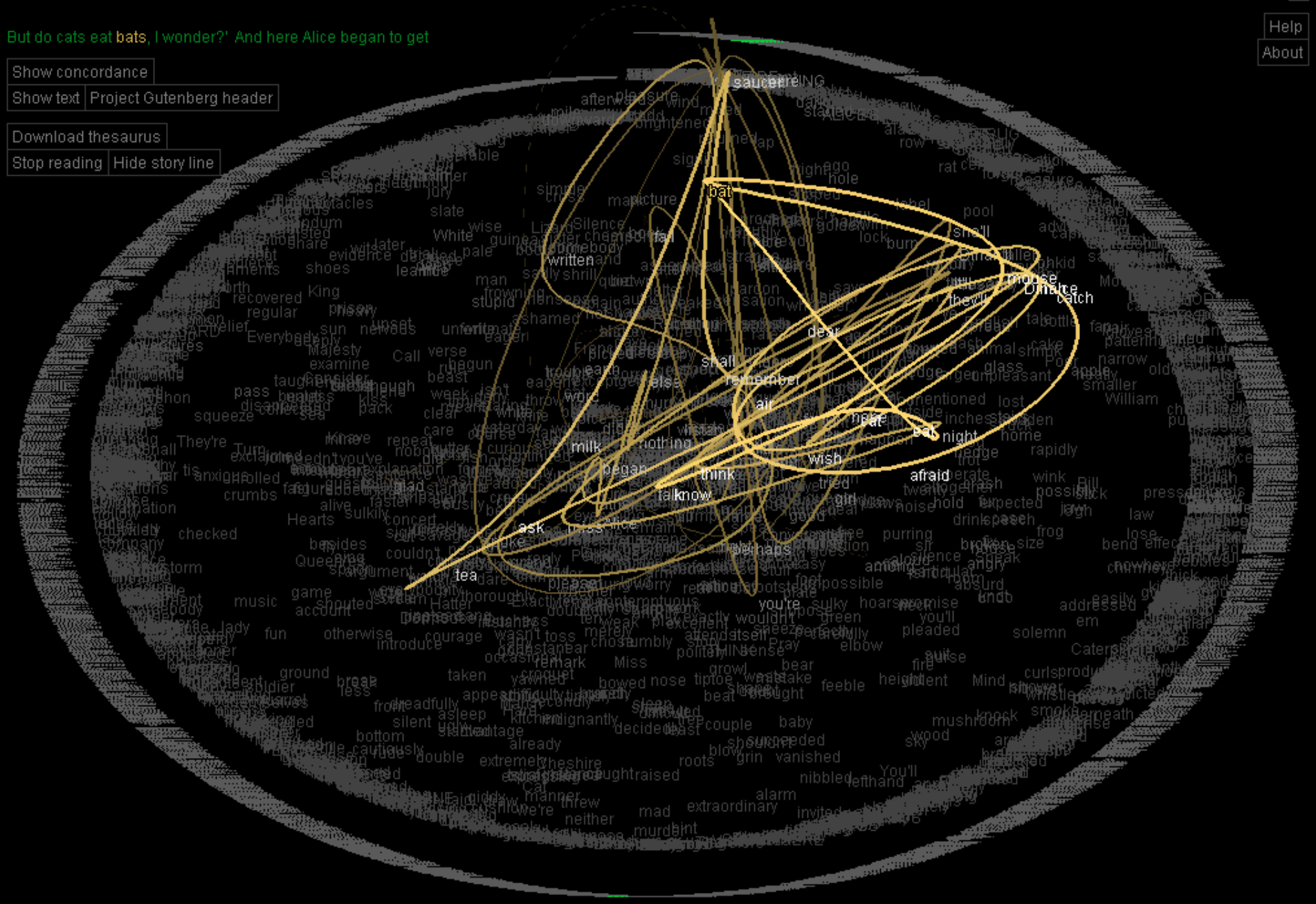


Help

About

Show text | Project Gutenberg header

Stop reading Hide story line



# Alice's Adventures In Wonderland



Help  
About

Text

☐ Show only KWIC index (Key Word In Context)

Show

Hide text Project Gutenberg header

Download thesaurus

Read

ALICE'S ADVENTURES IN WONDERLAND

Lewis Carroll

THE MILLENNIUM FULCRUM EDITION 3.0

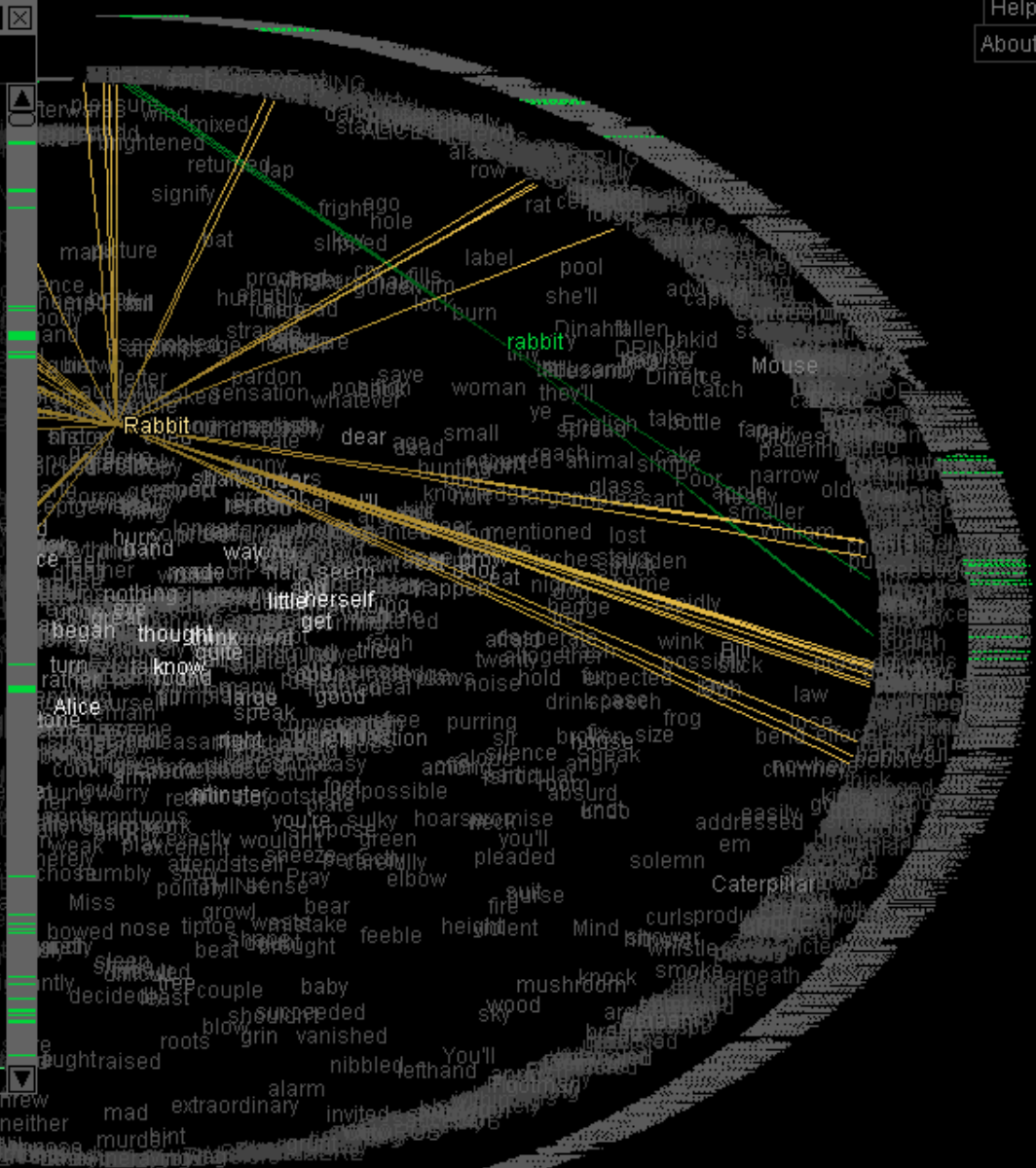
CHAPTER I

Down the Rabbit-Hole

Alice was beginning to get very tired of sitting by her sister, curious to see what was going on in the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book?' thought Alice 'without pictures or conversation?'

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

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



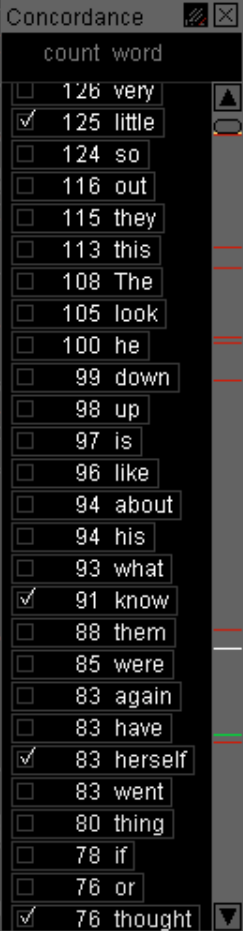
Help

☐ Hide concordance

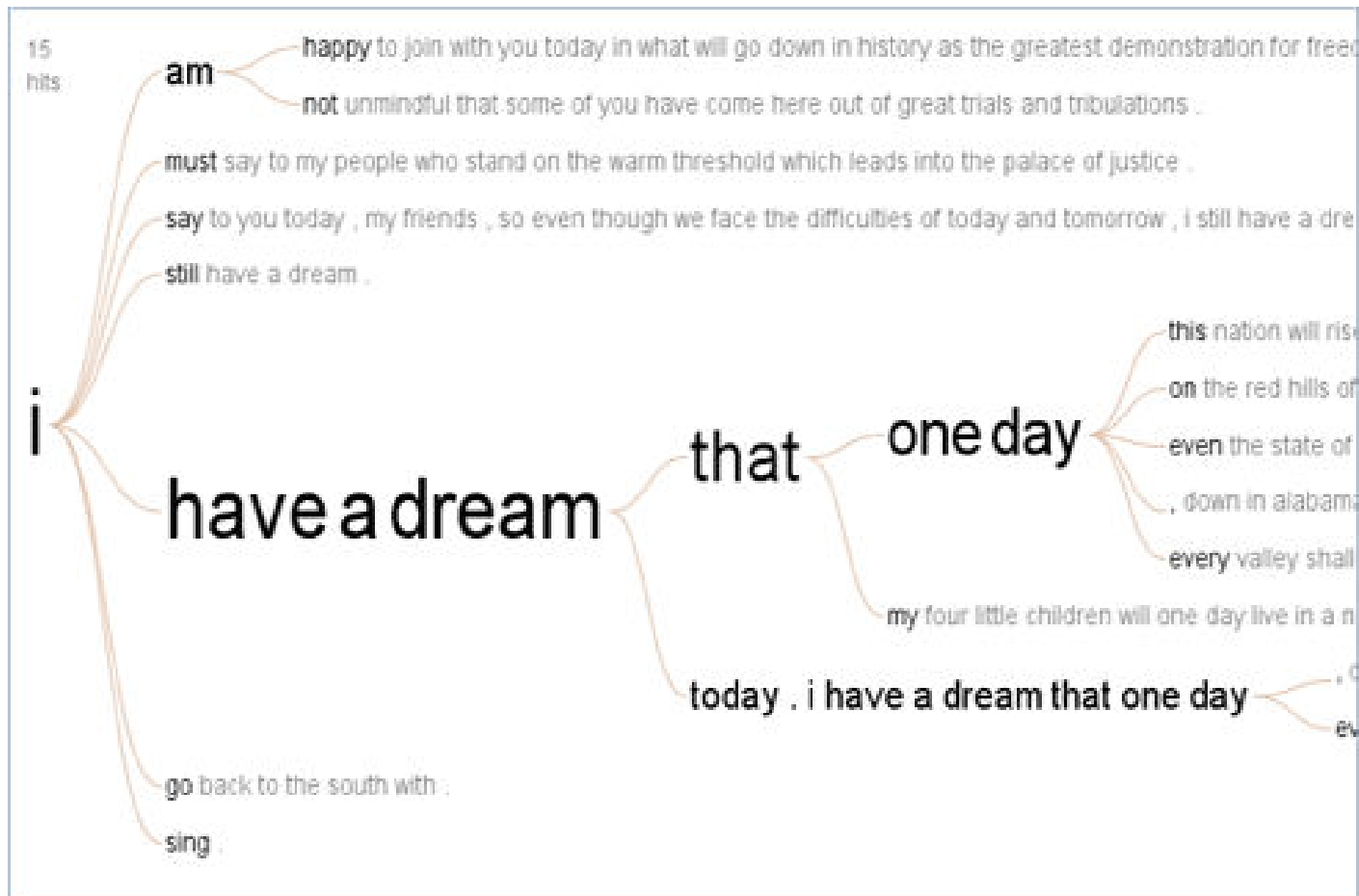
Show text | Project Gutenberg header

☐ Hide thesaurus lookup

Read



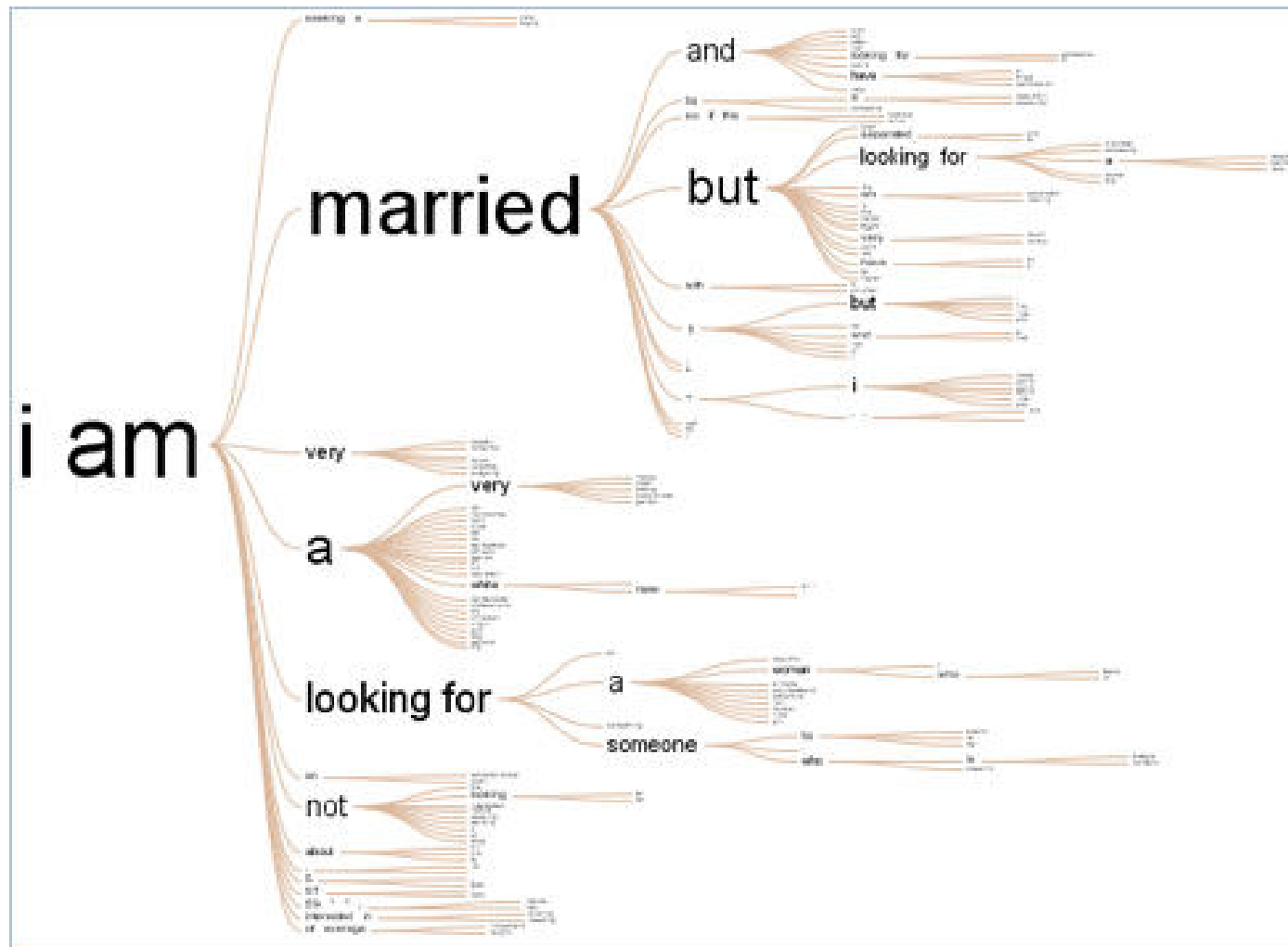
# Word Tree



<https://www.jasondavies.com/wordtree> based on M. Wattenberg and F. B. Viegas. The Word Tree, an Interactive Visual Concordance. IEEE Transactions on Visualization and Computer Graphics, 14(6):1221-1228, 2008.



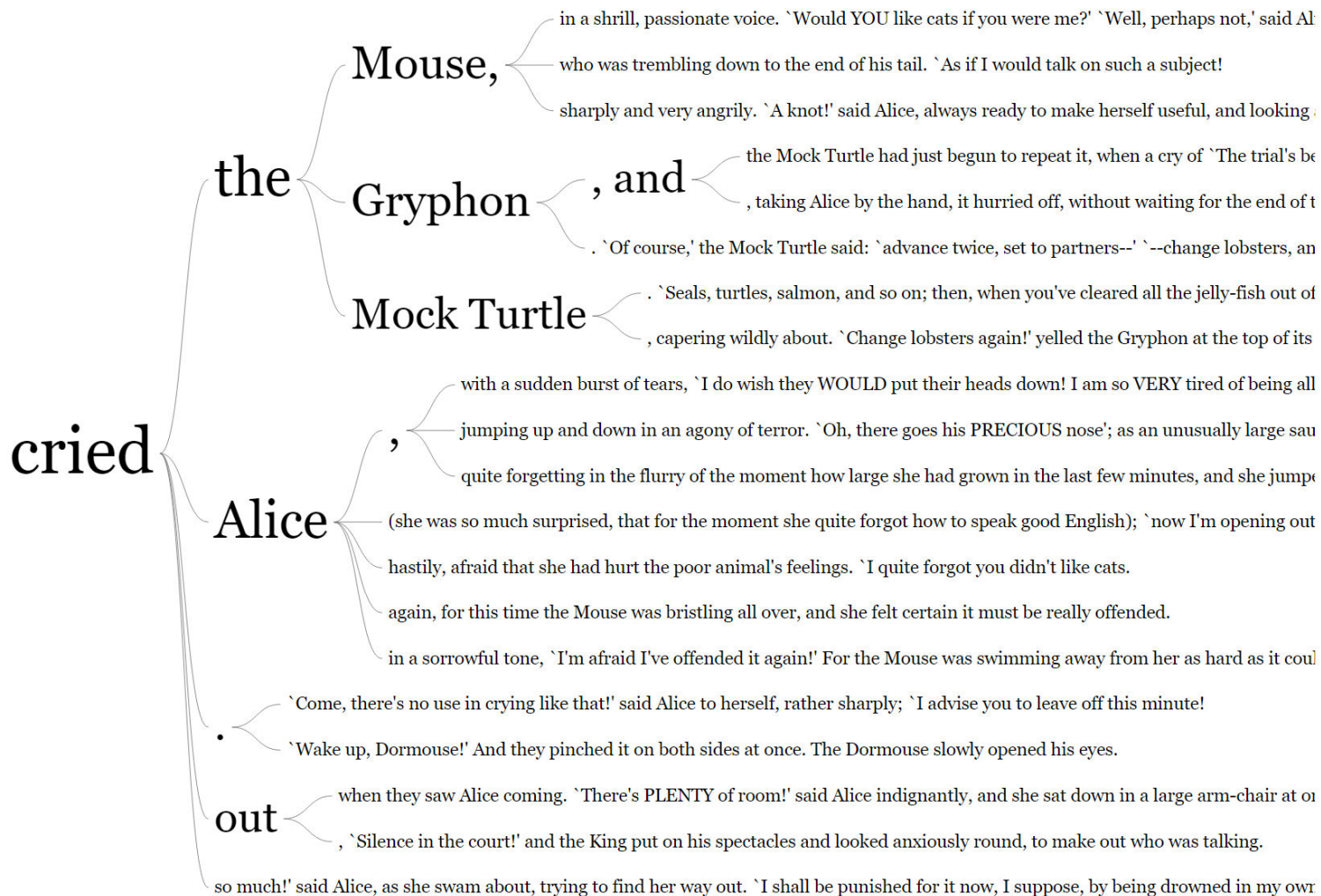
# Word Tree





# Word Tree: Alice

Shift-click to make that word the root.

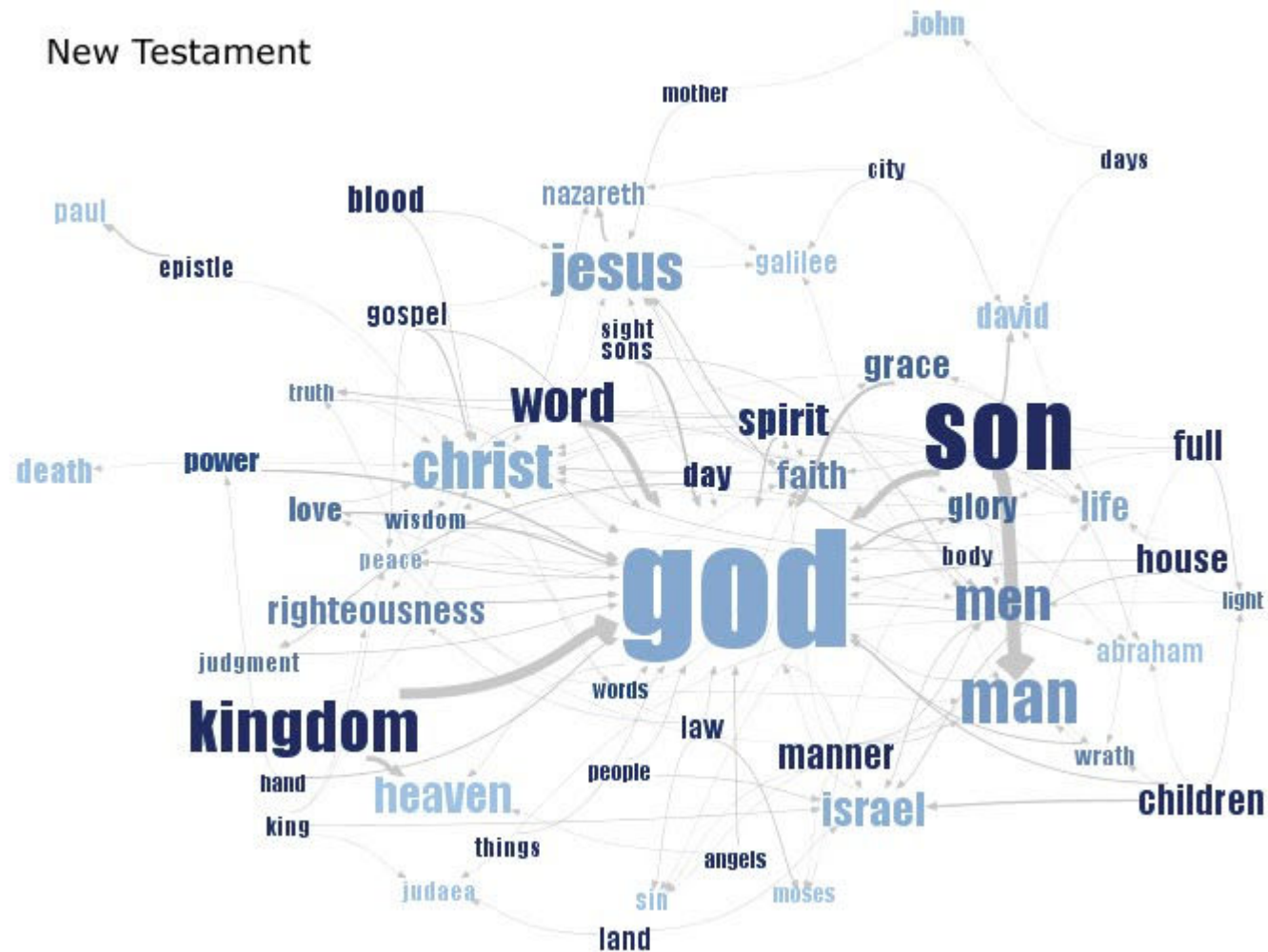


forgot you didn't like cats.

'Not like cats!' cried the Mouse, in a shrill, voice.n'Would YOU like cats you were me?'

'Well, perhaps not,' said iniaesoothing tone: 'don't b angry about it. And yet I wishould show you our cat Ditahak you'd take a fancy t cats if you could only see Bae. is such a dear quiet thAdge went on, half to herself, as she swam lazily about in the pool, 'and she sits purring so nicely by th fire, licking her paws and washing her face--and she is such a nice soft thing to nuaa she's such a capital eee catching mice--oh, I beg your pardon!' cried Alice ,gfor this time the Mouse wa bristling all over, and she felt certain it must be offended. 'We won't talk Aeouany more if you'd rather not.'

'We indeed!' cried the Mouse who was trembling down to th end of his tail. 'As if I talkdon such a subject! Our family always HATED cats: partW, vulgar things! Don't methear the name again!'



Frank van Ham, Martin Wattenberg, and Fernanda B. Viegas. 2009. Mapping Text with Phrase Nets. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (November 2009), 1169-1176.

# Phrase Net

Select a phrase

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

or enter your own

\* and \*

Submit

Filters

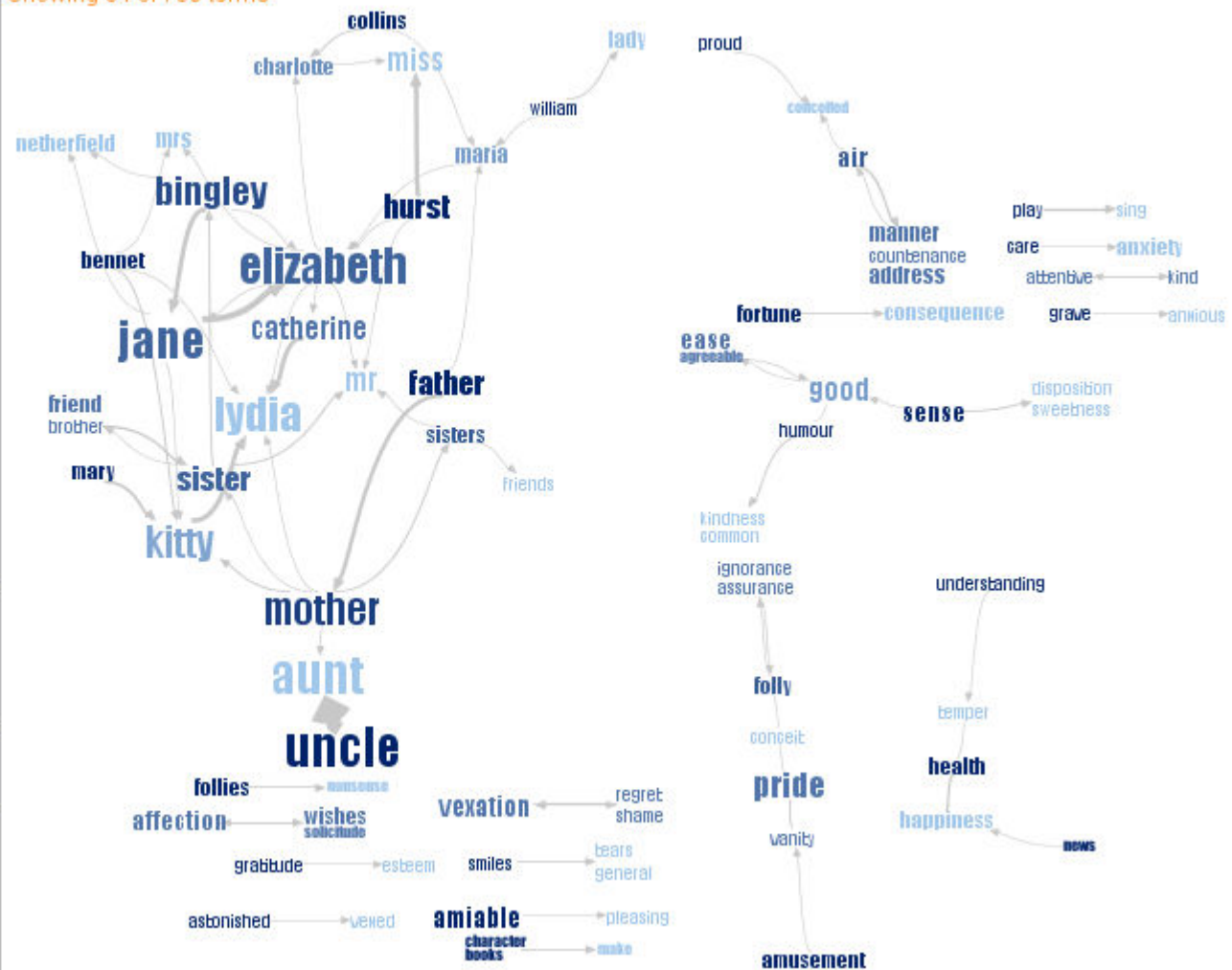
Show top: 100

Hide common words ☒

Zoom

In  Out  Reset 

Showing 84 of 759 terms





## Tools – Semantic data

- Entity Resolution
  - TextRazor
- Semantic Structure
  - DocuBurst



# Named Entity Recognition

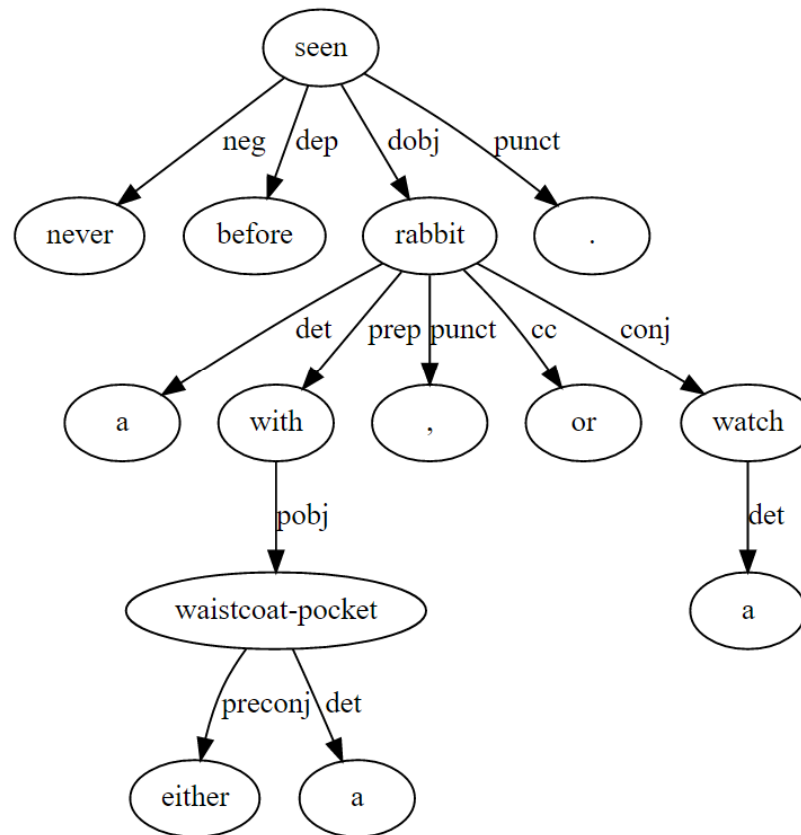
- Detect people, places, companies, etc. from text
- Heuristic (rule-based) and other approaches
- Entity resolution means to link the entity with a unique (shareable) ID, e.g. DBPedia, PermID

# TextRazor

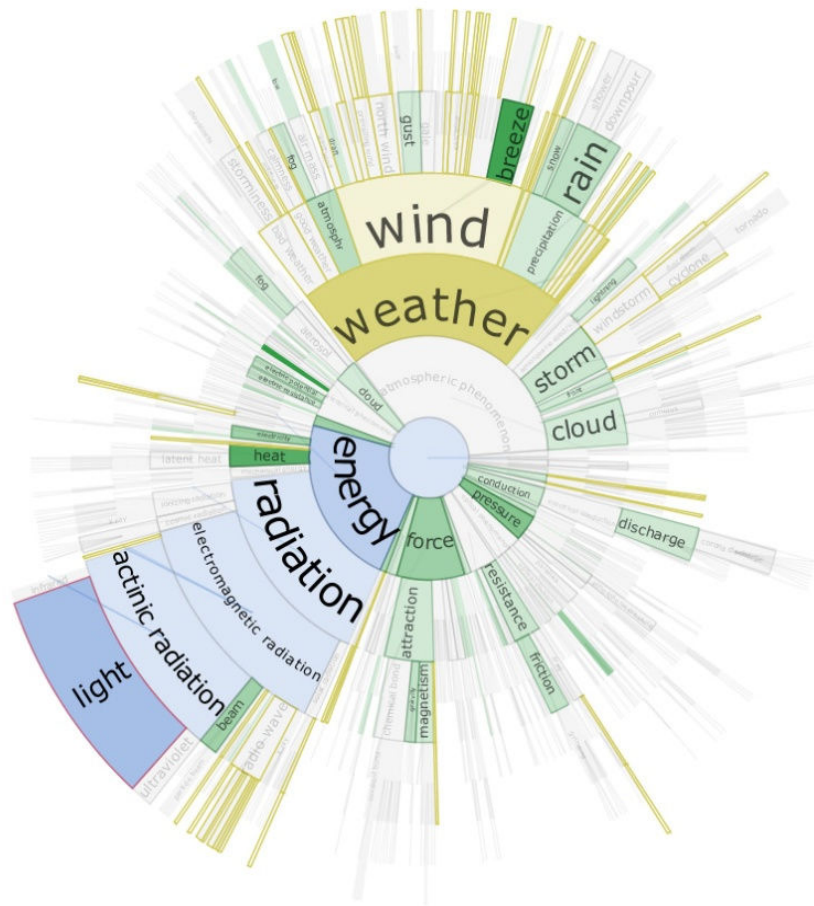


never before seen a rabbit with either a waistcoat-pocket, or a watch

[Words](#) [Phrases](#) [Relations](#) [Entities](#) [Meaning](#) [Dependency Parse](#)

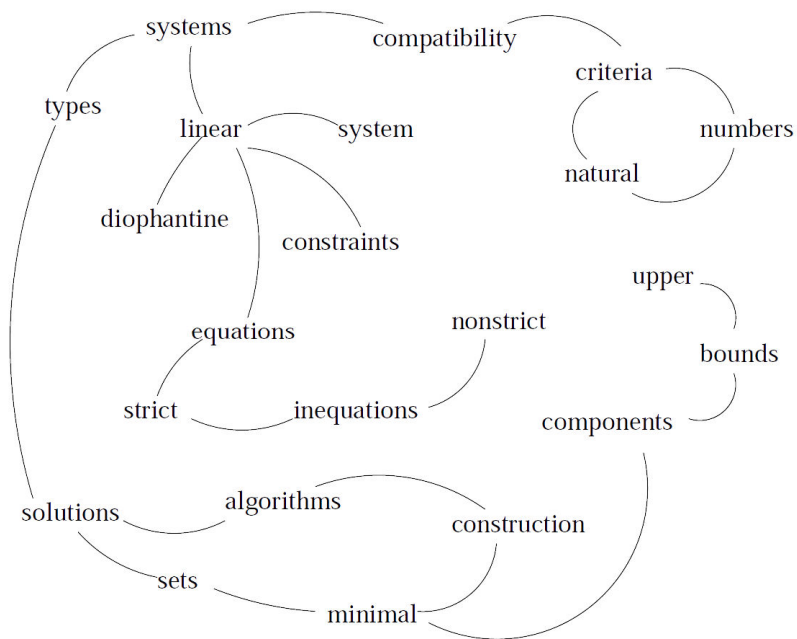


# DocuBurst

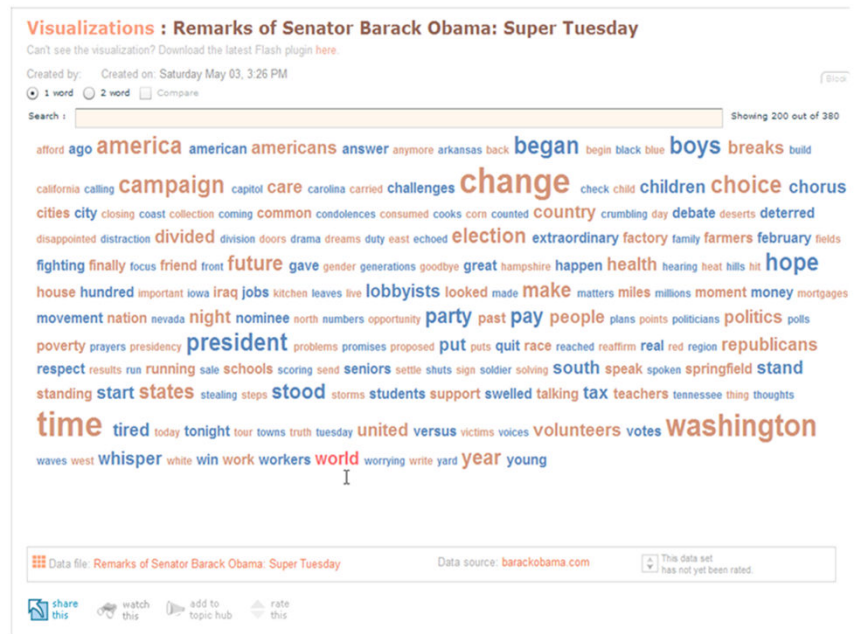


Collins, C.; Carpendale, S.; Penn, G. DocuBurst: Visualizing Document Content using Language Structure. Proceedings of Eurographics/IEEE VGTC Symposium on Visualization, June, 2009.



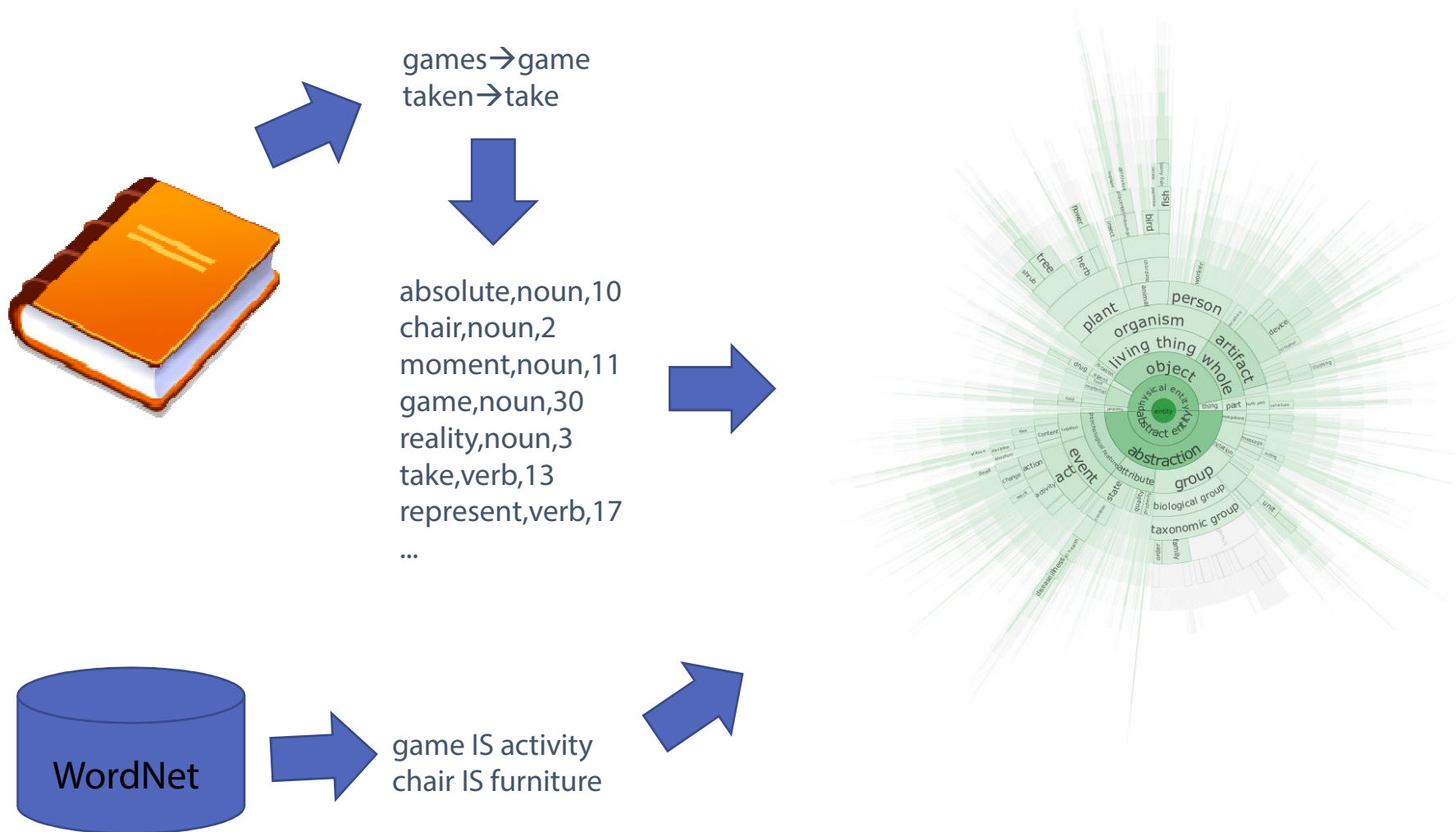


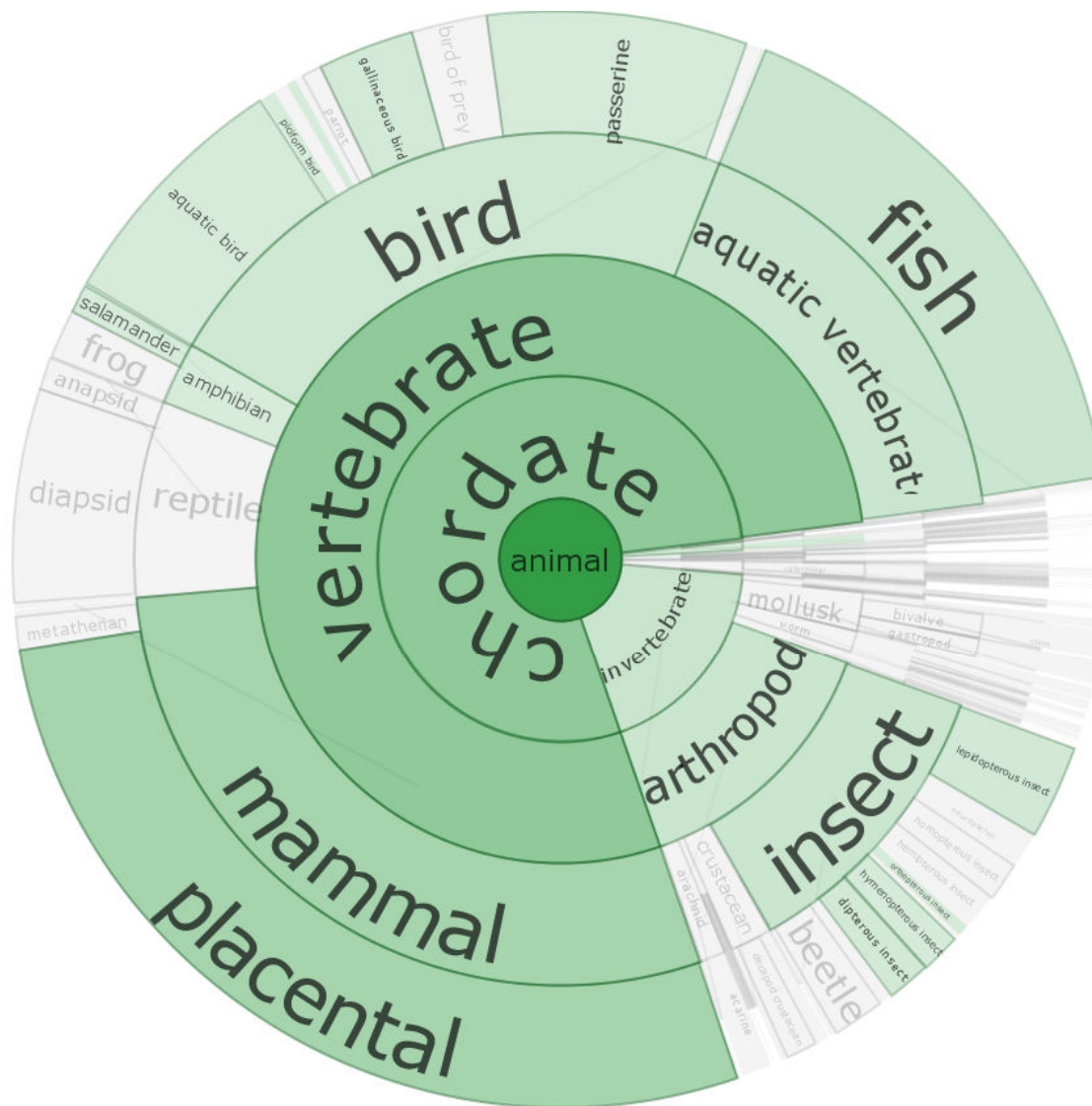
Mihalcea and Tarau, 2004



Wattenberg et al., 2008

# DocuBurst





0 119.86

Search Filter Options Text Segments Concordance Lines

Focus:

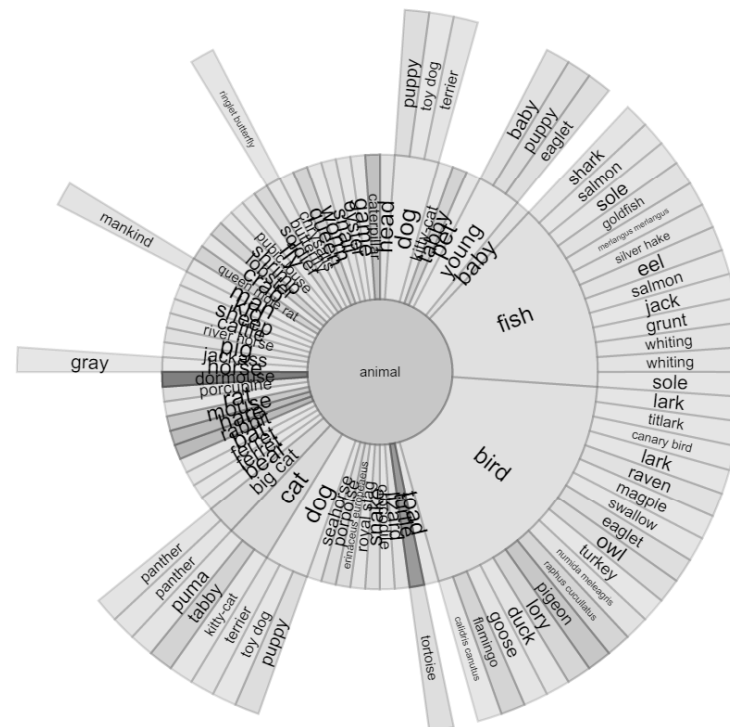
Word/Sense Details:

**POS:** noun

**Synonyms:** dipterous insect, two-winged insects, dipteran, dipterion


**Sense:** insects having usually a single pair of functional wings (anterior pair) with the posterior pair reduced to small knobbed structures and mouth parts adapted for sucking or lapping or piercing

100



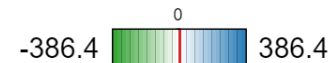
English  
Canterbury  
KING AND QUEEN  
King minute  
bank Alice Cat  
Mary Andthe Edgar  
Caterpillar  
even UN Maine  
the Queen  
Queen whole  
Alice didn  
farmer

Alice in Wonderland - fixed



0 39

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





# Tools - Document Collections

- Clustering documents
  - Lexos

# Lexos

## Lexos {loader} An Integrated Lexomics Workflow

[Workspace](#)[Reset](#)[Upload](#)[Manage](#)[Prepare ▾](#)[Visualize ▾](#)[Analyze ▾](#)

### File Upload

Max. file size: 250 MB -- Formats Supported: .txt, .html, .xml, .sgml, and .lexos [?](#)

[Browse](#)

Ready For Files To Upload

Or drop files here

Lexos v3.0 © 2016 Lexomics Research Group · Wheaton College (Norton, MA)

This site is licensed under the [MIT License](#).

[Tweet to @Lexomics](#)[Feedback and Support](#)

<http://lexos.wheatoncollege.edu/upload>



# Lexos {scrubber} An Integrated Lexomics Workflow

[Workspace](#)[Reset](#)[Upload](#)[Manage](#)[Prepare ▾](#)[Visualize ▾](#)[Analyze ▾](#)

## Scrubbing Options

- |  |   |
|--|---|
| <input checked="" type="checkbox"/> Remove All Punctuation | <input type="checkbox"/> Keep Hyphens ?                   |
| <input checked="" type="checkbox"/> Make Lowercase         | <input type="checkbox"/> Keep Word-Internal Apostrophes ? |
| <input checked="" type="checkbox"/> Remove Digits          | <input type="checkbox"/> Keep Ampersands ?                |
| <input type="checkbox"/> Remove Whitespace ?               |   |
| <input type="checkbox"/> Scrub Tags ?                      |   |

## Additional Options

[Stop Words/Keep Words ? >](#)[Lemmas ? >](#)[Consolidations ? >](#)[Special Characters ? >](#)[Preview Scrubbing](#)[Apply Scrubbing](#)[Download Scrubbed Files](#)

## Previews of Documents

### Alice\_in\_Wonderland\_without\_Gutenberg\_Terms.txt

ALICE'S ADVENTURES IN WONDERLAND Lewis Carroll THE MILLENNIUM FULCRUM EDITION 3.0 CHAPTER I. Down the Rabbit-Hole Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had ... ..ge tale, perhaps even with the dream of Wonderland of long ago: and how she would feel with all their simple sorrows, and find a pleasure in all their simple joys, remembering her own child-life, and the happy summer days. THE END



# K-Means Clustering

- Turn documents into vectors
- Cluster  $n$  items based on their similarity in  $k$  groups

# K-Means

Mouseover points or rows in the table to see document labels.

0	hellobarbie_lines_v2_2
0	hellobarbie_lines_v2_4
0	hellobarbie_lines_v2_22
0	hellobarbie_lines_v2_23
0	hellobarbie_lines_v2_24
0	hellobarbie_lines_v2_29
1	Alice_in_Wonderland_without_Gutenberg
1	Alice_in_Wonderland_without_Gutenberg
1	Alice_in_Wonderland_without_Gutenberg
1	Alice_in_Wonderland_without_Gutenberg
1	Alice_in_Wonderland_without_Gutenberg
1	Alice_in_Wonderland_without_Gutenberg
1	Alice_in_Wonderland_without_Gutenberg
1	Alice_in_Wonderland_without_Gutenberg
2	hellobarbie_lines_v2_13
2	hellobarbie_lines_v2_14
2	hellobarbie_lines_v2_15





# Comprehensive Tools


- Voyant
  - Developed by Digital Humanities Researchers
  - Provides many of the functions of previous tools online
  - Active development and user community


# VOYANT

see through your text

**Add Texts** ?

ALICE'S ADVENTURES IN WONDERLAND  
Lewis Carroll  
THE MILLENNIUM FULCRUM EDITION 3.0

 Open

 **Reveal**

*Voyant Tools is a web-based reading and analysis environment for digital texts.*

<https://voyant-tools.org/>

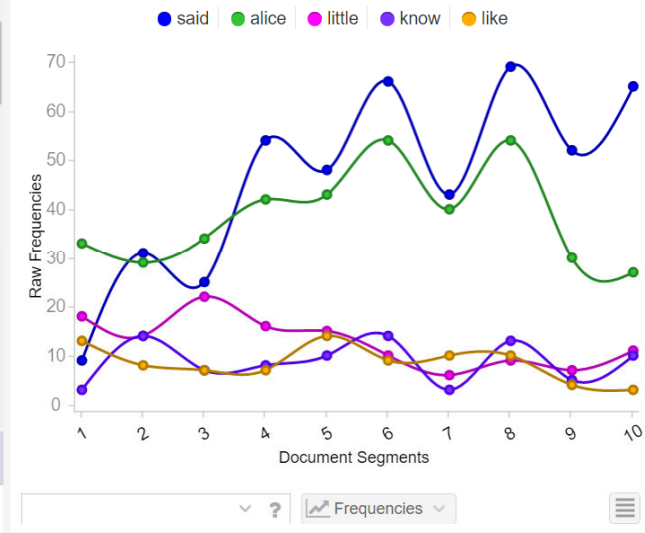


**ALICE'S ADVENTURES IN WONDERLAND** Lewis Carroll ...

ALICE'S ADVENTURES IN WONDERLAND  
Lewis Carroll  
THE MILLENNIUM FULCRUM EDITION 3.0  
CHAPTER I. Down the Rabbit-Hole

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

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



Summary

Documents

Phrases

This corpus has 1 document with 26,694 total words and 2,634 unique word forms. Created about 6 days ago.

Most **frequent words** in the corpus: **said** (462); **alice** (386); **little** (128); **know** (87); **like** (85)

items:

Contexts

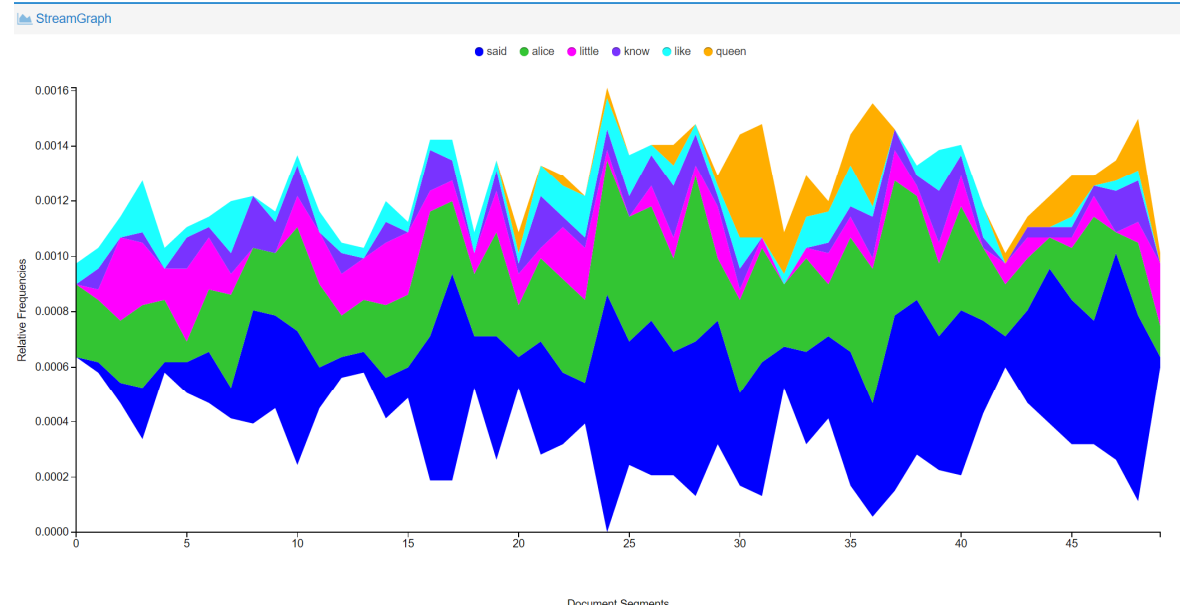
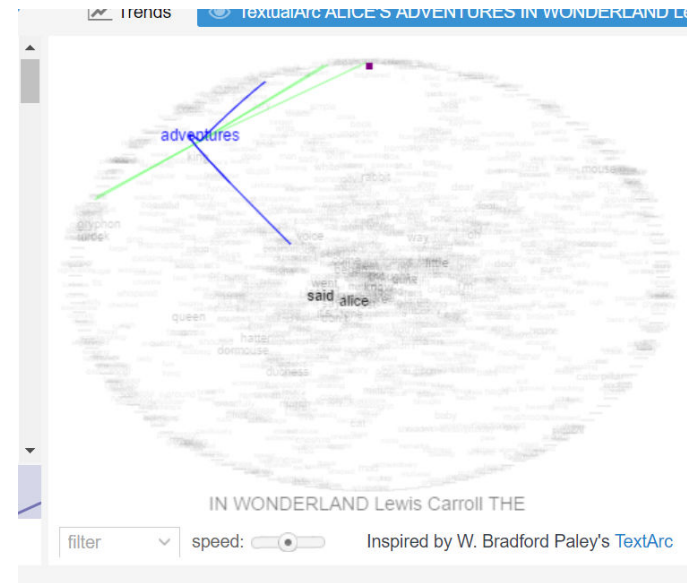
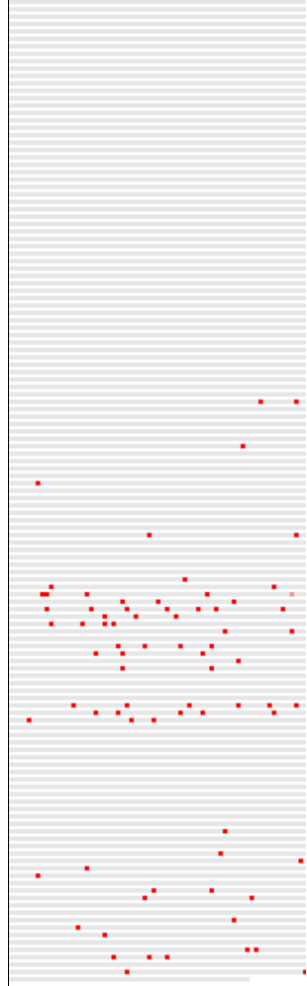
Bubblelines

Document	Left	Term	Right
1) ALIC...	fallen by this time?" she	said	aloud. 'I must be getting
1) ALIC...	certainly was not here before,'	said	Alice,) and round the neck
1) ALIC...	No, I'll look first,' she	said	, 'and see whether it's marked
1) ALIC... off. ***** 'What a...		said	Alice; 'I must be shutting
1) ALIC...	it might end, you know,'	said	Alice to herself, 'in my
1) ALIC...	use in crying like that!' said	said	Alice to herself, rather sharply
1) ALIC...	currants. 'Well, I'll eat it,'	said	Alice, 'and if it makes
1) ALIC...	ate a little bit and	said	anxious to herself. 'Which way

462 context expand

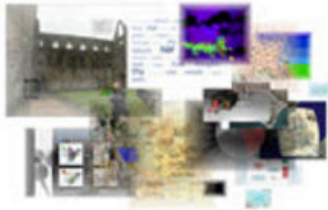
# Embedded Tools

ALICE'S ADVENTURES IN WONDERLAND





# More Resources



**DH | Digital Humanities Resources for Project  
Toychest | Building**

(curated by [Alan Liu](#))

[Guides to Digital Humanities](#) | [Tutorials](#) | [Tools](#) | [Examples](#) | [Data Collections & Datasets](#)

---

<http://dhresourcesforprojectbuilding.pbworks.com>

## TAPoR 3

## Discover research tools for studying texts.

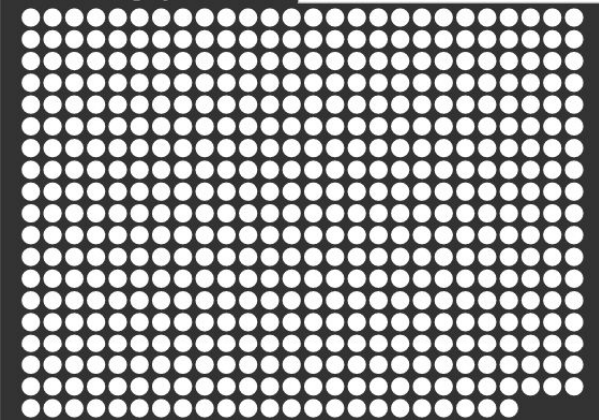


Voyant 2.0 is a complete rewrite of Voyant. It provides a suite of text analysis tools that will work with most texts you can upload or find on the web. These tools are combined in skins. For documentation see [Documentation for Voyant 2.0](#)



Select category

All



## Voyant Tools 2.0 (Corpus View)

# Text Visualization Browser

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

Provided by ISOVIS group

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

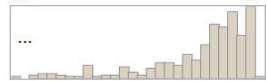
Techniques displayed:

341

Search:

Time filter:

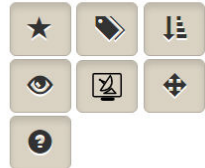
1976 2017



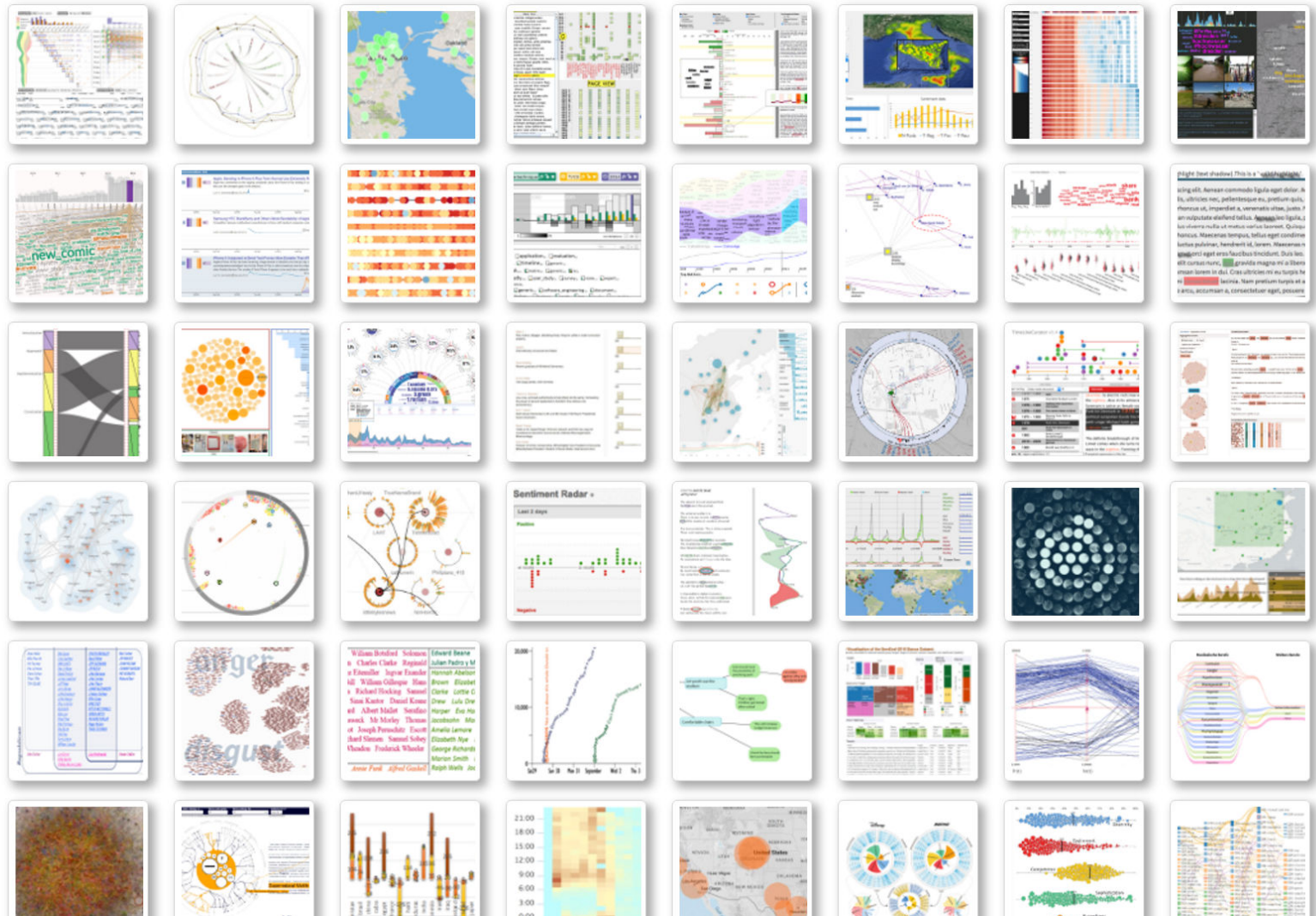
Analytic Tasks



Visualization Tasks



Data



<http://textvis.lnu.se>



## Next Time

- More visualization techniques:
  - Visualizing topic modelling, document variations, sonic patterns, readability, and more
- Case Studies in Visual Text Analytics for DH