

# **Lesson 10: Text Visualisation: Concepts and Methods**

**Dr. Kam Tin Seong  
Assoc. Professor of Information Systems**

**School of Computing and Information Systems,  
Singapore Management University**

**2021-07-04 (updated: 2021-07-05)**

# Content

- Introduction Text Visualisation
  - Text data
- Text Visualisation Methods
  - Tag Cloud
  - Wordle
  - Word Tree
  - Phrase Nets
- R Packages for Text Visualisation

# Introduction to Text Visualisation

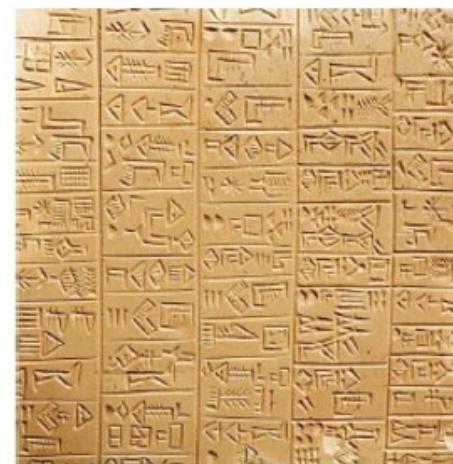
## History of text

Chauvet cave  
proto-writing



~20,000 years ago

Sumerian cuneiform  
logographic



~5,000 years ago

Phoenician abjad  
predecessor of alphabet

א	,	ת	P
ב	ג	י	S
ג	ד	ק	Q
ד	ה	ל	R
ה	ו	מ	š
ו	ז	נ	X
ז	ח	ס	T
ח	ו	ׁ	

~3,000 years ago

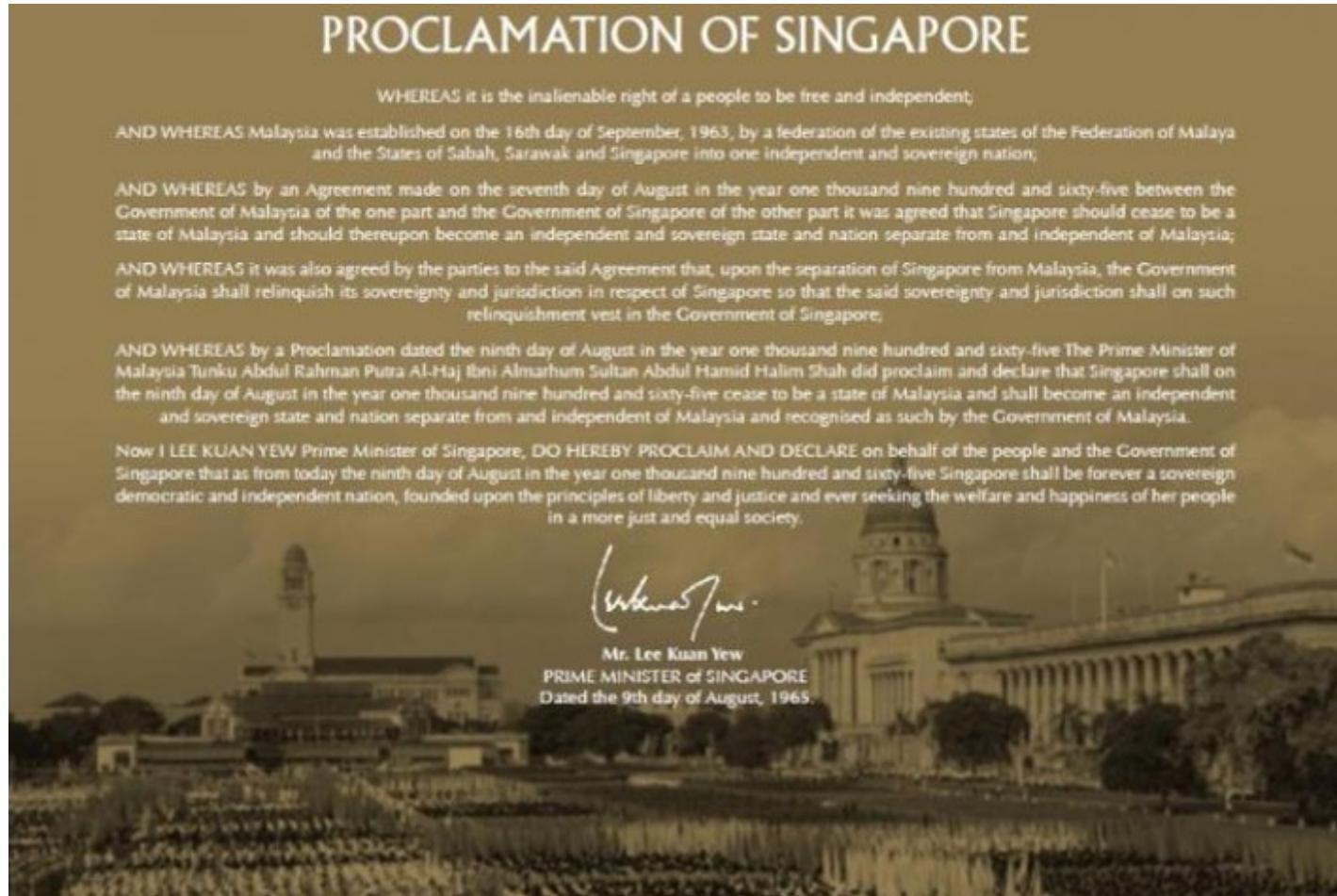
Latin letters

~2,500 years ago

A B C D E F  
G H I J K L M  
N O P Q R S  
T U V W X Y Z

# Introduction to Text Visualisation

## Text as historical archive



# Introduction to Text Visualisation

## Text as a mode of communication



# Introduction to Text Visualisation

## Textual data for business intelligence analytics

By Channel NewsAsia, Updated: 08/10/2011



### Market for new flats likely to stabilise in four years: PM Lee

Market for new flats likely to stabilise in four years: PM Lee

33% | 67% | Shared 5 times

Tweet 2 Recommend 22

SINGAPORE: Prime Minister Lee Hsien Loong said he expects the recent slew of housing measures to stabilise the market for new flats in four years. He also gave his assurance that the prices of new flats will remain affordable.

Mr Lee was speaking at a ceremony to mark the completion of upgrading works in a precinct in Ang Mo Kio.

In his speech, Mr Lee noted the concerns many Singaporeans have over buying new flats and he outlined several ways in which housing authorities are working to meet demand.

Some of these measures have started to take effect, and Mr Lee said he expects the market to stabilise in about four years' time.

twitter acryilinc Alena Yilin Lee  
"@STcom: Hundreds stranded at Tanjong Pagar MRT station after train breaks down [bit.ly/nZFEDx](http://bit.ly/nZFEDx)" Somebody needs to do sth to SMRT lah!  
1 hour ago



Sarah

Spacious and in a very quiet area. Very pleasant. Barbara and Nicolas are welcoming and kind. My daughter and I are delighted to stay everything was perfect. And the place is close to everything really perfect plan.

October 2014

Translated from Français



Vicky

Great location and cozy and clean room. Nicolas and Barbara were great hosts. We had a wonderful experience to stay in Nicolas and Barbara's house. And we would like to stay longer when we come to Paris again. Thank you very much, Barbara and Nicolas!

October 2014



Coralie

Second stay with Nicolas and Barbara, always nice! The room is spacious, airy and comfortable. Welcoming Nicolas and Barbara was very attentive. I thoroughly recommend it!

October 2014

Translated from Français

## Why Visualise Text?

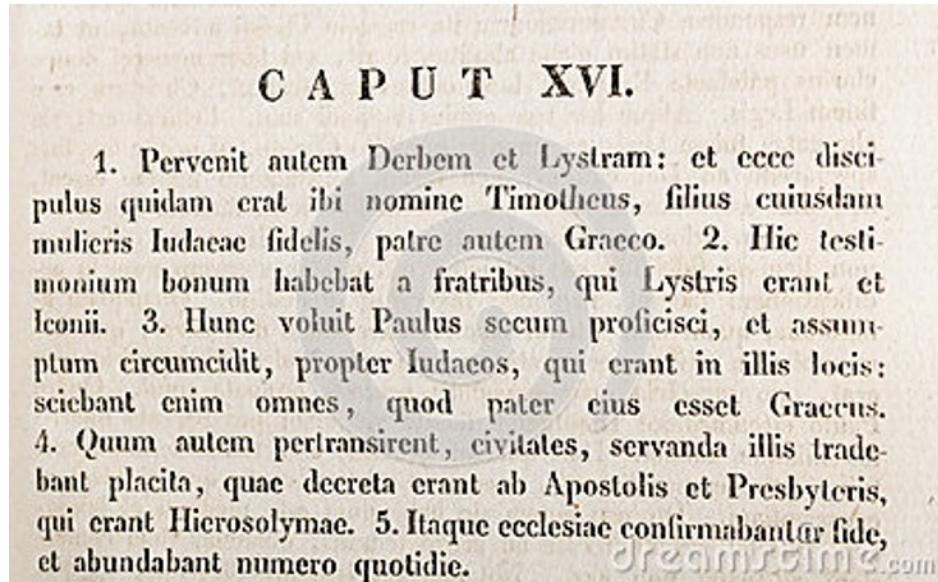
- Understanding – get the “gist” of a document
- Grouping – cluster for overview or classification
- Compare – compare document collections, or
- Inspect evolution of collection over time
- Correlate – compare patterns in text to those in other data, e.g., correlate with social network

## Levels of Text Representation

- Lexical level, transforming a string of characters into a sequence of atomic entities, called tokens.
- Syntactic level, identifying and tagging (anotating) each token's functions.
- Semantic level, extracting of meaning and relationships between pieces of knowledge derived from the structures identified in the syntactical level.

# Fundamental of Text Visualisation

Be warn, not all text are written in English and in digital forms!



# Text Visualisation Methods

- Tag Cloud
- Wordle
- Word Tree
- Phrase Nets

# Tag Cloud

- A tag cloud (word cloud, or weighted list in visual design) is a visual representation for text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text.
- 'Tags' are usually single words, normally listed alphabetically, and the importance of each tag is shown with font size or color.



Source: [Tag cloud](#)

## Application of Tag Cloud I: Branding

- One-word tag cloud of DBS's corporate values statement created using Many Eyes.

A tag cloud visualization showing the frequency of words from DBS's corporate values statement. The words are represented as colored rectangles, with larger rectangles indicating higher frequency. The colors range from light blue to dark red. The most prominent word is 'dbs' (blue), followed by 'business' (dark red), 'choose' (dark red), 'corporate' (dark red), 'customers' (dark red), 'heart' (dark red), 'people' (blue), and 'values' (dark red). Other visible words include 'bank', 'change', 'customer', 'culture', 'customer', 'embed', 'embrace', 'empowerment', 'enterprising', 'entrepreneurial', 'expand', 'financial', 'firm', 'full', 'goal', 'goals', 'grow', 'guided', 'helped', 'improve', 'individual', 'industries', 'intellectual', 'kindness', 'lead', 'long-term', 'make', 'mindset', 'modern', 'mutual', 'organisation', 'partnerships', 'pride', 'putting', 'region', 'respect', 'reward', 'sector', 'sense', 'service', 'set', 'short', 'standards', 'status', 'strive', 'strong', 'technology', 'time', 'transform', 'treat', 'true', 'trust', 'ultimately', 'understanding', and 'uphold'.

access accountable act asian asset attitude bank banking behave biggest business  
can-do capital change choose citizens common community corporate courage  
culture customer customers dbs delight develop difference dignity distinctive  
embed embrace empowerment enterprising entrepreneurial expand financial firm full goal  
goals grow guided heart helped improve individual industries intellectual kindness lead  
long-term make mindset modern mutual organisation partnerships people  
people-driven performance positive potential pride putting quo realise reasons recognise  
region respect responsible reward sector sense service set short standards status strive  
strong technology time transform treat true trust ultimately understanding uphold  
values

## Application of Tag Cloud I: Branding

- Two-word tag cloud of DBS's corporate values statement created using Many Eyes.



A two-word tag cloud where the size of each word indicates its frequency or importance in DBS's corporate values statement. The words are arranged in a cluster, with larger words like 'strong' and 'customer' being the most prominent.

accountable embrace asian service attitude strong banking sector  
biggest asset can-do attitude common goals corporate citizens  
corporate values customer mindset customers choose  
dbs performance delight customers distinctive dbs embrace change  
firm values full potential intellectual capital long-term partnerships  
modern technology mutual trust people-driven business  
performance culture positive can-do realise full recognise long-term  
responsible corporate service standards status quo strong sense  
time develop understanding reward

# Wordle

- A toy for generating “word clouds” from text that you provide

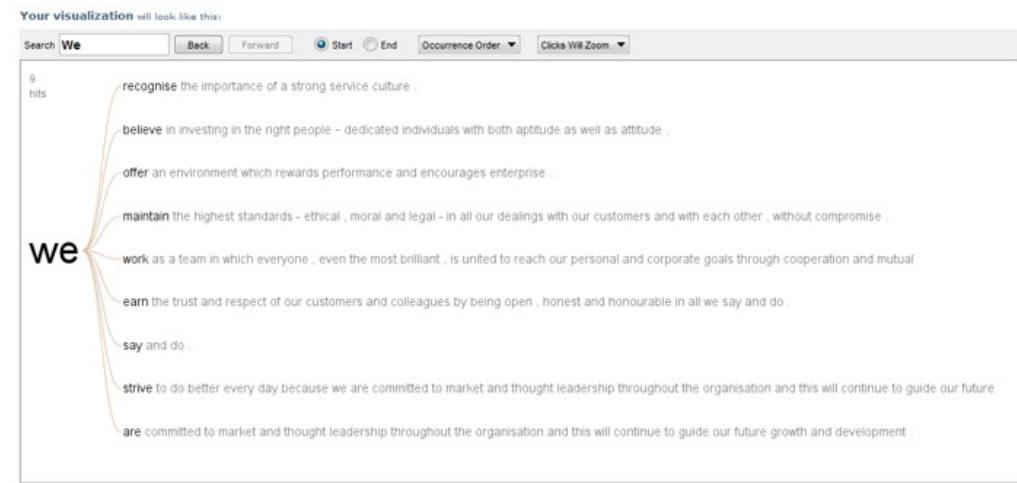


# Word Clouds of Corporate Values Statements



# Word Tree

- A visual search tool for unstructured text, such as a book, article, speech or poem. It lets you pick a word or phrase and shows you all the different contexts in which the word or phrase appears.
- The contexts are arranged in a tree-like branching structure to reveal recurrent themes and phrases.



Link: <https://www.jasondavies.com/wordtree/>

# Phrase Net

- A phrase net diagrams the relationships between different words used in a text. It uses a simple form of pattern matching to provide multiple views of the concepts contained in a book, speech, or poem.

## Customizing Phrase Net

Data set: Journey Back to Singapore (Version 1)

Your visualization will look like this:

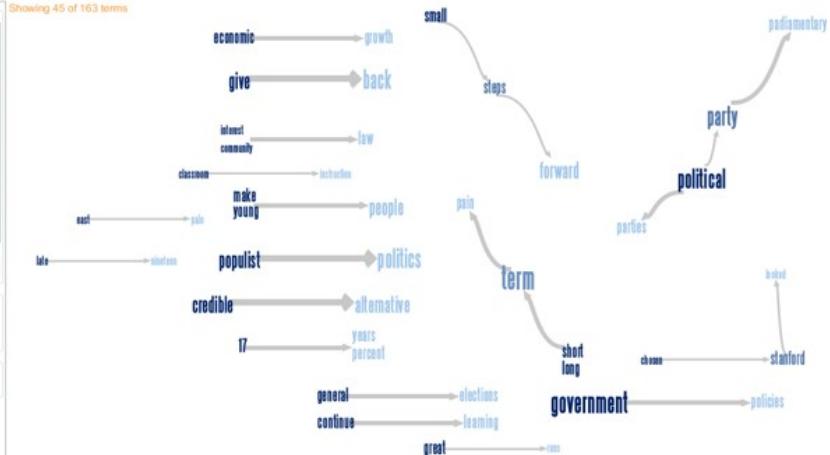
Select a phrase

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  
= =

Filters  
Show top: 50  
Hide common words

Zoom  
In  Out  Reset

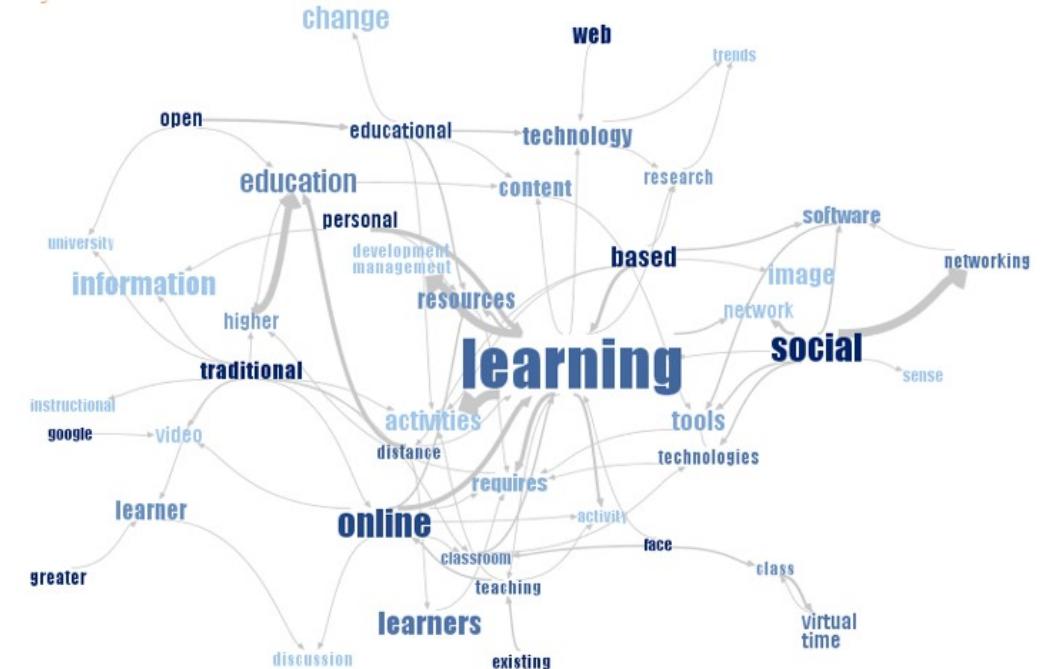


# Phrase Net

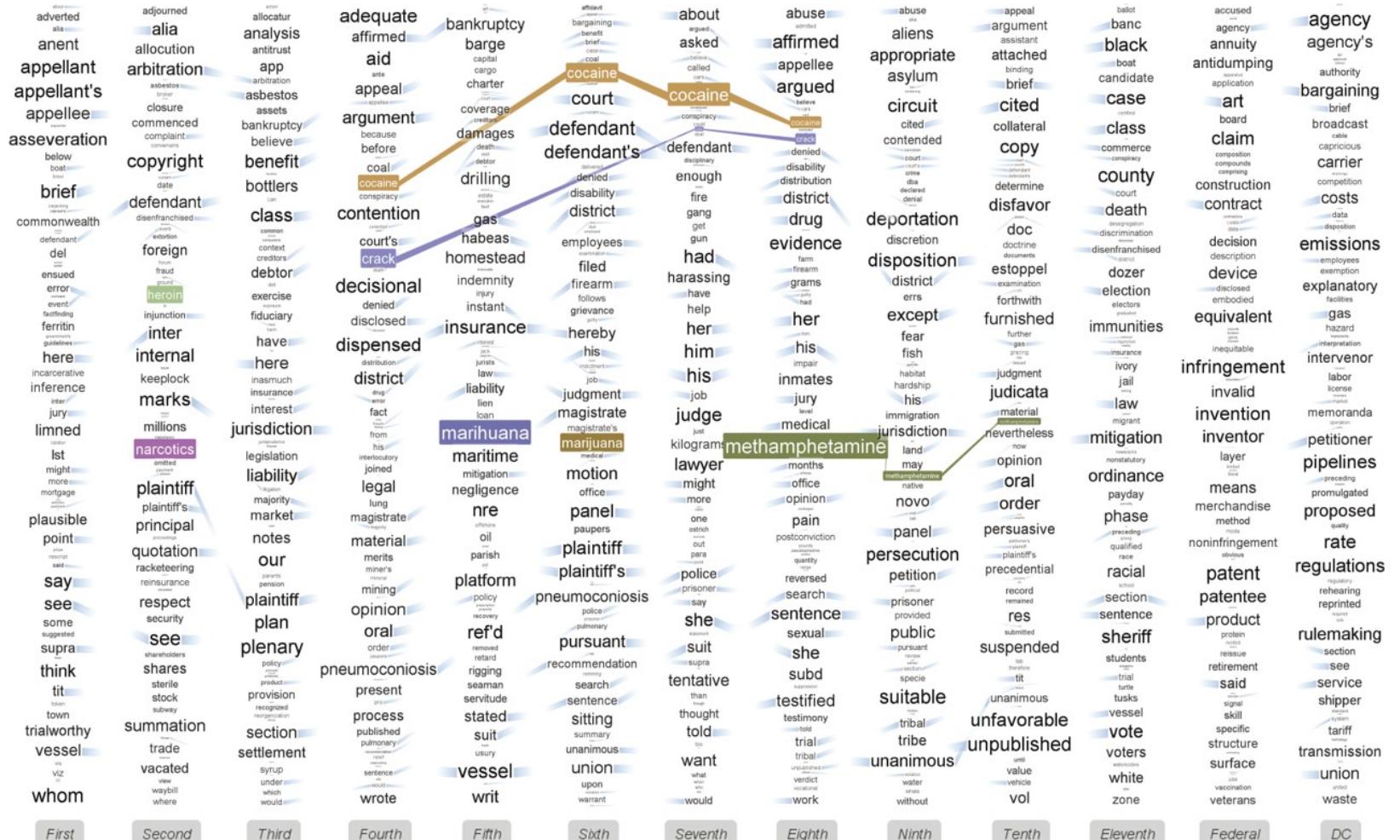
Words separate by the keyword "and"



## Words that directly follow one another

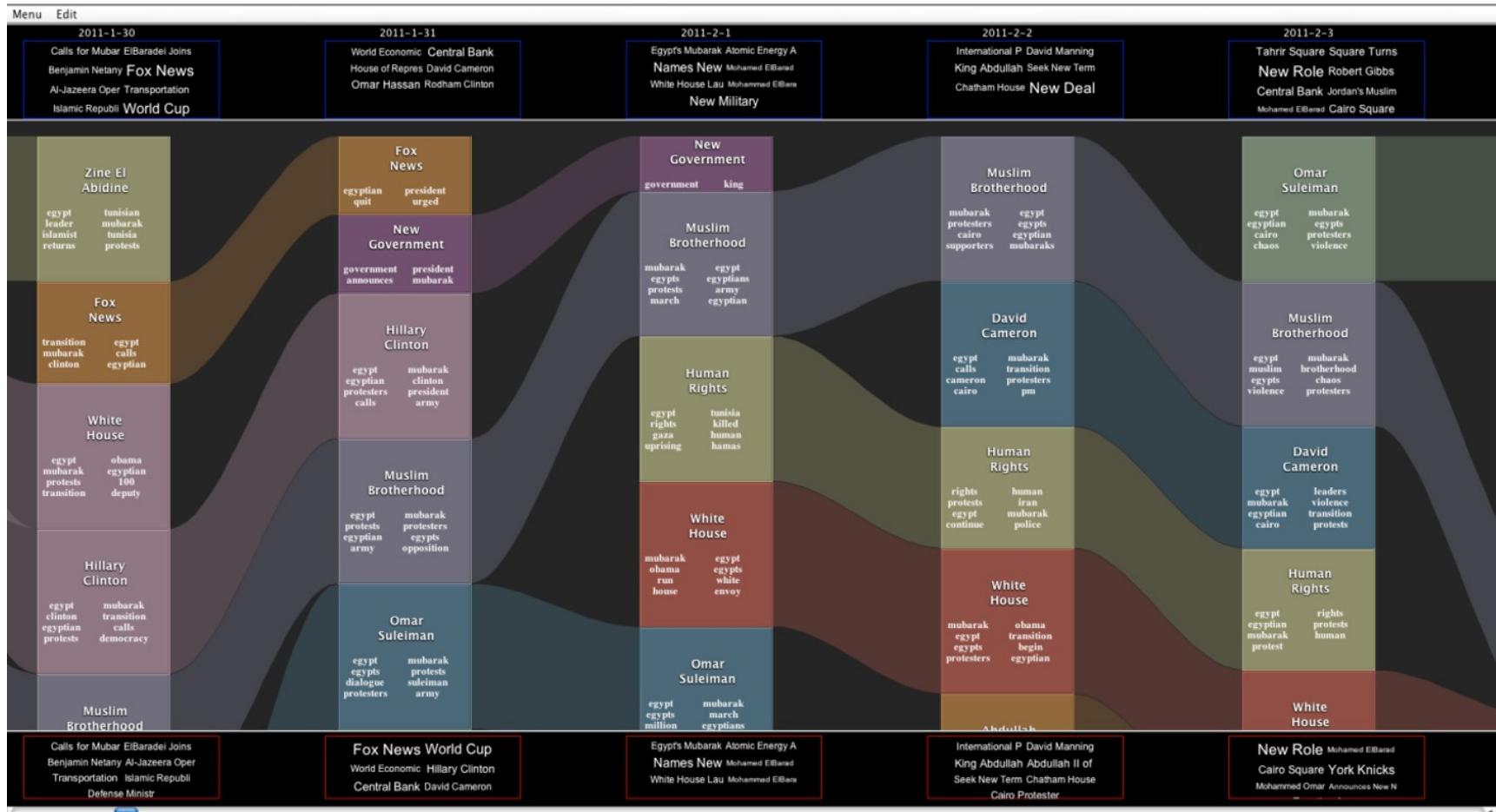


# Parallel Tag Cloud



Reference: Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora

# Story Tracker: Main View



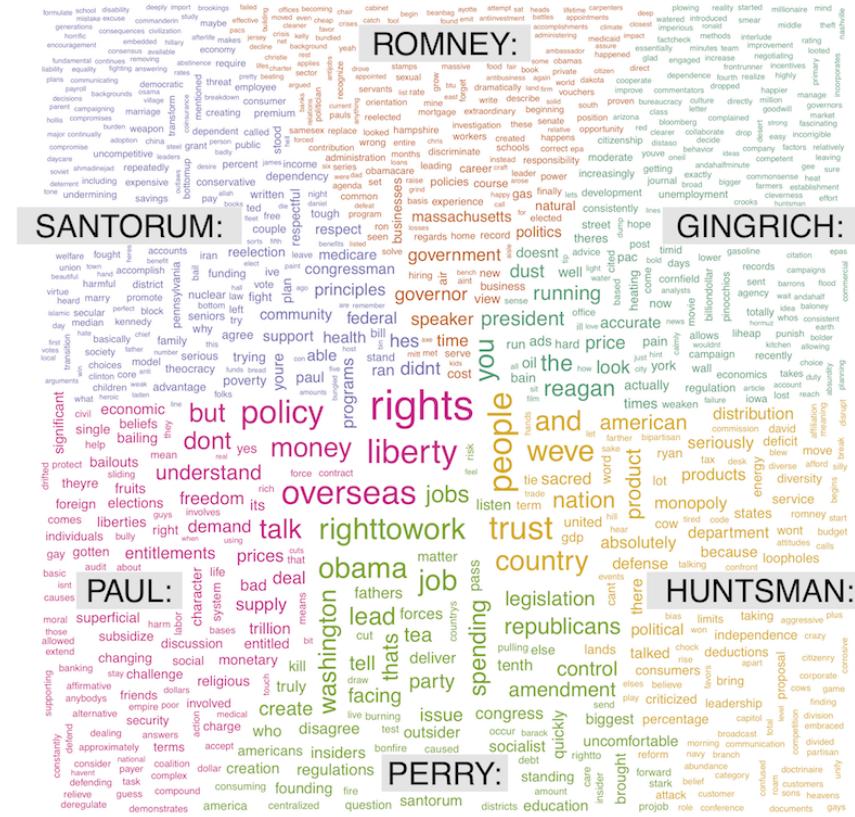
## Reference: Story Tracker: Incremental visual textanalytics of news story development

# R packages for Text Visualisation

- ggwordcloud: a word cloud geom for ggplot2
- TextPlot: R Library for Visualizing Text Data

# wordcloud

- Provides functionality to create pretty word clouds, visualize differences and similarity between documents, and avoid over-plotting in scatter plots with text.
  - Visit this [link](#) for more information.



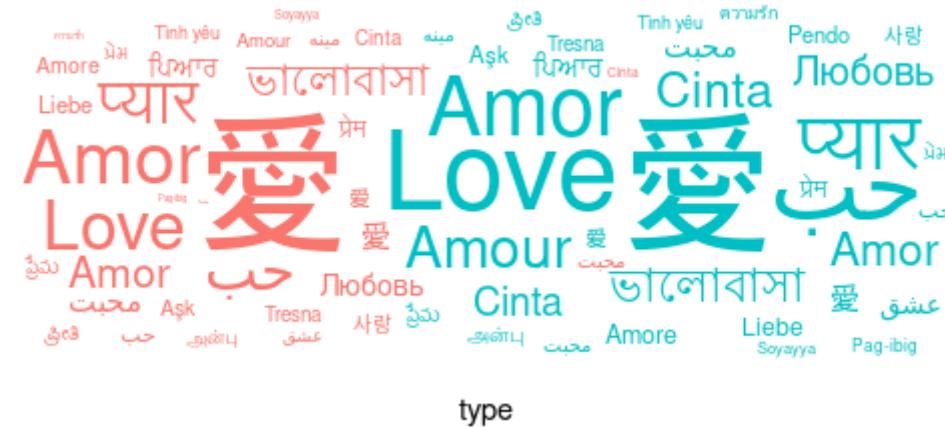
# wordcloud2: Create Word Cloud by 'htmlwidget'

- A fast visualization tool for creating wordcloud by using [wordcloud2.js](#), ia JavaScript library to create wordle presentation on 2D canvas or HTML.
  - It provides Shiny functions.
  - Visit this [link](#) for more information.



# ggwordcloud: a word cloud geom for ggplot2

- `ggwordcloud` provides a word cloud text geom for ggplot2.
  - as an alternative to `wordcloud` and `wordcloud2`.



# Wordcloud on Shiny

Shiny from R Studio

[Back to Gallery](#)

## Word Cloud

**Choose a book:**

**Change**

---

**Minimum Frequency:**



1      12      50

1 6 11 16 21 26 31 36 41 46 50

**Maximum Number of Words:**



1      300

1 31 61 91 121 151 181 211 241 271 300



global.R server.R ui.R

```
fluidPage(  
  # Application title  
  titlePanel("Word Cloud"),  
  
  sidebarLayout(  
    # Sidebar with a slider and selection inputs  
    sidebarPanel(  
      selectInput("selection", "Choose a book:",  
                 choices = books),  
      actionButton("update", "Change"),  
      hr(),  
      sliderInput("freq",  
                 "Minimum Frequency:",  
                 min = 1, max = 50, value = 15),  
      sliderInput("max",  
                 "Maximum Number of Words:",  
                 min = 1, max = 300, value = 100)  
    )
```

```
# Show Word Cloud
mainPanel(
  plotOutput("plot")
)
)
```

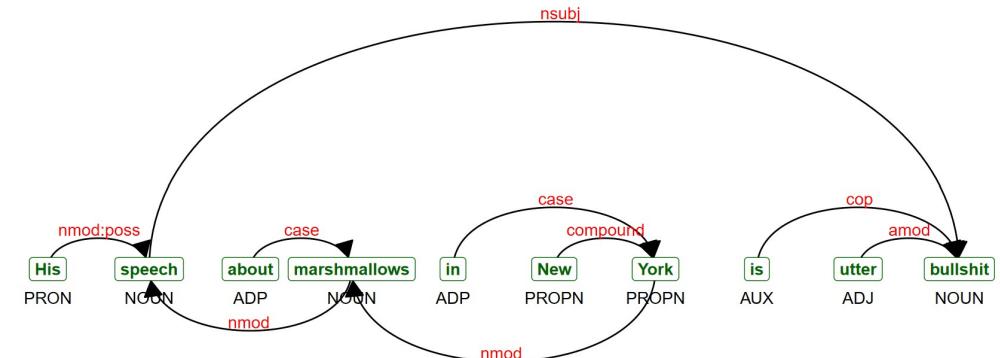
For live demo, visit this [link](#)

# TextPlot: R Library for Visualizing Text Data

- Aims to visualise complex relations in texts.
- Provides functionalities for displaying text co-occurrence networks, text correlation networks, dependency relationships as well as text clustering.
- Visit this [link](#) for more information.

## Dependency Parser

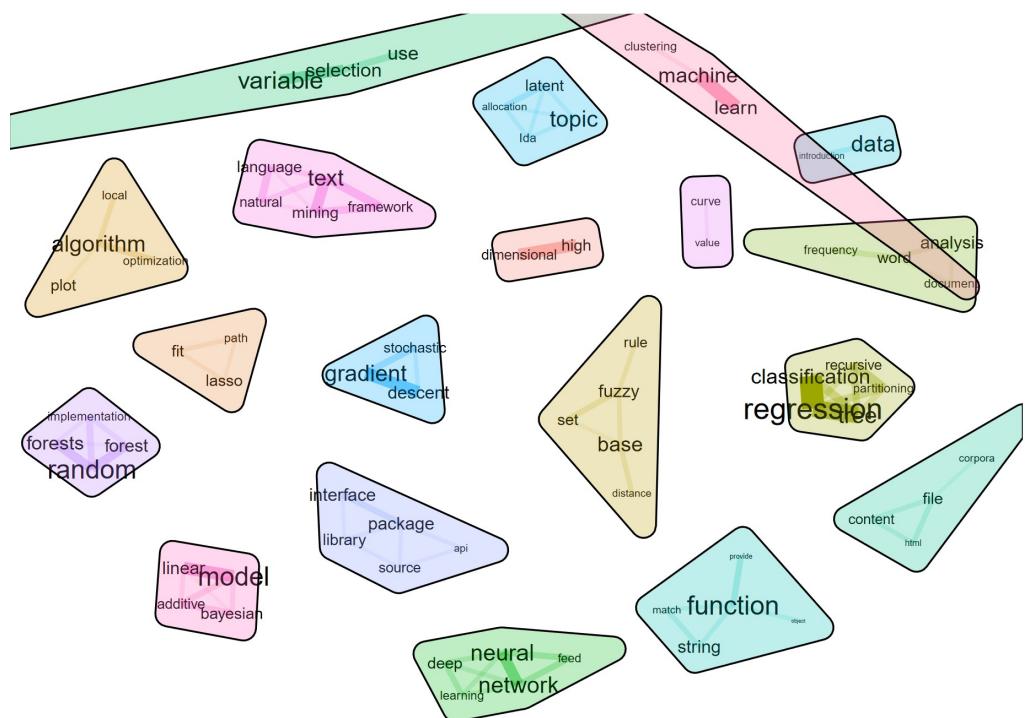
- This example visualises the result of a text annotation which provides parts of speech tags and dependency relationships.



# TextPlot

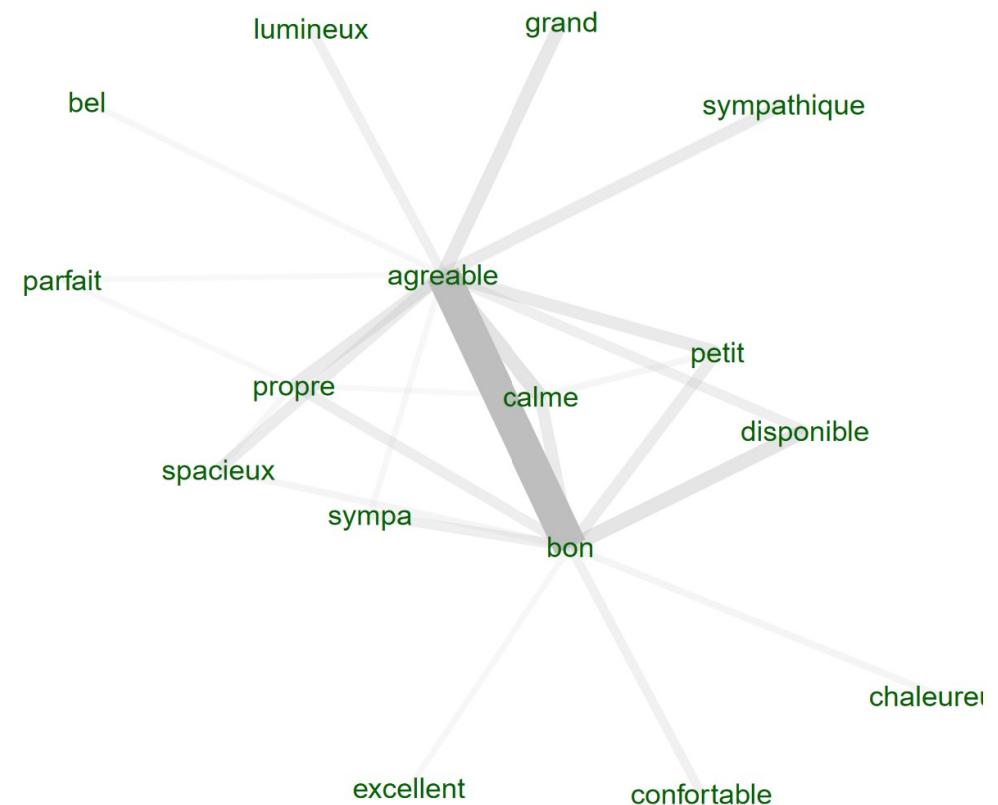
## Biterm Topic Model plots

- This example shows plotting a biterm topic model which was pretrained and put in the package as an example.



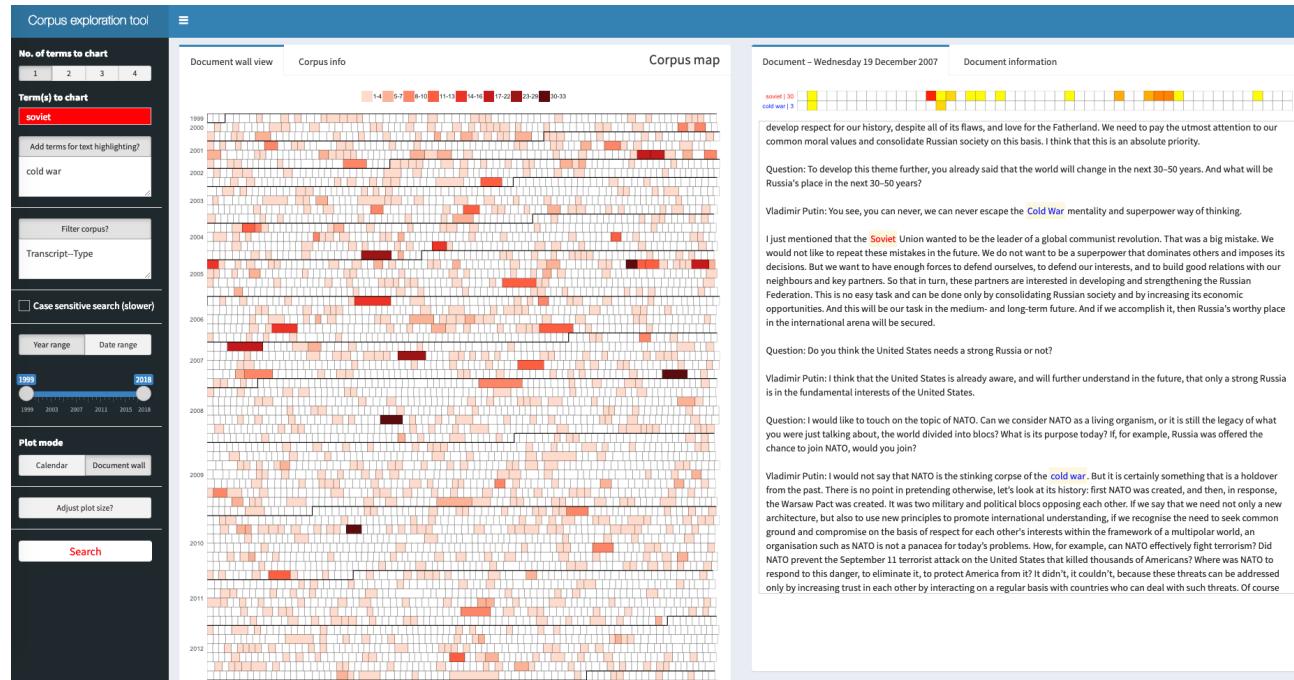
## Co-occurrence of texts

- The following graph shows how frequently adjectives co-occur across all the documents.



# corporaexplorer: An R package for dynamic exploration of text collections

- It is an R package that uses the Shiny graphical user interface framework for dynamic exploration of text collections.
- Its intended primary audience are qualitatively oriented researchers who rely on close reading of textual documents as part of their academic activity.

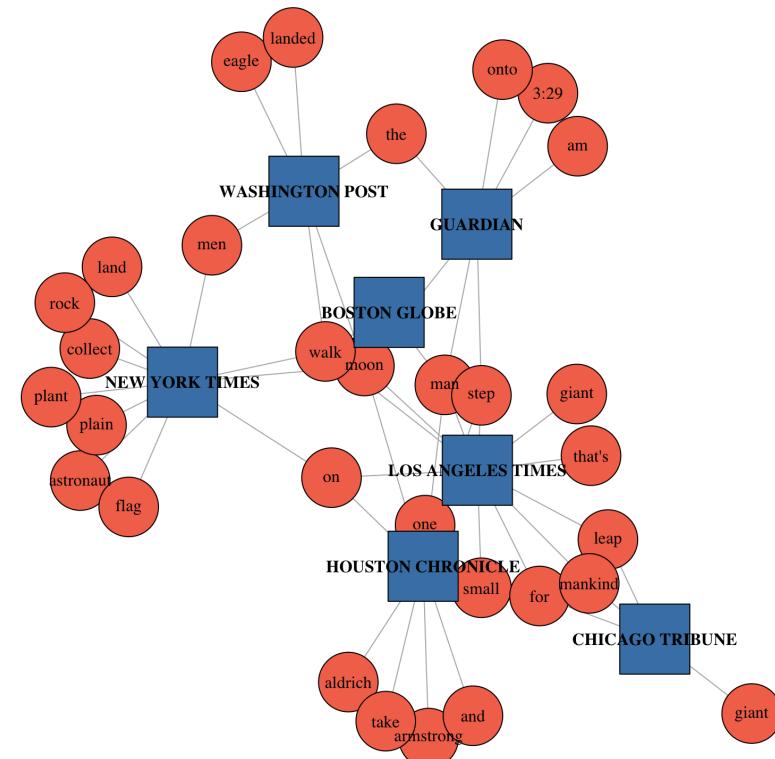


Visit this [link](#) for more information.

## textnets

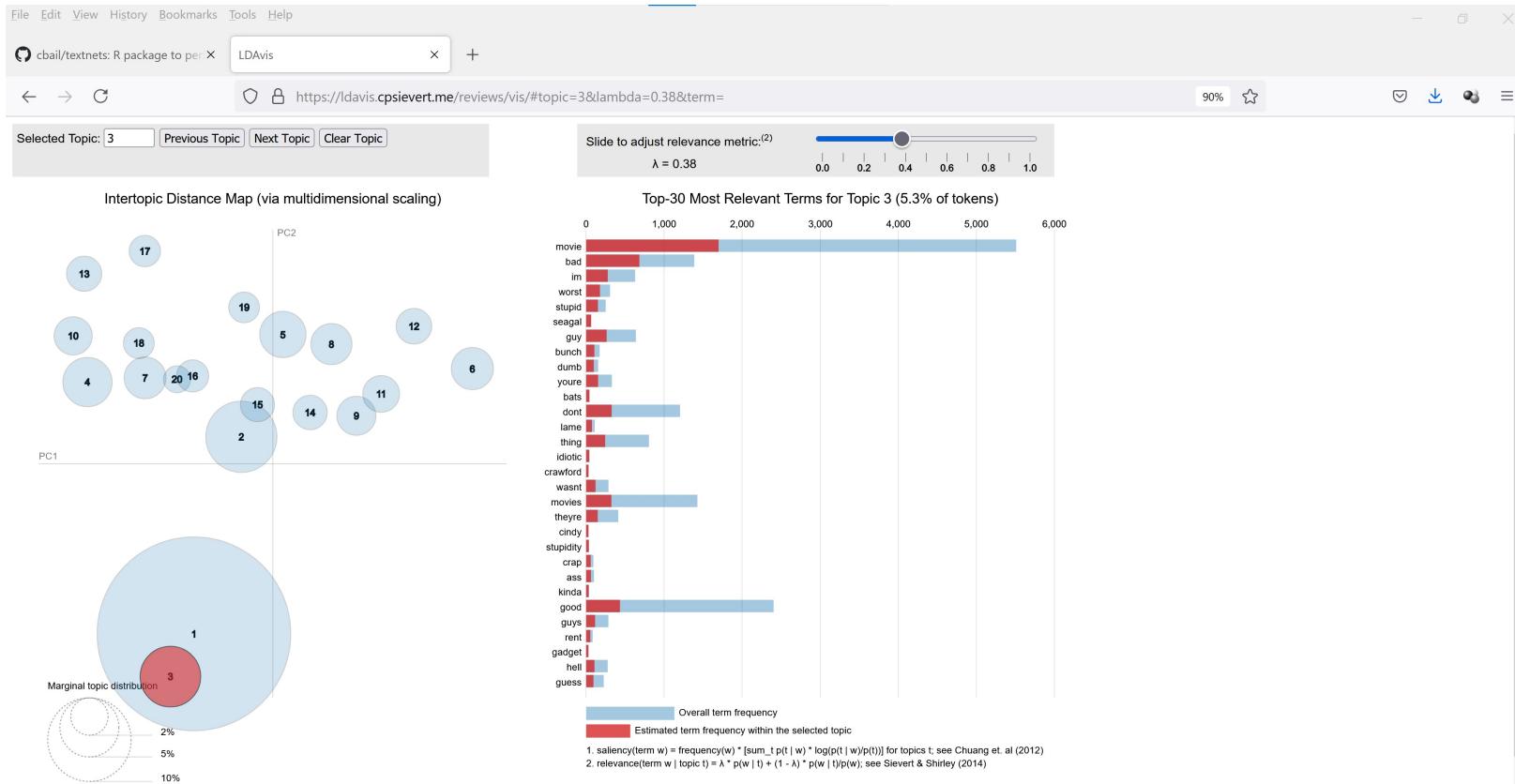
- R package for automated text analysis using network techniques.
- Visit the [github](#) repository for more information.
- Notice that this package is not on cran yet. You need to install it by using the code `install_github("cbail/textnets")`.

Reference: Bail, Christopher A. (2016) "Combining Network Analysis and Natural Language Processing to Examine how Advocacy Organizations Stimulate Conversation on Social Media." *Proceedings of the National Academy of Sciences*, 113:42.



# LDAvis

- R package for interactive topic model visualization.
- Visit the [github](#) repository and [cran](#) for more information.



Source: For live demo, visit this [link](#).

## References

Cao, Nan and Cui, Weiwei (2016) **Introduction to text visualization**, Springer. This book is available at smu [e-collection](#).