# Text Visualization

Mohan Raj Rajamanickam

CPSC 533c Presentation

Text Visualization

CPSC 533c Presentation

Mohan Raj Rajamanickam

CPSC 533c Presentation

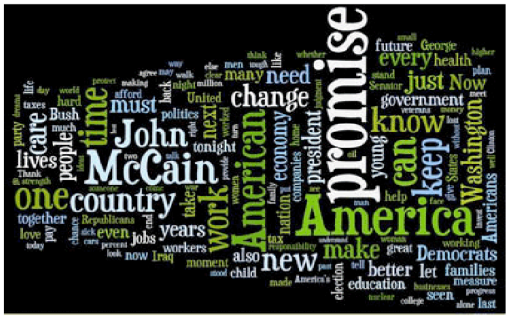
Text Visualization

Mohan Raj Rajamanickam

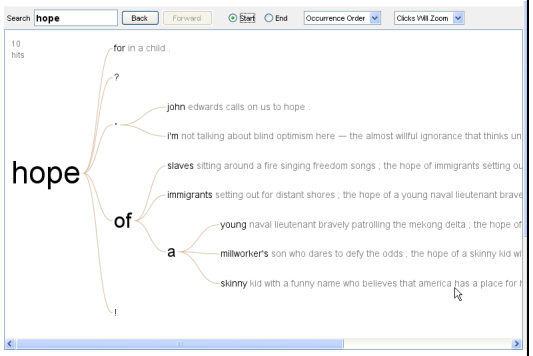
## Wordle

- Improvement over Tagclouds
  - Effectiveness of Tagclouds
- wordle = ¿Tagclouds
  - aesthetic combination of layout, type and colors

Tag Cloud of Barack Obama's speech at the Democratic Convention in 2008

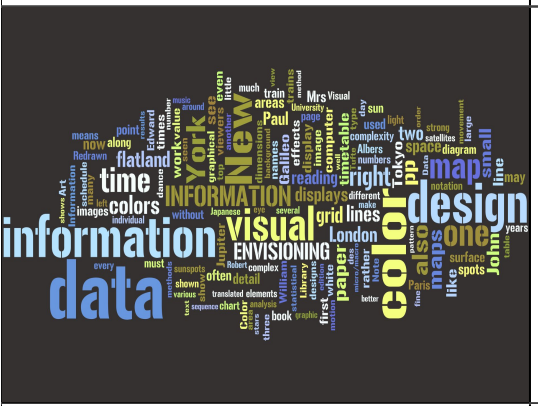


Wordle of Barack Obama's speech at the Democratic Convention in 2008



Wordtree of Obama's speech

## What is Wordle useful for ?



## What is Wordle useful for ?

- Gives a good Overview
  - as reminders, reinforces existing knowledge, as 'memento' (~50% users of wordle)
  - as learning aids
    - used by educators
- Emotional visualization !
  - marriage vows

## What is Wordle not useful for ?

- Standalone learning tool
  - prior knowledge required
- Discovering new information
- Interactive visualization

## Design implications of Wordle

- Creating mass visualizations
  - Low cost of entry
  - Portability or Sharing
  - Aesthetics ~= Usability

## Types of Text Visualizations

- Visualizing relationships among huge document collections
  - Targetted more at experts
  - Discovering previously unknown information is key
- Visualizing structure of a document
  - Unlikely to discover new knowledge
  - act as aids to reinforce existing knowledge
- Visualizing literature and citation relationships

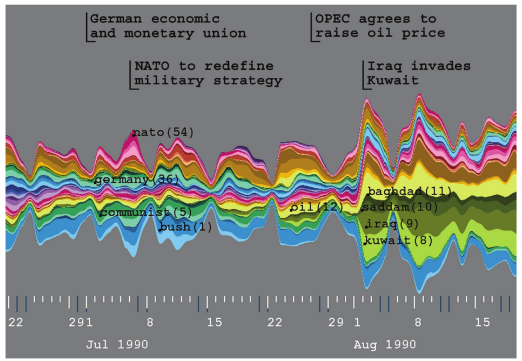



Fig. 3. ThemeRiver visualization of Associated Press news wire stores from July and early August 1990.

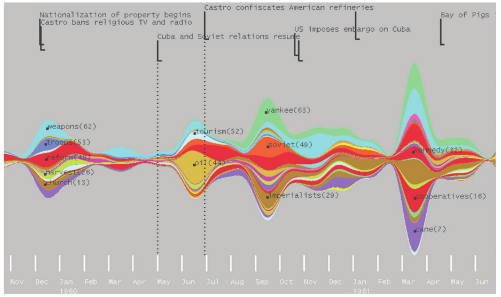
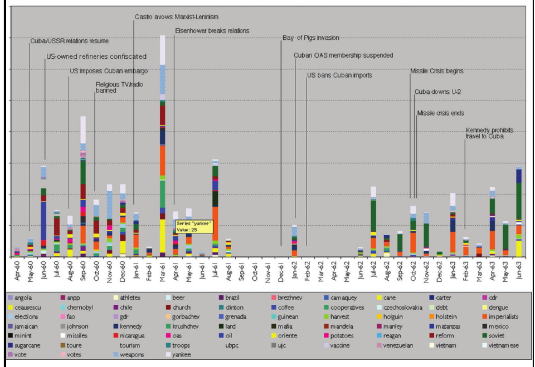
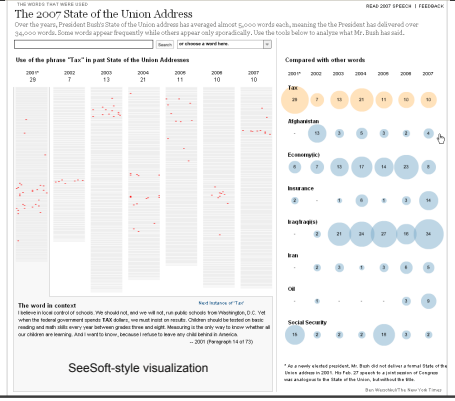


Fig. 2. ThemeRiver uses a river metaphor to represent themes in a collection of Fidel Castro's speeches, interviews, and articles from the end of 1959 to mid-1961.






SeeSoft-style visualization

# Themeriver

- can get good overview with help of markers

Drawbacks:

- not much useful without prior knowledge on the topics/markers
- scalability of number of topics - reuse of colors
- less avenues for user interaction
  - high cost of processing
- evaluation with 2 users

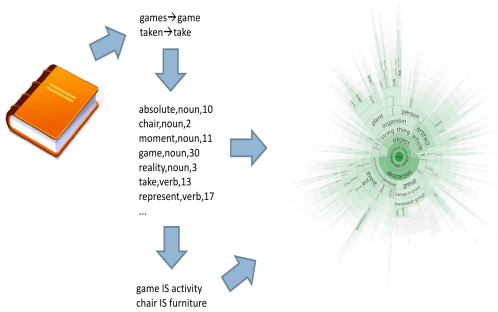
# Docuburst

- What if
  - tagclouds + lexical semantic hierarchy = ?

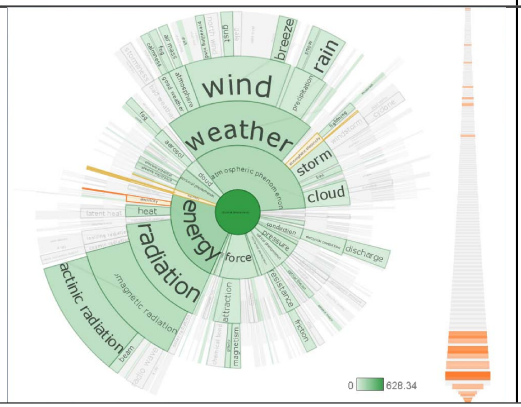
# Wordnet

- Sets of synonyms
  - {bank, financial institution}
  - {bank, slope, riverside}
- Sets are connected through IS-A relationship
  {robin, redbreast} IS-A
  {bird} IS-A
  {animal, animate_being} IS-A
  {organism, life_form, living_thing} IS-A
  {entity}

Source: Eurovis Docuburst presentation




Source: Eurovis Docuburst presentation




# Docuburst

- Answers the question: 'What is this document about ?'
- Goal: to bring the exploratory experience of physical libraries to digital libraries
  - Idea of small multiples: library interfaces could be enhanced with arrays of Docuburst icons
- Drawbacks:
  - Higher cognitive load, as well as system load

# Conclusion

- Nominal, categorical variables
  - difficult to visualize
    - no inherent ordering
- Text
  - categorical nature +
  - high dimensionality
  - = challeging to visualize

# References


- DocuBurst: Visualizing Document Content using Language Structure. Christopher Collins, Sheelagh Carpendale, Gerald Penn. EuroVis 2009.
- Participatory Visualization with Wordle. Fernanda B. Viegas, Martin Wattenberg, Jonathan Feinberg. InfoVis 2009.
- ThemeRiver: Visualizing Thematic Changes in Large Document Collections Susan Havre, Elizabeth Hetzler, Paul Whitney, Lucy Nowell. IEEE Transactions on Visualization and Computer Graphics, 8(1), pp 9-20, 2002.
- SEARCH USER INTERFACES, PUBLISHED BY CAMBRIDGE UNIVERSITY PRESS. COPYRIGHT © 2009 BY MARTI A. HEARST. http://searchuserinterfaces.com/book