



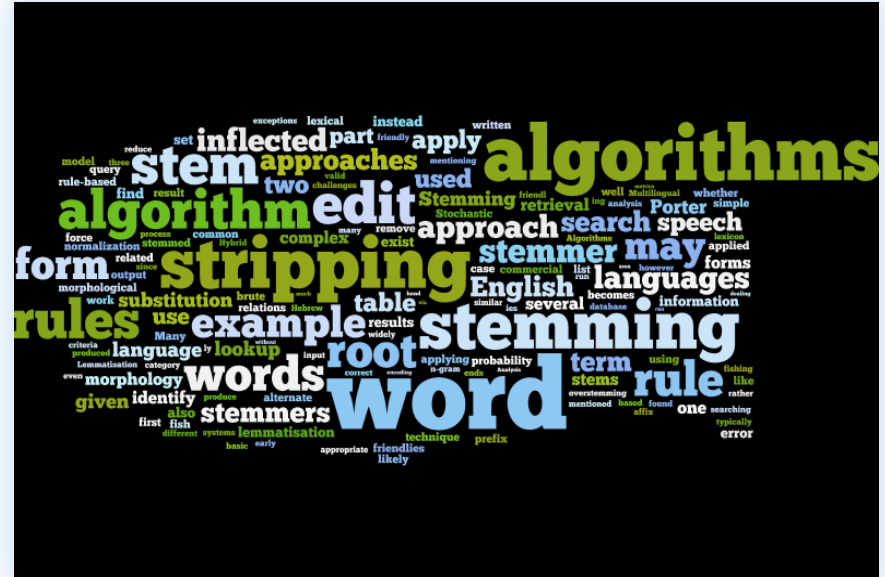
Visualizing Text

Gilbane Conference

Carl Lambrecht; VP, Engineering, Lexalytics, Inc.

carl.lambrecht@lexalytics.com / @cjlambre

Which word is gone?



How about now?

| | | | |
|-------------|----|------------|----|
| • stem | 86 | • language | 20 |
| • word | 53 | • edit | 20 |
| • algorithm | 49 | • example | 18 |
| • rule | 36 | • root | 18 |
| • suffix | 27 | • apply | 14 |
| • strip | 23 | • search | 13 |
| • approach | 21 | • inflect | 12 |
| • form | 21 | • english | 10 |

| | | | |
|-------------|----|-----------|----|
| • stem | 86 | • edit | 20 |
| • word | 53 | • example | 18 |
| • algorithm | 49 | • root | 18 |
| • rule | 36 | • apply | 14 |
| • strip | 23 | • search | 13 |
| • approach | 21 | • inflect | 12 |
| • form | 21 | • English | 10 |
| • language | 20 | • part | 10 |

Visualization vectors

Associated Metadata

- Language
- Geography
- Time
- Publication
- Author/@handle
- Socioeconomic
- Social associations

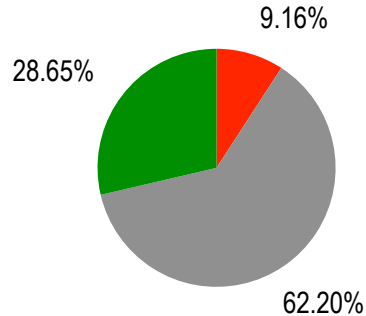
Content Derived

- Stemmed Words/Words/
Phrases
- Part-of-Speech
- Extracted Features
 - Entities
 - Themes
 - Topics
 - Intentions
- Sentiment/Emotions

Filters and Grouping

- Weighting factors
 - Counts
 - “Importance”
- Similarity
- Co-occurrence
 - Categories
 - Other words

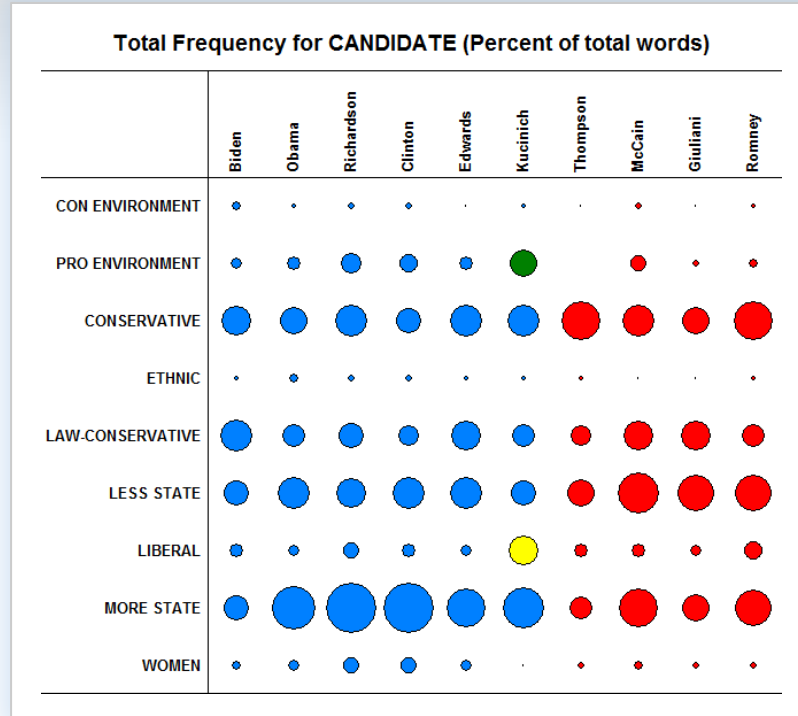
Pies (one axis)



Positive: 28.65%
Negative: 9.16%
Neutral: 62.20%

For any more than 3 data points pie charts become increasingly hard to read.
If you have 3 or fewer data points, why not just use a table?

Bubbles (4 axes)



Courtesy of Provalis Research

Stemmed Words vs. Words vs. Word Phrases vs. Relationships

- I was greatly satisfied with my dinner.

- Greatly satisfied

- Greatly

Satisfied(x1.5) → dinner

- Great

- I hate the cracked screen on my phone.

- Cracked screen

- Crack

Cracked Screen → phone

Stemming vs. Lemmatization

Stemming

Walking → Walk

Better → Better

I am meeting him tomorrow

Meeting → Meet

In our last meeting, we...

Meeting → Meet

Lemmatization

Walking → Walk

Better → Good

I am meeting him tomorrow

Meeting → Meet

In our last meeting, we...

Meeting → Meeting

Examples from Wikipedia

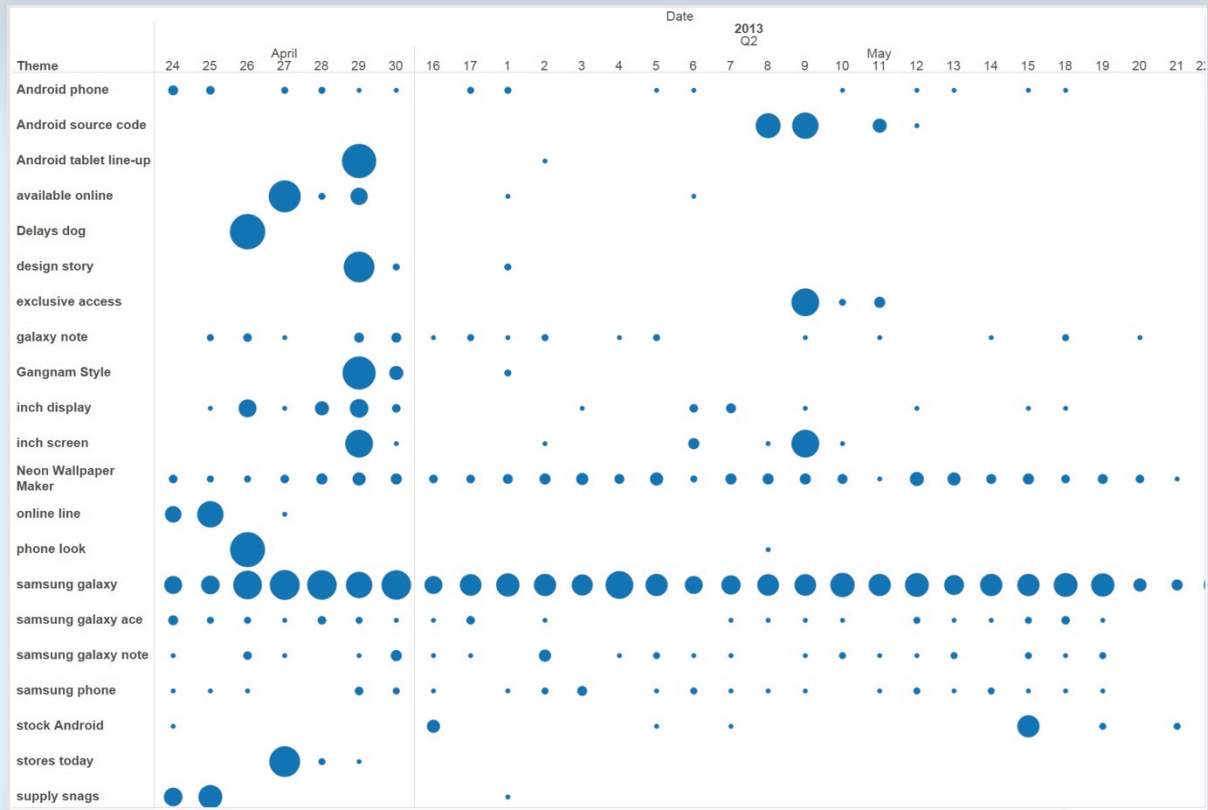
Top themes + relative occurrence



Plus Sentiment



+Time



Summary

- Don't use word clouds if you can avoid them.
- Really don't use word clouds for any sort of comparison over time.
- If you're going to use word clouds
 - use intelligent colors
 - use them either as a user-interface
 - or use them when you've already done a bunch of filtering
- Don't use pie charts when details are important, use tables instead.
- Many other chart types, such as bubble charts, have the visual appeal of word clouds while providing more information.

And check this out...

http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen

LEXALYTICS

Read  Between The Lines