

# **DYNAMIC TWITTER DATA ANALYSIS**

---

Rahul Vunnam, Krishna Reddy, Volodymyr Korshak

# DYNAMIC DATA ANALYSIS

---

- ✘ Discovering relationships between data is considered to be a relatively new field of computer science.
- ✘ The need for uncovering connections and structuring existing data appeared with growing amounts electronic of data in the world.
- ✘ What is the overall reception of the new Iphone? What topics are associated with “Samsung tablet”? Which words do people use when talking about “Barack Obama”?

# TECHNOLOGIES USED

---

- ✘ Python
- ✘ Twython (python Twitter API wrapper)
- ✘ JavaScript and JSON
- ✘ D3js Visualization Library

# THE FLOW

---

- ✘ First, query the Twitter APIs for all tweets containing the requested word using Twython.
- ✘ Second, parse the JSON results for words, hash-tags, etc.
- ✘ Third, analyse the results and compute statistical information.
- ✘ Forth, render the results on screen using D3js.

# PARSING AND ANALYZING DATA

---

- ✘ Parsing by definition means Syntactic analysis or search for a specific part of text that matches the given pattern. One of the most popular ways to do text-parsing is using Regular Expressions.
- ✘ Analysing twitter messages could uncover an immense amount of information. This makes it possible to answer more and more questions about people's thoughts at any given time.

# TWEET PARSING EXAMPLE



**applenws** Apple News

Hunting Park Man Robbed of iPod and iPhone - NBC  
Philadelphia [bit.ly/vxsBtr](http://bit.ly/vxsBtr) [#apple](#)

18 hours ago ☆ Favorite ↻ Retweet ↩ Reply

"`\s?(\\w+)\s?`" pattern matches for words present in the tweet. '/s?' matches white space characters if they exist. '/w+' matches any amount of word characters

"`(http:[^\\s]+)`" pattern works same way as the one above, it searches for links starting with 'http:'. By placing '^' before '/s' we specify that there should not be any white spaces.

"`([#]{1}\\w+)`" pattern searches for Strings starting with '#'. By placing '{1}' after '#' symbol we specify that '#' should occur only once.

# DATA VISUALIZATION

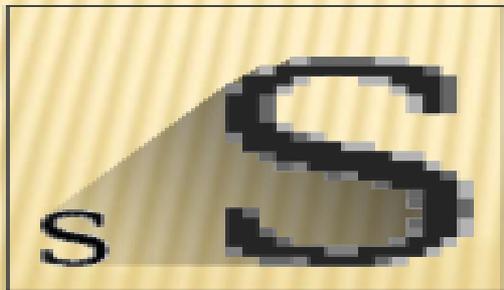
---

- ✘ What kind of information can we get from analysis?
- ✘ Popularity of words used in a specific country or in a specific topic
- ✘ Message analysis uncovers connections between words and topics. However, it is impossible to represent all this using one single graph
- ✘ What is the clearest way to show connection between topics?

# D3JS

---

- ✘ D3js is a JavaScript visualization library which uses Scalable Vector Graphics (SVG) to represent data in various ways.
- ✘ SVG concept helps to facilitate zoom-in, zoom-out functions.



**BITMAP**  
-jpeg .gif .png



**OUTLINE**  
-svg

# D3JS

---

- ✘ D3JS gets the data from the backend in JSON format .
- ✘ It Identifies nodes and edges in the data.
- ✘ Creates the document with nodes and edges using SVG.

# CONCLUSION

---

- ✘ Dynamic Data Analysis is a powerful tool to uncover hidden connections between various things. It is closely connected to Big Data Analytics and becomes widely used in everyday life.

**THE END**

---

**Thank You**

Feel free to ask Questions