Visualizing social media content with SentenTree

Mengdie Hu, Krist Wongsuphasawat, and John Stasko
Visualizing textual content

• Keywords
  • Text cloud, Wordle, semantic word cloud, parallel tag cloud, etc.

• Text structures
  • Word Tree, Double Tree, etc.
  • PhraseNet
Wordle, Viegas et al. 09

Seam-carving word cloud, Wu et al. 11

Parallel tag cloud, Collins et al. 09
love the

Phrased net, Van Ham et al. 09

Word tree, Wattenberg and Viegas 08
hey Jude

make it bad
be afraid
let me down

don’t

remember to
let her into your heart
let her under your skin

then you
can start
begin

better better better better better waaaaa

na
Recurring segments collapsed and larger

Non-recurring segments aligned

http://loveallthis.tumblr.com/post/166124704
Design goals of SentenTree

• Highlight topics with keywords
• Show keywords in context
• Coverage and conciseness
  • Show fragments instead of full sentences
  • Remove stopwords and punctuations
  • Collapse common fragments
• Summary and Details
  • Show original context on demand
  • Enable drill-down
SentenTree visualization demo
Building a SentenTree vis
Constructing frequent sequential patterns

- I love the World Cup!
- RT if you are watching the World Cup today!
- Marcelo opens the World Cup scoring with this own goal to give Croatia the Lead.
- The first goal of the World Cup is an own goal by Brazil!
- This is the best goal so far in the World Cup: Marcelo’s own goal.
Constructing frequent sequential patterns

- I love the World Cup!
- RT if you are watching the World Cup today!
- Marcelo opens the World Cup scoring with this own goal to give Croatia the Lead.
- The first goal of the World Cup is an own goal by Brazil!
- This is the best goal so far in the World Cup: Marcelo’s own goal.
Constructing frequent sequential patterns

- world cup (5)
- world cup own goal (3)
- goal world cup own goal (2)
Constructing frequent sequential patterns

- world cup (5)
- world cup own goal (3)
- goal world cup own goal (2)

goal – world cup – own goal
Force-directed layout with constraints

- Implemented with the CoLa package
  - [http://marv1.infotech.manash.edu/webcola/](http://marv1.infotech.manash.edu/webcola/)
- Constraints:
  - Words in a pattern: left to right order
  - Bigrams: shorten edge, center words vertically
  - Child branches: center horizontally
Performance

- Time cost
  - Pattern generation: linear to number of sentences in dataset
  - Graph layout: depends on complexity of graph
- Javascript implementation, Chrome browser, Macbook Air
- Test results:
  - 10k – 2 sec
  - 100k – 20 sec
  - 1M – 2 minutes
Observations and initial feedback
Natural content clustering

take tour safaris best new features os x yosemite

wildfire weakened evacuations end evacuations end wildfire dies near giant

yosemite national park california apple pay huge money give faster ios 8 yosemite downloads
Data analysts feedback

• Like
  • Frequent words “pop out” with frequency count
  • Context available through hovering
  • Similar content grouped together
  • Shows major topics
  • Less frequent content discoverable
Data analysts feedback

- Learn
  - Same word appears multiple times
  - Drill down, search, etc.
- Suggest
  - Improve packing algorithm to remove whitespace
  - Reduce noise
    - Stemming and spell checks
    - Filters

Show more information: meta-info on user, sentiment, etc.
Takeaway

- SentenTree is a novel visualization technique for social media text that both summarizes topics (through keywords) and shows context (through sentence fragments).
This research is supported in part by the DARPA XDATA program and the National Science Foundation via award IIS-1320537.
SentenTree: Make TextVis great again

- SentenTree is a novel visualization technique for social media text that both summarizes topics (through keywords) and shows context (through sentence fragments).