

Report: US Presidential debate analysis

Presented by :
Tanmay Thanker

1) Model your report after the case studies in Chapter 9, using Munzner's table format for summarizing your final design as that at the end of section 9.1

Answer:

System	Animated Scattered plot(Cumulative Sentiments score of the candidates)
Data type	Dataset Type: Table Attribute Type: 1 Text data (Speech) 1 Categorical data(Candidate names) 1 Ordinal data (line number)
Derived Data	Extracting sentiments score (by using sentimod.js)
View comp.	Marks: points, Lines Channel: Color hue, Motion, Position
View Constr	Scatterplot with animation

Reduction	Dynamic Filtering , Dynamic aggregation
Abstract task	Comparison of sentiment score.
Scalability	Candidate name: 4 Text data: Thousands Statements

Word Cloud:

System	Word Cloud
Data type	Dataset Type: Table Attribute Type: Text data Speaker :- Categorical
Derived Data	Extracting important keywords from speech data(job, tax, china, money)
View comp.	Marks: Text Channel: Size, Color hue.
View Constr	Word cloud

Reduction	Dynamic Aggregation, Dynamic filtering
View Coord.	Highlighting the keywords
Abstract task	Highlighting the Most spoken keyword by word size
Scalability	Thousands of words

Lexical :

System	Lexical Episodes
Data type	Dataset Type: Table Attribute Type: 1 Text data (Speech) 1 Categorical data(Candidate names)
Derived Data	Extracting important keywords from the data for our visualization and analysis
View comp.	Marks: Lines, Area Channel: Size, Color hue
Reduction	Dynamic filtering ,dynamic aggregation

View Coord.	Lexical Tree
Abstract task	Abstract view and Explaining the text in more detail.
Scalability	The whole text in abstract way

2) Describe any algorithmic challenges that you faced. Clarify how you may have adjusted the scope of your design as a result of implementation challenges.

Algorithm Specification:

Basic text processing by building Regular Expressions .for example, find all the instances of keywords in speech.

We have done the **restructuring of data** so that it can be used in our visualization. In our Lexical Tree visualisation we have filtered the data in a step way so that it will show the detailed data to the end user in a sensible way.

From debate → keyword → Speaker → Sentiments → Statement (along with highlighted keyword)

In the below screenshot we can see the level of details which we structured to show in our lexical tree visualization.



Word Normalization | Filtering | tokenization (word cloud):

- We have defined some words like adjectives ,self centric words (i, me,we)etc... should not display into visualization
- We are removing some special characters like “ , ,!,’ ..etc” from our text by using regular expression to find the frequency of the words.

Sorting of words based on count

Then we are sorting the words by frequency to catch the most used words in the speech.

3) Present validation results, including a *detailed scenario* about how to use your tool to solve the problem, and *any feedback (e.g., pros and cons) about your tool from your recruited users.*

Answer: As we can see in the below screenshots where we have demonstrated the search text in the structured way to represent the data where we can see the result. So it's shows that the detailed data representation of an overview.

From debate → keyword → Speaker → Sentiments → Statement (along with highlighted keyword)

This way we are resolving the issue of complex data understanding and keeping the keyword highlighted so that end user will able to get better understanding.

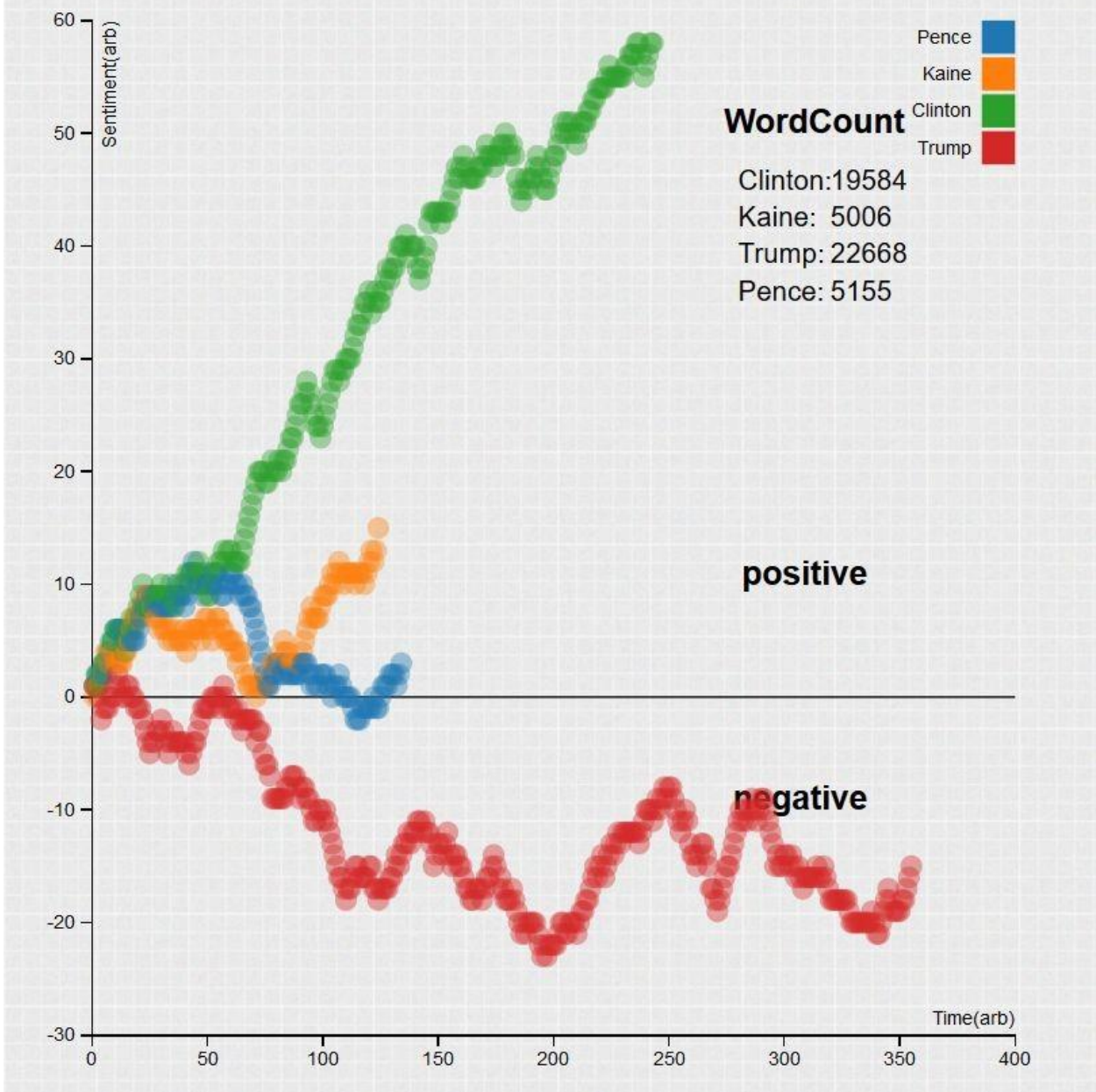


Sentiment Analysis(Scatter plot)

In this Visualization we have analyse the sentiments score pattern for whole debate for this we arbitrarily assigned a score of +1 for positive statement and +2 for very positive statement , -1 for negative statement and -2 for very negative statement and finally 0 for neutral statement.

For X axis we are considering a line number for whole debate and on Y axis we have sentiment score.

Sentiment Analysis (Animation)



Word Cloud:

This visualisation shows the word usage by each candidates during their speech. The blue (left side) word cloud is of Clinton and the red (right side) is of Trump. AS we can see that their are some big size words and some are small which are representing the frequencies during the speech.

[illegible]

Cons: The sentiment score is depends on **sentimod.js** which use **AFINN-111 Dictionary** which have been manually labeled a word by sentiment score from -5 to +5.

- 1) We can manually add or remove a tag which we are using as a keyword for tree.
- 2) In word cloud, If you click on any word it will automatically add to the tree search data and act as a keyword.
- 3) We used animation to find out the pattern in sentiments for the debate in the scattered plot. It also shows the word count.

Solution: All contributed equally.

As we have tested the tool with couple of our friends and get the feedback from them regarding the experience and the detailed information which they got from the visualization.

We have done some UI changes based on the feedback which we got from the users, Like changing the background color of the text in lexical tree. We have also changed the word cloud format/background.

References:

- 1) <https://www.kaggle.com/mrisdal/2016-us-presidential-debates>
- 2) <https://github.com/d3/d3/wiki/Gallery>
- 3) <https://github.com/soops/sentimood>