

Report

HW 7

Environment:

- Bokeh 0.12.7
- Python 3.6

Please note:

- Run bokeh server via command
(bokeh serve Wiki.ipynb --show)
- The code is written entirely in Jupyter notebook. No need to external packages apart from Anaconda.
- Launch Jupyter notebook with command **"jupyter notebook --NotebookApp.iopub_data_rate_limit=10000000"**.

All observations and techniques are written down in notebook itself.

Observations based on Mid-Term and VAST challenge:

Mid-term included taking live Twitter data and analysing trends. It was done by 3 approaches- clustering, time-frequency and Wordcloud.

In the VAST challenge, wiki-edit page was given. Although the two problems involved text, they were inherently 'solved' by different approaches. VAST challenge involved looking at Wiki-users and their interactions based on edits. A graph was drawn subsequently. This was not done in Twitter data.

The structure of data, tokenization, clustering of data was done in same way for both problems. But from there on approaches, diverged.