

Weekly Progress Report

Detecting Conversations Going Viral

Matt Chapman

Week 15, 10/04 - 14/04

1 General Notes

Not the most productive of weeks. Time lost on Monday - Thursday due to university and personal commitments. However, progress has been made in the following areas:

- Generating test data is reasonably simple - the only issue of concern here is time to complete the export.
- Consensus at the moment (within the team) is that 2 months of data, limited to Dutch language, will provide the most useful data.
- Data Export format is CSV, easily customised read and processed using the Python Pandas package.
- I have formulated some interesting queries that filter the data and produce some human readable spikes that can be manually annotated.
- An epic and stories within it has been created in the Buzzcapture JIRA instance, to allow for sprint planning. Sprints at Buzzcapture are (at the moment) 2 weeks in length.
- I have started building a Mendeley bibliography to manage sources and citations through the project.
- Regarding the bibliography, I have added some more useful sources from a couple of the suggested Special Interest Groups. This will continue through the project.

2 Following Week

The following tasks are due to be completed in the next week:

- Finalisation of dataset upon agreement with supervisor
- Initial processing of dataset to filter on agreed terms

- Plotting and annotation of data set
- Build familiarity with Python data science tools. Feedback on this appreciated.

In addition to these tasks, the following issues need to be addressed regarding the project in general:

- Research questions need solidifying. I am not currently happy with the generality of the research questions I have, and would like to refine these into concrete subquestions.
- The hypothesis I have does not feel suitable. I would also like to refine this further.