

Initial Plan

Social Media Visualization Tools and Techniques

Author: Steven Oakley

Supervisor: Nick J Avis **Moderator:** Xianfang Sun

Module Code: CM0343 - 40 Credits

Project Description

This project will focus on two aspects of toolkits used to harvest, analyse and visualize social media data, such as content posted on Facebook and Twitter. Firstly, I will analyse and evaluate the usefulness of the data provided by current systems, and consider if it would be possible to further capitalize on this data to produce more detailed information. Following this, I will attempt to amalgamate a variety of different social media harvesting toolkits into a single system, which in combination with software I will develop myself, should enable me to provide a more meaningful outlook on the social data harvested.

Project Aims and Objectives

Whilst the data types and toolkits that I ultimately attempt to amalgamate cannot be confirmed until I complete the investigation stage of the project, I can state that one of my primary aims that I believe will be a recurring theme throughout the project, is my desire to lead the project toward some kind of detailed 'early warning system'. For example, if I could gather the geographical metadata from Twitter users tweeting about Floods in their vicinity, and combine it with data from a reliable source of weather information, such as the Met Office, then a more dynamic and responsive overview of the situation could potentially be produced. This could be particularly useful in less economically developed countries, whereby the current implementations for obtaining information relating to exactly which areas are most affected by an environmental crisis are less sophisticated. Following is a list of varying aims that I currently feel could be feasible, within the scope of the project.

- Investigate current data harvesting and visualization tools.
 - Analyse and compare the capabilities and accessibility of similar software toolkits in order to determine which would be most beneficial for further use in this project.
- Combine large amounts of data from numerous sources into a single database of linked files.
 - Query this data, independently or in real-time, with other sources of information to determine if there is any particularly interesting information.
 - If so, provide a way for a human user to review the information, so that in a real-life scenario, appropriate action could be taken.

Work Plan – First Term

My general work plan consists of researching, designing and implementing the program in the first term, and then using the system to capture data and analyse it appropriately in the second term. However due to the nature of the project, if an event occurs which I feel could produce useful or interesting results earlier in the year, I will attempt to record as much data I can about the incident immediately, for later analysis.

Step 1: Initial Meeting + Planning	- 03/10/12 - 10/10/12
Step 2: Research Existing Data Mining Solutions	- 10/10/12 - 17/10/12
Step 3: Begin Integrating Twitter data with Weather APIs	- 12/10/12 - 31/10/12
Step 4 - Milestone: Submit Initial Plan	- 19/10/12
Step 5: Create a draft GUI + Popup Box notifications	- 30/10/12 – 07/11/12
Step 6: Add visual data to notifications	- 07/11/12 – 14/11/12
Step 7 – Milestone: Weather Integration Complete	- 15/11/12
Step 8: Research other APIs	- 14/11/12 – 19/11/12
Step 9: Implement new API into existing program	- 19/11/12 – 06/12/12
Step 10: Produce Interim Report	- 15/11/12 – 13/12/12
Step 11: Submit Interim Report	- 14/12/12

I have produced a Gantt Chart (uploaded separately to 'Other Files') containing a visualised timeline of these steps for the first term.

Work Plan –Second Term

The primary focus of the second term will be in collecting, analysing and reviewing data, using the system I have developed. Whilst I currently plan to focus on amalgamating Twitter, location and weather information, I believe that I should be able to incorporate additional related data, which should become more apparent as the implementation stage progresses. As such, I cannot yet state clearly what information I will be analysing in the second term. Listed below are the expectations of what I will achieve in this second term, and once I am sure of what information I have available, I will create a detailed Gantt chart to visualize this information.

Weeks 1 – 3: Actively searching for and recording data from any unusual events in areas that my program can collate information from i.e. flooding.

Weeks 3 – 4: Appropriately formatting this data into human-readable formats, for human analysis.

Weeks 5 – 7: Evaluating exactly what use this information has, and potentially editing the current program state to incorporate any additional features deemed necessary.

Weeks 8 – End of Project: Producing the final report, incorporating all notes made thus far.