Analysing Twitter Posts to Discover and Review Software and Tools

The aim of this project is to text mine Twitter posts to find new software/tools, such as development tools, and to discover public perception of these releases.

I am using Twitter as it is a rich source of data for opinion mining and sentiment analysis with over 200 million tweets every day[[1]](#footnote-1).

Objectives:

1. Collect tweets.
2. Filter through tweets initially by keyword to find software-related posts. E.g. names of software, operating systems, games.
3. Create linguistic preprocessing tools.
4. Analyse tweet sentiment.
5. Structure and integrate data.
6. Visualise data through GUI.
7. Evaluate system.

Deliverables:

1. Integrate Twitter Streaming API into stub application.
2. MySQL database with tables holding Twitter users, tweets and discovered tools.
3. Named-entity recogniser for tools.
4. Create or integrate existing sentiment classifier.
5. Improve information retrieval and filtering algorithms based on data.
6. GUI

1. http://blog.twitter.com/2011/06/200-million-tweets-per-day.html [↑](#footnote-ref-1)