CS533 – INFORMATION RETRIEVAL SYSTEMS

PROJECT PROPOSAL

ANALYSIS OF LOCATION-BASED TWEETS

Group Members:

Elif DAL

Fatma BALCI

Havva Gülay GÜRBÜZ

Mert Emin KALENDER

1) MOTIVATION / IMPORTANCE

Social media via web and mobile applications has enabled people to communicate and share information with others and adopted by many. The rapid growth in social media usage among people has added new aspects and dimensions to interpersonal interactions [1]. Twitter is one of the most popular social media services. It is a microblogging application used by millions of people who are supplied with ability to remain socially connected with others and share ideas, thought etc. [2] After its launch, Twitter users have increased rapidly, and as of June 2012 the number of Twitter users was estimated 517 million generating billions of tweets per month [3].

Regarding this vast usage of social media, the streams shared by social media services provide access to huge amounts of information describing ongoing incidents in real-time. Twitter API is a stream service enabling developers to retrieve tweets with corresponding user, date and location information. This service provides ability to retrieve and analysis vast amount of information for certain places and time ranges.

2) DESCRIPTION OF THE PROBLEM

Rising of the social media has created a digital landscape where people are able to socialize and express themselves through a variety of means and application [5]. In this project, we aim to analyze location-based tweets to share information easier.

Twitter API provides access to tweets with related information. Tweets for a certain location can provide significant information about ongoing incidents for that area in real-time. Further, this can be extended by sentimental analysis of tweets and used to get a fast and effective way to monitor the event in details throughout the feelings of people. For example, retrieving tweets posted around Eskişehir Caddesi during the rush hours or tweets within Şükrü Saraçoğlu Stadium area during a football match would give important information for the event going on there.

3) METHODOLOGY

Twitter supplies developers with a public API[1], currently version 1.1, to retrieve collections of tweets with time and location information. This service is by itself sufficient to get tweets for certain location and time range. Considering the sentimental analysis of tweets, there are various existing sentiment classifiers for Twitter datasets with different results [4]. Although the approach to be used for analysis is not yet decided, the approaches and datasets will be

[1] https://dev.twitter.com/docs/api/1.1

considered for sentimental analysis part. Probably one or more of these approaches will be implemented, compared, or a new approach will be defined.

4) EXPECTED RESULTS

The expected result is a compact web application that can be used to search for locations and retrieve tweets with analysis result for a certain location.

REFERENCES

[1] - A. L. Hughes and L. Palen. Twitter Adoption and Use in Mass Convergence and

Emergency Events. In proc. of the 6th ISCRAM Conference, Sweden, May 2009.

[2] - S. Milstein, A. Chowdhury, G. Hochmuth, B. Lorica, and R. Magoulas. Twitter

and the micro-messaging revolution: Communication,connections, and immediacy.140

characters at a time. O'Reilly Media, 2008.

[3] - http://techcrunch.com/2012/07/30/analyst-twitter-passed-500m-users-in-june-2012-140m-of-them-in-us-jakarta-biggest-tweeting-city/

[4] - Saif, Hassan, Yulan He, and Harith Alani. "Semantic sentiment analysis of

twitter." The Semantic Web-ISWC 2012. Springer Berlin Heidelberg, 2012. 508-524.

[5] - G.Paltoglou, M.Thelwall. Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media, ACM Transactions on Intelligent Systems and Technology (TIST), September 2012