DD2476 Search Engines and Information Retrieval Systems

Project 6: Virality Prediction

Contact: Sahar Asadi, Meltwater (sahar.asadi@meltwater.com), Roelof Pieters, Meltwater (roelof.pieters@meltwater.com)

This project is worth 3 ECTS credits. This means that it is expected to require 80 hours of work for each person in the group. The project formulation, method, and results are presented in a report as well as in a poster session. For more details, look at the course homepage, under Project in the menu.

Problem

Twitter has become a huge source of information in today's world. Analysis of the behavior of social network and how the information spreads quickly can be an important factor in opinion mining and analysis of trends. In this projects we are interested to investigate on the spread of information on Twitter. We want to predict which tweets will be re-tweeted the most. In other words, we are interested in predicting virality of tweets.

Assignment

The task is to develop a framework to show virality prediction in tweets. Your framework at the end should visualize the result for two use-cases how the virality is investigated. This task involves

- collecting data from twitter APIs (you will get some help and guide from us).
- studying key elements in analyzing tweets virality (we will help you with introducing some related articles).
- extract usernames and hashtags from each tweet.
- analyze life cycle and distribution of a data in the tweet dataset that you collected. For instance, how many retweets happens, what is the distribution of a mention over time in the social network that you are investigating.

About Meltwater

Meltwater is an international media monitoring company established in Oslo, Norway in 2001. The company is growing in the media monitoring domain and has offices in more than 27 countries and the HQ is located in San Francisco. We provide insight from outside data in the web to help the customer to make an informed decision. The

engineering teams are located at Stockholm, Gothenburg, Berlin, Budapest, Bangalore, San Francisco, and New Hampshire.

The defined project is related to extracting knowledge to build better insights. The project is meant to be fun and a team work. The project outcome can possibly be used as part of our knowledge base projects in our data science team.

Please feel free to reach out if you have any questions.