DD2476 Search Engines and Information Retrieval Systems

Project 5: Social Media Powered Recommendations

Contact: Roelof Pieters, Artificial Experience (roelof@kth.se, 076-3173848)

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 major source of information in today's world. Analysis of the behavior of social networks, which kinds of information is shared and by whom can be used as proxy for finding information or experts.

Natural language processing, graph theory and social network analysis can be used to give insight into knowledge communities, influencers, and experts: which users are known to be expert in a particular domain?

Assignment

As your assignment you will be given a twitter dump of some one million tweets.

Your assignment:

- Create a simple application (or web application) that lets a user search for a keyword/topic and get an ordered list of recommended twitter users to follow back as a result, as well as potential top matching tweets for that topic

This assignment can be expanded by trying out various ways of making sense of twitter data:

- through modeling it in a graph database[1],
- using natural language processing or topic modeling[2] for deriving topics from tweets and/or users,
- and dealing with out-of-vocabulary keywords by matching these to synonyms, for instance with word2vec[3].

[1] As inspiration see http://nicolewhite.github.io/neo4j-jupyter/twitter.html or https://github.com/ptwobrussell/Mining-the-Social-Web-2nd-Edition/tree/master/ipynb
[2] http://radimrehurek.com/gensim/wiki.html#latent-dirichlet-allocation
[3] http://rare-technologies.com/word2vec-tutorial/

Furthermore, a wide selection of implementations/code of machine learning research papers can be found at GitXiv[4].

About Artificial Experience

Artificial Experience is a distributed research team and open network that is currently forming. It brings together machine learning, HCI, design and creative industry people, with the shared goal of researching "CreativeAI"[5] and its its applications, implications and opportunities for academia, industry and culture.

If you have any questions, don't hesitate to ask (in English).

[4] http://gitxiv.com/

[5] An extensive thought piece on the discipline of CreativeAI is at https://medium.com/@ArtificialExperience/creativeai-9d4b2346faf3