

DD2476 Search Engines and Information Retrieval Systems

Project 3: Twitter Page Rank

Contact: Simon Stenström, Findwise (simon.stenstrom@findwise.se, 073-616 35 34)

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

Google (and probably many of the other big online search engines) uses among other things Page Rank. On forums, page rank is not very useful. There are not enough linked content to decide if a document is relevant or not.

The idea behind Page Rank is that linked content is good content. On twitter, references from other tweets are probably a measure on the quality.

Assignment

Your assignment is to implement the Page Rank algorithm but use references as links to decide if a tweet is relevant or not.

1. Crawl a part of twitter
2. Detect references to other tweeters.
3. Run the Page Rank algorithm.
4. Consider if the hashtags (ie #politics) can be used to decide if a reference is more relevant than another (Is a reference to a tweeter that tweets about the same keyword more relevant?).

If you have time, use this information and index a part of twitter into Solr (<http://lucene.apache.org/solr/>) with boost your value.

About Findwise

Findwise is a growing IT consultancy company, founded in 2005 by a team of experts from the enterprise search industry. The company currently employs about 70 people (January 2012) and have offices in Sweden, Denmark, Norway and Poland.

The project is meant to be fun but could possibly be used as a demo of what our customers could do with their data more than just making it searchable. The

implementation idea should NOT be considered the “correct” solution, but may be used as a pointer to one way of solving the problem.

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