TweetRank

TweetRank is an attempt to apply the PageRank algorithm on Twitter statuses (tweets).

It uses a different rank calculation which considers attributes such as number of replies/retweets and hashtags.

|  |  |
| --- | --- |
| Victor Hallberg | [victorha@kth.se](mailto:victorha@kth.se) |
| Johan Stjernberg | [stjer@kth.se](mailto:stjer@kth.se) |
| Joan Puigcerver I Perez | [joanpip@kth.se](mailto:joanpip@kth.se) |
| Alexander Hjalmarsson | [alehja@kth.se](mailto:hjalle@sgh.se) |
| Christoffer Rydberg | [chrryd@kth.se](mailto:chrryd@kth.se) |

# Components



**Ranker:** computes TweetRank from crawler data

**Solr:** indexes and provides query interface for end user

**Crawler:** collects data via the Twitter API

# Crawler

# Ranker

* Uses the complete path Monte Carlo algorithm
* Does random walks of length N started from each node
* Ranker uses 10 threads to compute walks in parallel
* Final rank is the total number of visits for each node after all walks are done

|  |  |  |  |
| --- | --- | --- | --- |
| Thread 1 | Thread 2 | Thread 3 | Thread 4 |
| https://github.com/mogelbrod/TweetRank/blob/master/doc/graphs/mc-threads-0.png?raw=true | https://github.com/mogelbrod/TweetRank/blob/master/doc/graphs/mc-threads-1.png?raw=true | https://github.com/mogelbrod/TweetRank/blob/master/doc/graphs/mc-threads-2.png?raw=true | https://github.com/mogelbrod/TweetRank/blob/master/doc/graphs/mc-threads-3.png?raw=true |
| {2, 1, 0, 1} | {1, 1, 1, 1} | {2, 1, 1, 0} | {1, 1, 1, 1} |

{6, 4, 3, 3}

**Normalize rank**

{6, 4, 3, 3}

# Solr / Lucene

* Used as the query interface, index and weighting (including tf-idf scoring).
* Avoiding precision problem with TweetRank from 0-10 instead of the summing them to 1.
* Product of lucenes tf-idf and TweetRank + boost of hashtags and usernames.
* Receives tweets through POST requests with XML messages.
* Using Solr's external file field to get the TweetRank scores.

# Results