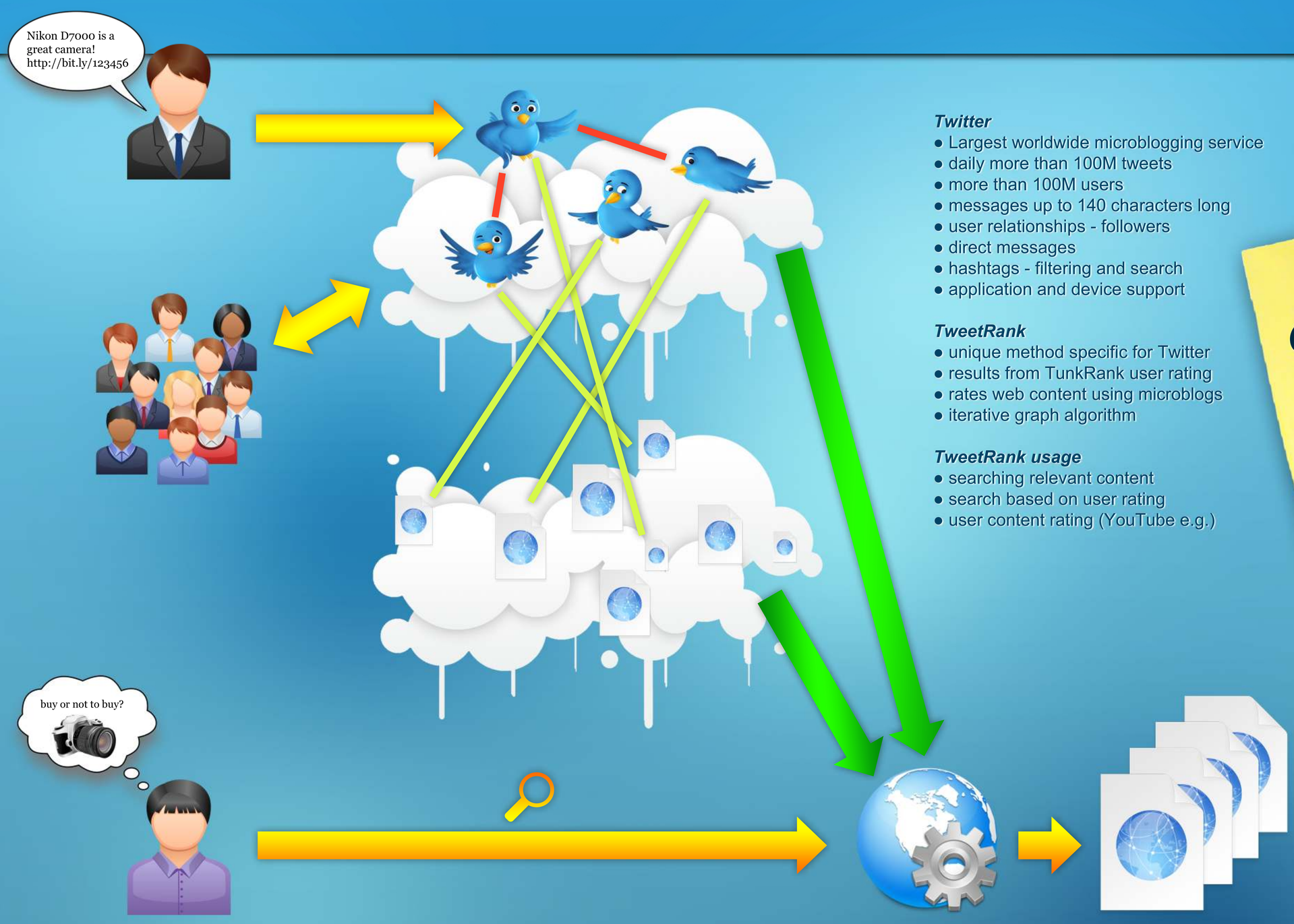
Leveraging Microblogs for Resource Ranking

Web 2.0 • huge amounts of user generated data • search improvement

Tomáš Majer

Supervisor: Marián Šimko

PeWe@FIIT personalized web group



Twitter
- Largest worldwide microblogging service
- daily more than 100M tweets
- more than 100M users
- messages up to 140 characters long
- user relationships - followers
- direct messages
- hashtags - filtering and search
- application and device support

TweetRank
- unique method specific for Twitter
- results from TunkRank user rating
- rates web content using microblogs
- iterative graph algorithm

TweetRank usage
- searching relevant content
- search based on user rating
- user content rating (YouTube e.g.)

Conclusions

New & unique method

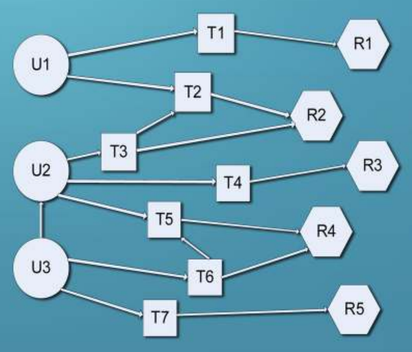
Considering Twitter structure – Retweets

User ranks based on user followers and tweets

Utilize in sorting search results

Evaluation based on user ranks on YouTube videos

Prefers current and relevant resources



$$TweetRank(url) = \sum_{t = TWU(url)} \left(TR(t) + \sum_{r = RETW(t)} TR(t) * TR(r) \right)$$

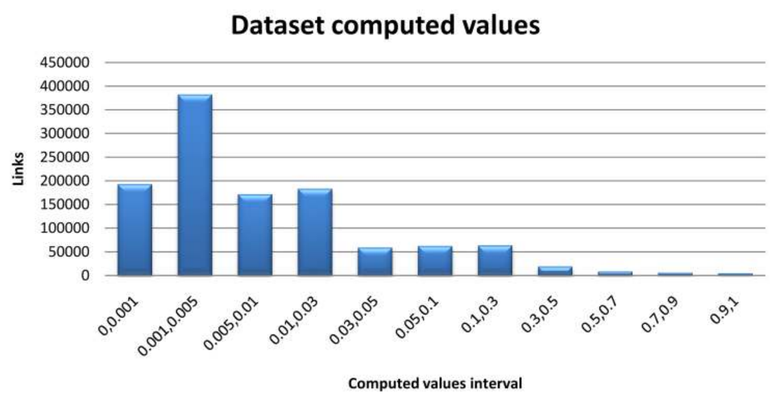
$$TR(t) = \frac{UR(A(t))}{|T(A(t))|}$$

$$UR(u) = \sum_{f = F(u)} \frac{1 + K(u) * UR(f)}{|F(f)|}$$

Dataset

Tweets and profiles:
- 1 997 446 tweets
- 367 824 user profiles

Followers:
- 1 468 365 182 connections
- 40 103 281 nodes

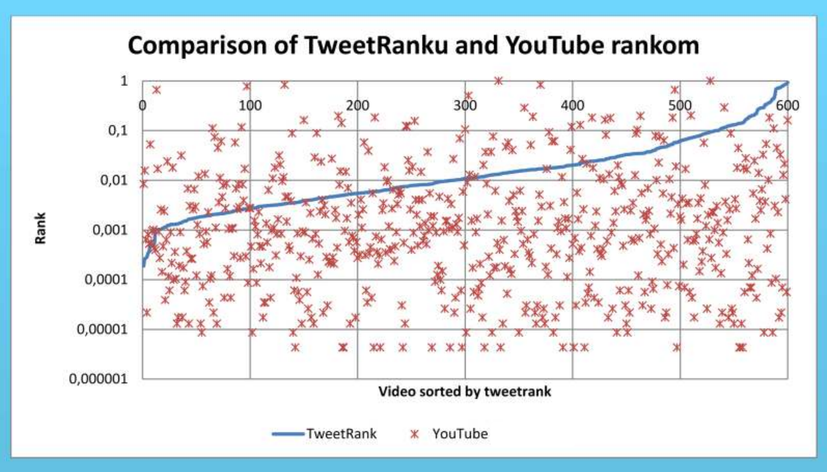


Experiments

a

Comparison of TweetRank results with user rating on YouTube. Used 600 samples.

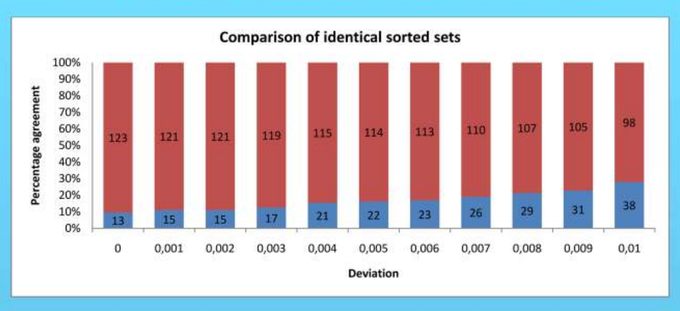
- no corelation was found among results
- limited tweet time in dataset
- change of rating algorithm on YouTube

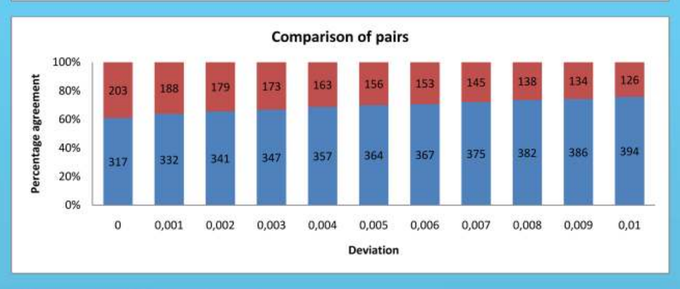


b

Comparison of calculated rating of YouTube videos with user rating using own application.

- 600 videos.
- users had to rate a set of 5 videos using marks (1 to 5)

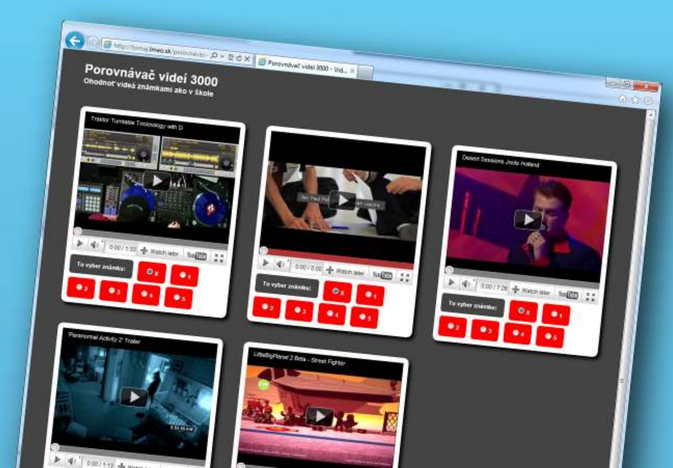




70 users, 680 ratings

corelation: 0,387

c

Utilizing TweetRank in search result sorting

- 20k web documents indexed from dataset
- comparison based on keyword count and computed TweetRank value
- TweetRank does not return relevant results on large sets of documents
- when sorting smaller sets, TweetRank prefers relevant data
- possibility to use for search result sorting