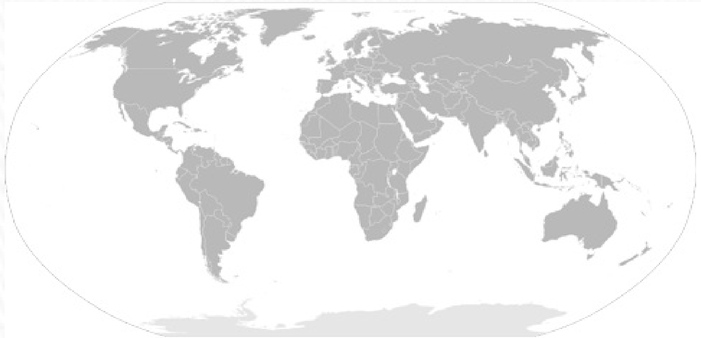
Marcel Kanta
Supervised by
Marián Šimko

Trend-Aware User Modelling with Location-Aware Trends

Motivation

Information overload



Solution

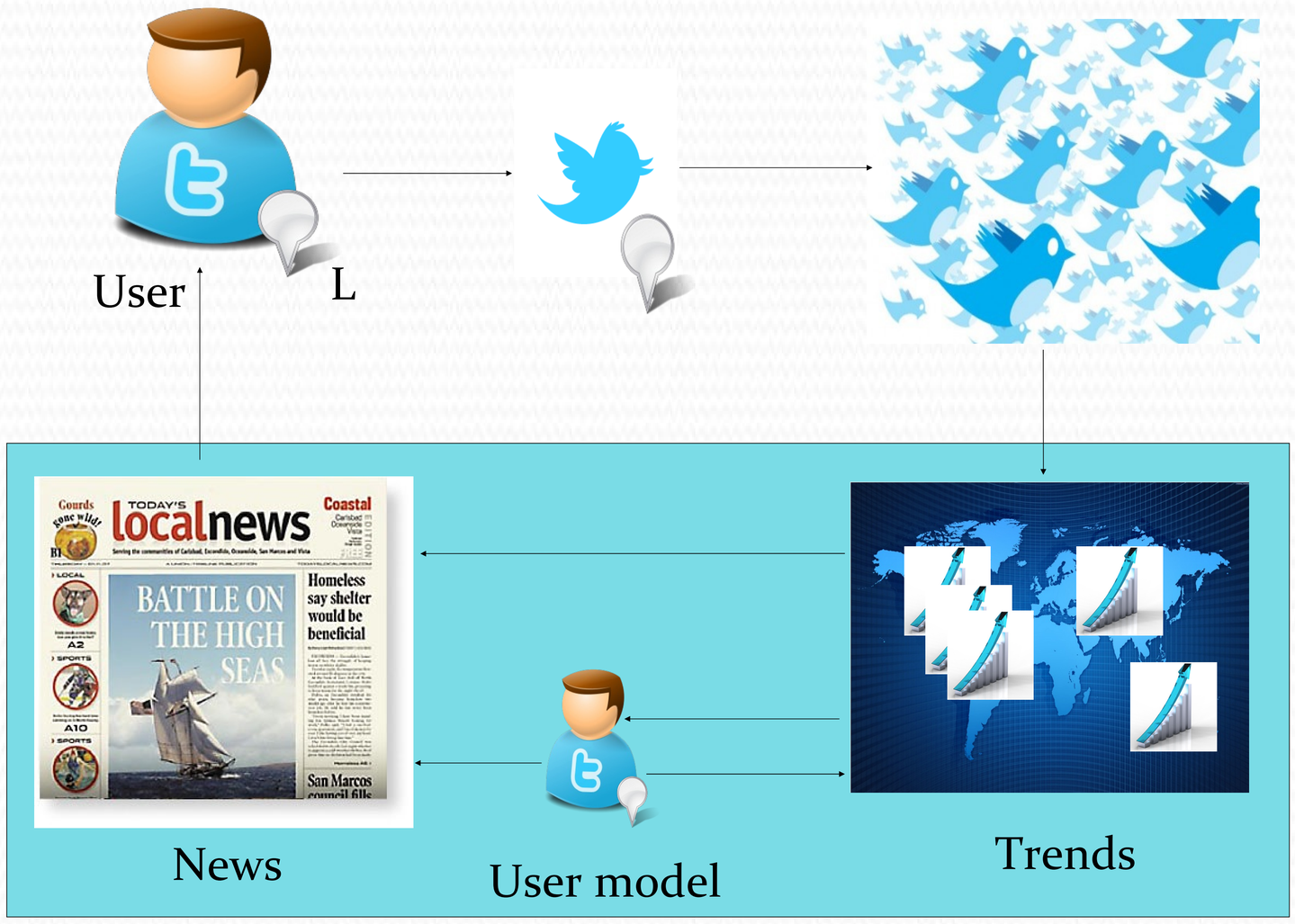
- Content recommendation
- User model
- Location awareness of trends

Model

- TF-IDF, tTF-IDF
- Filter
- User model
  - User, Region, Word, Weight
- Trend model
  - Period, *Region*, Word, Weight

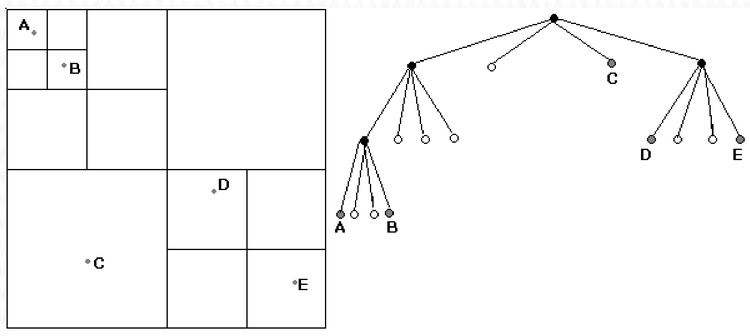
Our method

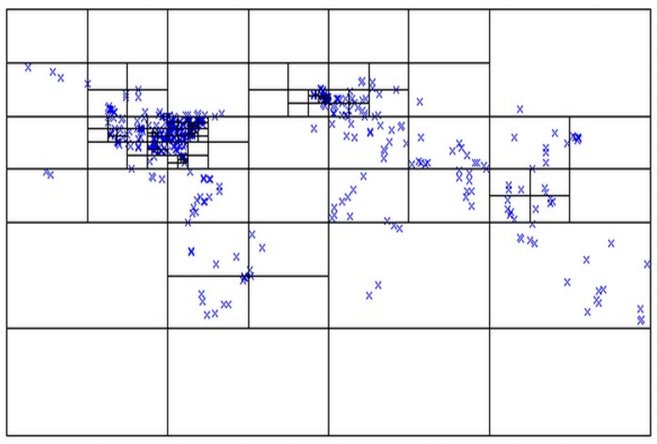
Input: enriched entities, topics from tweets, *Output*: links



Region

- PR-Quadtree
- Locations from profile





Implementation

- Map-Reduce
- Hive
- User defined functions

```
45_2010›0 → 3078990193496064 → 21196127 → 1 → Entertainment_Culture
45_2010›0 → 3078990193496064 → 21196127 → 0.535 → Technology_Internet
45_2010›0 → 3079109739548672 → 117525637 → 1 → Education
45_2010›78 → 3079235040186368 → 41901869 → 1 → HumanInterest
45_2010›17 → 3079235040186368 → 41901869 → 1 → HumanInterest
45_2010›15 → 3079235040186368 → 41901869 → 1 → HumanInterest
45_2010›11 → 3079235040186368 → 41901869 → 1 → HumanInterest
45_2010›6 → 3079235040186368 → 41901869 → 1 → HumanInterest
45_2010›3 → 3079235040186368 → 41901869 → 1 → HumanInterest
45_2010›0 → 3079235040186368 → 41901869 → 1 → HumanInterest
```

```
Add jar udfs.jar;
Create temporary function GROUP_NUMBER_ROWS
as 'sk.fiit.kanta.hive.UDFNumberRows';

create table enriched(cas string, region int,tweetid
string, userid int,
relevance double, word string) stored as sequencefile;
insert overwrite table enriched
select a.cas, a.region, a.tweetid, a.userid,
sum(a.relevance), a.word from
(select * from train
union all
select t.cas, t.region,
t.tweetid,t.userid,u.relevance,u.word from
(select * from train where word like 'http://%') t
join urlWordUnique u on t.word=u.url
) a group by a.cas,a.region,a.tweetid,a.userid,a.word;
```

Dataset

- UMAP – Abel et al. 2011
  - 2M Tweets, *Topics*, *Entities*
  - 2k Users
- 550k Links: topics & entities

Validation

Train & Test
Train: User model creation
Test: Recommendation
- Cosine similarity
- Comparison w. Gao et al. 2011

Results (small set)*

Precision P@n	1.5%/1.6%
Recall	9%/10%
F-measure	0.027/0.029
Mean Reciprocal Rank	0.194/0.198
Mean Average Rank	38/41

*top 100. local vs. global model (no local trends)

CALAIS Powered by Thomson Reuters
SEMANTIC PROXY
batchgeo Mapping should be this easy
PeWe@FIIT personalized web group
fiit.src