MPI-INF at the NTCIR-11 Temporalia Task

Robin Burghartz, Klaus Berberich
rburghar@mpi-inf.mpg.de, kberberi@mpi-inf-mpg.de
Max Planck Institute for Informatics Saarbrücken, Germany

INTRODUCTION

Knowing the target time of a query (such as weather in tokyo) might help to present more relevant results to the user. **But how can we determine the target time of a query?** In this year's NTCIR-11 Temporalia TQIC subtask, we look at the problem of putting queries in four different classes:

1. Past (History of Coca-Cola)
2. Recency (apple stock price)
3. Future (release date for ios7)
4. Atemporal (lose weight quickly)

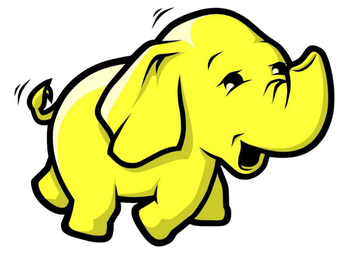
APPROACH

We **apply a selection of established off-the-shelf components** to **derive features from a broad spectrum**. We use standard classifiers on:

- Time series statistics (cf. [2])
- Linguistic properties of the query string
- Temporal dictionary entries of query terms
- …and others

As tools and libraries we used:

- WEKA 3 (for learning and classification)
- StanfordCoreNLP (for all kinds of NLP tasks)
- Hadoop/MapReduce (for the construction of the temporal dictionary)

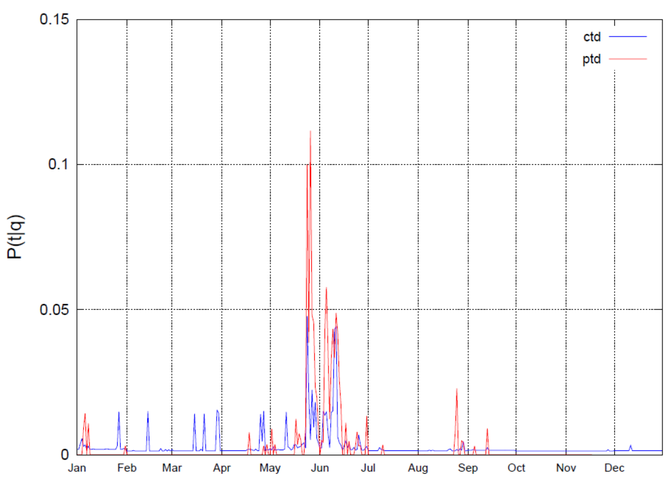


FEATURES

Time Distributions

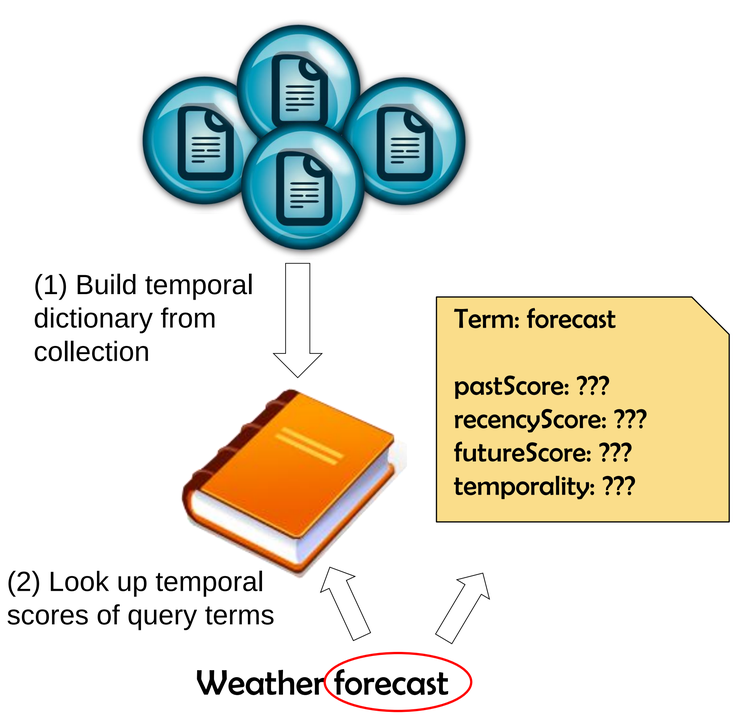
We investigate the effectiveness of temporal distribution features derived from documents relevant to a given query and distinguish between:

1. **Publication times** (timestamps of relevant documents)
2. **Absolute content times** (temporal expressions in relevant documents)
3. **Relative content times** (time orientation of temporal expressions relative to the relevant document's timestamp)



Temporal Dictionary Features

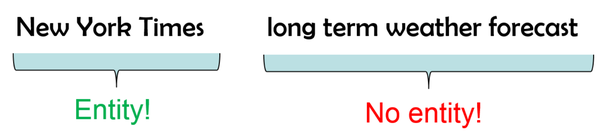
We measure the temporal orientation of words by answering the question: **how often does a given word co-occur with dates from the past, present and future?** We build a temporal dictionary of terms that contains this information and that links terms with their most likely time orientation, similar to [1].



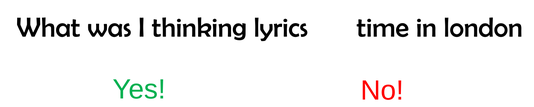
Linguistic Query Properties

We try to find correlations between NLP features of the query, such as POS tags or named entities, and the query's temporal class. Examples:

Named Entities.
Is the query string an entity?



Part of speech.
Does the query contain a personal pronoun?



Temporal expressions in the query.
Exploiting occurences of dates:

Did the Pirates win today
Disney prices 2014

RESULTS

	Run 1	Run 2	Run 3
Past	0.53	0.60	0.60
Recency	0.57	0.49	0.44
Future	0.65	0.71	0.63
Atemporal	0.73	0.76	0.80
Overall (in %)	**62.33**	**64.00**	**61.67**

Observations:

- We achieve a slight improvement over the baseline
- Queries of recent time interest are especially hard to classify
- Time distributions similar to [2] can only partially help at determining the time orientation of a query
- Linguistic features are helpful if queries are formulated as sentences
- Our temporal dictionary can completely replace word-vector features

REFERENCES

[1] Adam Jatowt, Ching-Man Au Yeung and Katsumi Tanaka. *Estimating document focus time*. CIKM 2013.

[2] Rosie Jones and Fernando Diaz. *Temporal profiles of queries*. ACM Trans. Inf. Syst., 25(3), July 2007.

[3] R. Burghartz. *Temporal query classification*, B.Sc. Thesis, Saarland University, 2014.

MAX-PLANCK-GESELLSCHAFT

max planck institut informatik