

# Temporal Query Classification at Different Granularities

Dhruv Gupta and Klaus Berberich

Max Planck Institute for Informatics

June 25, 2015

## Data

There are six query sets with their respective sizes are given in Table 2.

File Name	Description	Set Size
amb_y.q.keywords.csv	Periodic & ambiguous at year	113
aperi.q.keywords.csv	Aperiodic	118
amb_m.q.keywords.csv	Ambiguous at month	64
amb_d.q.keywords.csv	Ambiguous at day	74
unamb.q.keywords.csv	Unambiguous	142
atemp.q.keywords.csv	Atemporal	154

Table 1: Query set sizes for our evaluation setup

Each Comma Separated Value (CSV) files contains query id ( $q_{id}$ ) and the corresponding query keywords ( $q_{keywords}$ ).

## Description

**Queries.** The challenging aspect of evaluating our approach was compiling a list of queries for temporally ambiguous class at different granularities. To this end we use various previously published resources [3], TREC time-sensitive queries [1], NTCIR Geo-Time queries [2], and also manually compiled some of them from the Web. Next, we describe the query sets that we have used for each temporal class. Table 2 summarizes the query workload.

**Temporally Unambiguous Queries**, were constructed with keywords describing historical events. Sources used were [1, 2, 3]. Example queries are : `american civil rights activist rosa parks died, and concorde crash`.

**Temporally Ambiguous Queries**, were constructed for three different subsets at year, month and day granularity. For *year granularity* we accumulated international award events in various domains such as Sports, Science, Arts. Lists of

such events is available in Wikipedia <sup>1,2,3</sup> pages. Since, these events are periodic in nature, it also serves us the purpose of periodic queries. For example: us presidential elections and nobel prize in literature.

*Temporally Ambiguous Queries at Month Granularity*, considered major sports events in the United States with year of the season. E.g. nba playoffs 1990.

*Temporally Ambiguous Queries at Day Granularity*, were various sporting event competitions at a specific Summer Olympic games. Since each event competition spans multiple days but occur within the same month of the year. An example query would be, summer olympics 1992 archery.

*Aperiodic Temporally Ambiguous Queries* were constructed via keywords that were either broad category of accidents or natural calamity (for example mercy killings, or earthquakes united states of america). Reliable sources for these queries were from related work [1, 2, 3, 4]. Since events concerning known figures tend generally not be periodic in nature (e.g. abraham lincoln), we also considered several famous personalities as keyword queries in this category. A reliable source was obtained from Biography Online <sup>4</sup> website.

**Atemporal Queries**, consist of general vocabulary words with no temporal significance. E.g. apple, sardine, guitar, etc. An English dictionary listing common food items and musical instruments was used for this purpose.

Table 2: Query set sizes for our evaluation setup

	Set Id	Description	Size	
TX	TA	TAY	Periodic & ambiguous at year	113
		TAYP	Aperiodic & ambiguous at year	118
	TAM	Ambiguous at month	64	
	TAD	Ambiguous at day	74	
	TU	Unambiguous	142	
AT		Atemporal	154	

## Citation

This data set was compiled for our publication at SPIRE 2015 :

Gupta D. and Berberich K. :  
*“Temporal Query Classification at Different Granularities”*,  
 SPIRE 2015.

Kindly, cite our work when using the dataset in your research.

<sup>1</sup>[http://www.wikipedia.org/wiki/List\\_of\\_prizes,\\_medals\\_and\\_awards](http://www.wikipedia.org/wiki/List_of_prizes,_medals_and_awards)

<sup>2</sup>[http://www.en.wikipedia.org/wiki/List\\_of\\_multi-sport\\_events](http://www.en.wikipedia.org/wiki/List_of_multi-sport_events)

<sup>3</sup>[http://www.en.wikipedia.org/wiki/List\\_of\\_literary\\_awards](http://www.en.wikipedia.org/wiki/List_of_literary_awards)

<sup>4</sup><http://www.biographyonline.net/people/famous-100.html>

## References

- [1] W. Dakka, L. Gravano, and P. G. Ipeirotis. Answering general time-sensitive queries. *IEEE Trans. Knowl. Data Eng.*, 24(2):220–235, 2012.
- [2] F. Gey, R. Larson, N. Kando, J. Machado, and T. Sakai. Ntcir-geotime overview: Evaluating geographic and temporal search. In *Proc. NTCIR-8 Workshop Meeting*, pages 147–153, 2010.
- [3] D. Gupta and K. Berberich. Identifying time intervals of interest to queries. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM 2014)*, 2014.
- [4] R. Jones and F. Diaz. Temporal profiles of queries. *ACM Trans. Inf. Syst.*, 25(3), 2007.