

# Event Search and Analytics

## Detecting Events in Semantically Annotated Corpora for Search & Analytics

Dhruv Gupta

Max Planck Institute for Informatics, Saarbrücken, Germany  
Saarbrücken Graduate School of Computer Science, Saarbrücken, Germany  
dhgupta@mpi-inf.mpg.de

### CCS Concepts

•Information systems→Information systems applications;  
Data mining; Information extraction;

### Keywords

Information Retrieval; Text Analytics; Text Summarization;  
Semantic Annotations; Diversity & Novelty; Semantic Search

### ABSTRACT

Information retrieval systems have largely relied on word statistics in text corpora to satisfy information needs of users by retrieving documents with high relevance for a given keyword query. In my PhD research I hypothesize that information needs of users can be satisfied to a greater extent by using *events* as means of navigation in text corpora. An event, in my context is an act performed by certain actor(s) at a specific location during a specific time interval. With the availability of tools that can provide us with accurate semantic annotations in form of named entities, geographic locations, and temporal expressions; we can leverage growing number of knowledge resources such as Wikipedia and ontologies such as Freebase [1] to understand natural language text and mine important events. Formally we can state the central hypothesis as follows:

**Central Hypothesis** Given text corpora with semantic annotations; traditional information retrieval models can be improved by utilizing knowledge about *events* and using *events* as proxies for information needs.

Consider the text <sup>1</sup> below. Having semantic annotations in this text we can now devise algorithms that can deduce that the event is that of Usain Bolt winning Olympic competition in Beijing, China.

... Beijing<sup>(Geo:(39.55,116.23))</sup> where Bolt<sup>(Wiki:Usain-Bolt)</sup> announced himself to the world with two Olympic golds and two world records in 2008<sup>(Time:[01-01-2008,31-12-2008])</sup> ...

<sup>1</sup><http://www.bbc.com/sport/0/athletics/34032366>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WSDM 2016 February 22-25, 2016, San Francisco, CA, USA

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-3716-8/16/02.

DOI: <http://dx.doi.org/10.1145/2835776.2855083>

**Research Objectives.** Given text corpora with semantic annotations, I describe three important research problems: *i.* identifying important events; *ii.* using events for retrieval tasks; and *iii.* using events for analytics.

**Identifying Important Events.** *Events* are the proposed building blocks for further text analysis. Given semantically annotated text corpora, the objective shall be to report important events given a multidimensional query consisting either of keywords, time, geographical location or named entities. For example, given keyword query **summer olympics**; the reported events will be the various summer Olympics with their location, time interval, participating named entities, and keywords describing the events.

**Diversifying and Summarizing Search Results** are retrieval tasks that try to address the information need underlying an ambiguous query at different levels of textual granularity. I propose to use the mined set of *events* as information intents. This will allow for automatic creation of *event timelines* or *entity biographies*.

**Semantic Search and Analytics.** The mined set of *events* can further be utilized for search and analytics. The objective would to first model the mined set of events as a *data cube* (see Figure 1) and subsequently provide *data cube operations* [2]: roll, slice, dice, drill up, and drill down.

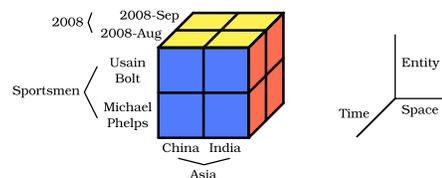


Figure 1: Example data cube for events

**Conclusion.** In this abstract I laid out an outline of the research work that I envisage to carry out for my PhD dissertation. The research in its culmination shall provide us methods to computationally extract world history as sequence of temporally ordered events and portray future events to take place from semantically annotated corpora. The research would also provide ways to perform semantic search and large scale event analytics on these annotated corpora.

### REFERENCES

- [1] Bollacker K. et al. Freebase: A collaboratively created graph database for structuring human knowledge. SIGMOD'08.
- [2] Han J. et al. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., 3<sup>rd</sup> edition, 2011.
- [3] Gupta D. *Event Search and Analytics*. WSDM'16 Doctoral Consortium.