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

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

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

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., 3rd edition, 2011.
[3] Gupta D. Event Search and Analytics. WSDM’16 Doctoral Consortium.