WikipEvent: Temporal Event Data for the Semantic Web

Ujwal Gadiraju, Kaweh Djafari Naini, Andrea Ceroni, Mihai Georgescu, Dang Duc Pham, Stefan Dietze, and Marco Fisichella

L3S Research Center, Leibniz Universität Hannover, Germany
{gadiraju, naini, ceroni, georgescu, pham, dietze, fisichella}@L3S.de

Abstract. In this demo we present *WikipEvent*, an exploratory system that captures and visualises continuously evolving complex event structures, along with the involved entities. The framework facilitates entity-centric and event-centric search, presented via a user-friendly interface and supported by temporal snippets from corresponding Wikipedia page versions. The events detected and extracted using different mechanisms are exposed as freely available Linked Data for further reuse.

Keywords: Events; Temporal Evolution; Wikipedia; RDF; Interface

1   Introduction

Exploratory search systems help users to search, navigate, and discover new facts and relationships. We detect and extract events from different sources and merge these events into a unique *event repository*. Each event is described primarily in terms of (i) a list of entities (Wikipedia pages) participating in the event, (ii) a textual description of the event, (iii) start and end dates of the event, and (iv) the extraction method used to obtain the event. We further classify entities as people, organizations, artifacts, and locations by exploiting the class hierarchy defined in YAGO2 [2], since different entity categories play different roles while participating in an event.

In this demo, we showcase two popular usecase scenarios. First, we show that *WikipEvent* can be used in order to explore the events in which particular entities have been involved. Secondly, we show the suitability of *WikipEvent* to explore the evolution of entities based on the events these are involved in. Both these scenarios can be additionally surveyed based on the temporal dimension. In addition to the events that are presented using a *timeline*[1], we adopt a versioning approach introduced previously[1] in order to display significant versions of wikipages corresponding to the entities involved in the event.

WikipEvent facilitates the understanding of events and their related entities on a temporal basis. Instead of exploration of isolated knowledge bases, WikipEvent takes advantage of the complimentary nature of sources contributing to the underlying event repository.

---

[1] Interface: `http://wikipeventdemo.l3s.uni-hannover.de/WikiEventEntity/`

2   WikipEvent Data

We extract events from three sources; Wikipedia Current Events portal[3], Yago[2], and an event detection method called *Co-References*[4]. Firstly, the WikiTimes project[2] provides an API for 50,000 events between 2001-2013, acquired from the Wikipedia Current Events portal. The second event source is the YAGO2 ontology including entities which describe events, e.g. 2011_Australian_Open as well as facts connecting entities, e.g. $< BobDylan > wasBornIn < Duluth >$. The Co-References method introduced in [4] extracts Wikipedia pages (entities) related to an event by using the Wikipedia edit history. The edits corresponding to a Wikipedia page (entity) are analysed for indications of the occurrence of an event involving that entity.

The resulting repository contains more than 2.6 million events, extracted from the different sources as shown in Table 1. The contribution from the sources is skewed due to two reasons. Firstly, they cover very different time periods: YAGO2 contains events and temporal facts spanning over thousands of years, Current Events captures events since 2001 only, and, for performance reasons, the Co-References method has been restricted to edits in 2011. In addition, Co-References exclusively analyses entities of type *politician*, while YAGO2 and Current Events contain almost all the Wikipedia pages. To facilitate comparisons across entities occurring in all three sources, in the rest of this section we will consider only those events that occurred in 2011 and involved politicians.

Source	Total	Politicians	Politicians 2011
All	2,629,740	50,168	1,401
YAGO2	2,578,547	42,399	360
Current Events	50,951	7,527	799
Co-Reference	242	242	242

Table 1: Number of events within the event repository, split by different sources.

Co-References is able to detect events with different duration and granularity (from a wrestling match to the Egypt Revolution). YAGO2 mostly contains high-level and well-known events represented through temporal facts regarding entities, often lacking textual descriptions. Current Events portal contains daily events which have a self explanatory textual description and are reliable, thanks to the high level of control within Wikipedia. The complimentary nature of the different sources in terms of complexity (number of participants), duration, and granularity of events is evident. Due to these reasons, also the schemas used to represent events across different sources are distinct, yet overlapping. While certain properties (for instance for time points) are overlapping yet follow different conventions, we lifted the events from different sources into a unified dataset following Linked Data principles and deploying a joint RDF schema.

Exposing Events Data as RDF.
We have exposed the *WikipEvent events* data through a public Linked Data

---

[2] http://data.l3s.de/dataset/wikitimes

interface using D2R Server[3], enabling URI dereferencing via content negotiation and providing a public SPARQL endpoint. This data can be accessed and queried via `http://wikipevent.l3s.uni-hannover.de/` and using our SPARQL endpoint[4].

Property	Value
lode:atPlace	Virginia
we:event_description	A Virginia judge rules the states prohibition of gay marriage unconstitutional.
we:event_end	Thu Feb 13 01:00:00 CET 2014
we:event_id	Wikitimes_169403
we:event_source	Wikitimes
we:event_start	Thu Feb 13 01:00:00 CET 2014
we:event_story	NULL
lode:involved	Virginia,Gay marriage
lode:involvedAgent	NULL
rdfs:label	events #1
we:reference	http://www.usatoday.com/story/news/nation/2014/02/13/virginia-same-sex-marriage/5473687/
rdf:type	we:Event

Fig. 1: Example of event related data as an RDF repository.

We represent events through established event RDF vocabularies, to facilitate reuse, interpretation and linking of our data through third parties. In particular, we use properties from the LODE ontology[5] to map different properties pertaining to events in our dataset, for instance, property `lode:atPlace` is used as predicate for stating venues of events. Figure 1 presents an example of an event and its entailing properties in our event repository.

3 Use Case Scenarios

WikipEvent can help students, scholars, historians, or journalists by facilitating temporal search focussed on either the entities at hand or the events. Thus, the WikipEvent framework can be used to satisfy two primary scenarios - entity-based and event-based information needs. Results are presented through a user-friendly interface, that supports faceted search, query-path tracing, query completion, temporal settings, and is abridged with events' sources as well as filters for related entities. The underlying versioning system [1], helps us to identify *significant* wikipage revisions of the entities involved in the event. These significant revisions of wikipages are also presented to the user in addition to the timeline of events. As introduced in our previous work, a *significant revision* is one where the edits between preceding and succeeding revision are above a certain threshold. Here, the notion of significance is modeled based on the Cosine distance between successive revisions [1].

Entity-based Search. Users may want to learn about entities of interest with respect to their temporal participation in events. For example, journalists might aim at studying political affiliations of individuals, and the campaigns

---

[3] `http://d2rq.org/`

[4] http://wikipevent.l3s.uni-hannover.de/snorql/
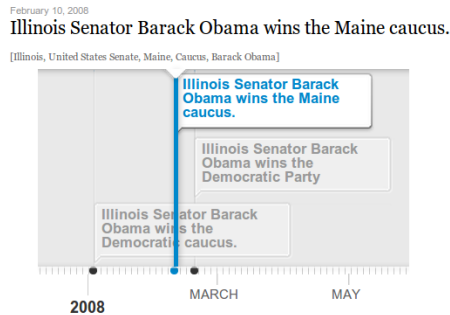
[5] `http://linkedevents.org/ontology/`

Fig. 2: Entity-centric search:*Barack Obama.*   Fig. 3: Event-centric search:*Iraq War.*

they participated in. Existing systems however, make it cumbersome to easily access this information. The WikipEvent interface overcomes this challenge by presenting a timeline of events that an entity is involved in. Additional filters for relevant entities help users to navigate through the retrieved results. Figure 2 presents an example of an entity-based search on WikipEvent.

Event-based Search. Historic events are a subject of interest to a wide array of people, ranging from students to archivists. WikipEvent facilitates a free-text search for events. Events relevant to a given query are presented in the form of a continuous timeline (Figure 3), while highlighting the entities involved. WikipEvent enables users to sift through event related information on a temporal basis, in order to learn more about the events and the participating entities.

To gain a complete understanding of the WikipEvent framework, we point the reader to a demo video [6].

References

1. Andrea Ceroni, Mihai Georgescu, Ujwal Gadiraju, Kaweh Djafari Naini, and Marco Fisichella. Information evolution in wikipedia. In *Proceedings of the 10th International Symposium on Open Collaboration*, OpenSym '14. ACM, 2014.
2. Johannes Hoffart, Fabian M Suchanek, Klaus Berberich, and Gerhard Weikum. Yago2: a spatially and temporally enhanced knowledge base from wikipedia. *Artificial Intelligence*, 194:28–61, 2013.
3. Giang Binh Tran and Mohammad Alrifai. Indexing and analyzing wikipedia's current events portal, the daily news summaries by the crowd. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 511–516, 2014.
4. Tuan A. Tran, Andrea Ceroni, Mihai Georgescu, Kaweh Djafari Naini, and Marco Fisichella. Wikipevent: leveraging wikipedia edit history for event detection. In *Web Information Systems Engineering - WISE 2014, Thessaloniki, Greece, 12-14 October 2014, Proceedings*, 2014.

---

[6] A demo video is available on the home screen of the web interface.