

# EventWiki: A Knowledge Base of Major Events

Tao Ge<sup>1,2</sup>, Lei Cui<sup>2</sup>, Baobao Chang<sup>1</sup>, Zhifang Sui<sup>1</sup>, Furu Wei<sup>2</sup>, Ming Zhou<sup>2</sup>

<sup>1</sup>MOE Key Laboratory of Computational Linguistics, Peking University, Beijing, China

<sup>2</sup>Microsoft Research Asia, Beijing, China

{tage, lecu, fuwei, mingzhou}@microsoft.com, {chbb, szf}@pku.edu.cn

## Abstract

This paper introduces a new resource called *EventWiki* which is, to the best of our knowledge, the first knowledge base resource of major events. In contrast to most existing knowledge bases that focus on static entities such as people, locations and organizations, our *EventWiki* concentrate on major events, in which all entries in *EventWiki* are important events in mankind history. We demonstrate that *EventWiki* is a very useful resource for information extraction regarding events in Natural Language Processing (NLP), knowledge inference and automatic knowledge base construction.

**Keywords:** EventWiki, event knowledge base, major event, resource, event-driven, information extraction, knowledge base construction, knowledge inference

## 1. Introduction

Knowledge bases (KBs) are large collections of facts about entities such as people, organizations and locations, which have attracted growing interest in recent years because they are useful in various studies and applications such as question-answering, semantic parsing, information extraction and knowledge inference.

Most existing knowledge bases (e.g., Freebase) focus on static entities (e.g., people, locations, and organizations) and their relations. Few of them, however, pay much attention to events and event-related knowledge; instead, they just treat events as normal entities and do not specially distinguish them, which makes it hard to identify event entries. Even though some facts in KBs can be treated as events, they tend to entity-centric and insufficient to describe major and complicated events (e.g., *2011 Tohoku earthquake* and *2014 World Cup*). In fact, event knowledge is very important for many applications and its related tasks (e.g., event extraction and event detection) have been extensively studied in both NLP and Data mining community in the past decades. Unfortunately, due to the lack of rich event-related knowledge and annotations, studies related to event knowledge discovery are either limited (e.g., in event extraction evaluation under Automatic Content Extraction project<sup>1</sup>, only events of 33 predefined event types are studied) or shallow (e.g., clustering-based event detection).

To provide rich event knowledge and a dataset for event-related studies, this paper introduces a new knowledge base resource – *EventWiki* is a knowledge base of events, which collects 21,275 major events of 95 types from Wikipedia<sup>2</sup>. We will show in the following section that *EventWiki* contains rich information about events (e.g., event taxonomy, event schemas and event relations), which makes it useful for a variety of studies and applications such as information extraction, knowledge inference and knowledge base construction. To the best of our knowledge, *EventWiki* is the first knowledge base of events which covers such large

numbers of major event entries.

## 2. EventWiki

Compared to static entities (e.g., person, location and organization), there is less work focusing on event knowledge discovery due to a lack of event-centric knowledge bases. Although Wikipedia and some existing knowledge bases (e.g., DBpedia) contains information of major events, there is no explicit indicator that indicates whether a page refers to an event or an static entity (e.g. a person), or event-event relation links, which makes it difficult to distinguish these events from overwhelming static entities and directly use these resources for event-related knowledge discovery.

For better studying event knowledge, we manually identify event pages from Wikipedia, build links between events to develop an event knowledge called *EventWiki*. As introduced before, entries in *EventWiki* are all major events that happened throughout mankind history. For an event entry, there are 4 kinds of information in *EventWiki*: event type, event infobox, event summary and full text description of the event. In addition, for some events, there are relations between them. Table 1 shows the structure of an event entry where  $e \in E$  is an event entry in *EventWiki*,  $T$  is the set of 95 event types, and  $rel_i : e_{k_i}$  means that  $e$  are related with  $e_{k_i}$  and their relation is  $rel_i$ . A concrete example is shown in Figure 1 that presents the basic structure of *EventWiki* where event information and event-event relations are clearly illustrated.

Event: $e$	
event type	$t \in T$
infobox	$\{slot_1:value_1, slot_2:value_2, \dots\}$
summary	text
full text	text
relation	$\{rel_1:e_{k_1}, rel_2:e_{k_2}, \dots\}$

Table 1: Structure of an event entry.

In this section, we introduce our resource – *EventWiki* as follows: we first introduce meta-data (i.e., event types and event schema), and then present the information of an event

<sup>1</sup><http://projects.ldc.upenn.edu/ace/>

<sup>2</sup><https://en.wikipedia.org>

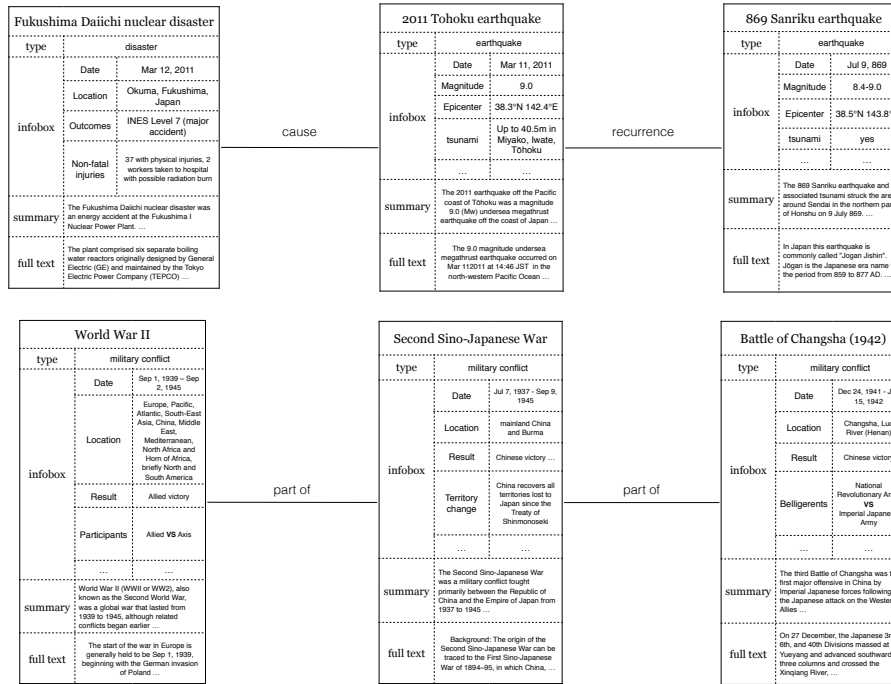


Figure 1: A (small) part of events in *EventWiki*, showing the basic schema of *EventWiki*.

entry in detail in *EventWiki* and the application of this resource.

## 2.1. Meta-data: Event types and schemas

As meta-data for describing event information, event types and schemas are very important information in many applications such as event extraction. However, there is no large amounts of annotated data for event typing or event schema extraction. As a result, most studies regarding event extraction are either limited (only focused on a small number of pre-defined event types) or shallow (just consider a verb as an event).

In *EventWiki*, there are 95 event types in total, derived from infobox’s template types in Wikipedia by manually screening from English Wikipedia dump on December 3, 2015. Every event entry in *EventWiki* belongs to one of these 95 event types.

For each event type, we extract the schema of this event type by finding out the most frequent non-empty slots from the infobox of events belonging to this type. The most frequent slots of an event type can be used to describe key information of an event of this type and thus can be considered as the schema of the event type. Table 2 shows the schemas of 6 event types.

With the information, it is easy to learn a classifier for event classification and event schema extraction. For event classification, each event entry in *EventWiki* with its event type can be treated as training data. By using such information, it is possible to train an event classification model which can be used to classify events reported in news articles every day for further event knowledge discovery. Likewise, *EventWiki* provides valuable resources for event schema extraction. For events that belong to one of 95 types in *EventWiki*, the event schema of the event type in *EventWiki* can be directly used as the schema for event extraction. Even

if an event does not belong to any of those 95 types, we still can use rich information in *EventWiki* to infer its event schema. For example, text information and slots in the infobox of an event entry in *EventWiki* can be used as training data to learn what slots an event should have given its text description. For example, a model may learn from *EventWiki* that if “magnitude” frequently appear in text description of an event, “magnitude” should be one slot of this event in its schema.

## 2.2. Events in *EventWiki*

As mentioned before, each event entry has 4 kinds of information (i.e., event type, event infobox, event summary and full text description) and event-event relation information. We call those 4 kinds of information intra-event information because they describe the information of an event while we call event-event relation inter-event information since it presents relations between events.

### 2.2.1. Intra-event information

In this section, we introduce the intra-event information in detail.

- **Event type:** event type presents the type of an event which must be one of 95 event types. In *EventWiki*, an event’s type is extracted by analyzing the template type of the first infobox of its Wikipedia article.
- **Event infobox:** as most static entities in Wikipedia, events also have infoboxes which describe most important information of events. We extract events’ infoboxes from Wikipedia. Note that slots (i.e., attributes) in an event’s infobox depends on the event’s type.
- **Event summary:** a summary of an event is a short text that summarizes the event, which is obtained by ex-

aircraft accident	olympic event	wildfire	military conflict	earthquake
date	date	date	place	date
site	teams	location	date	location
occurrence_type	gold	fatalities	combatants(1,2)	magnitude
crew	silver	area	result	depth
fatalities	bronze	source	partof	casualties
operator	venue	injuries	commanders(1,2)	countries affected
passengers	competitors	building	strength(1,2)	origintime
origin	nations	landuse	casaulties(1,2)	tsunami
destination	win_value	cost	territory	intensity

Table 2: Examples of schemas of 6 event types

tracting the first paragraph of an event’s text description at Wikipedia.

- Full text description: In contrast to event summary, full text description contains all text information in an event page at Wikipedia.

### 2.2.2. Inter-event information

Inter-event information is relations between events. For example, in Figure 1, we can see the relation between *Fukushima Daiichi nuclear disaster* and *2011Tohoku earthquake*.

For the current version *EventWiki*, inter-event information is mainly obtained by analyzing events’ infoboxes in which some event-event relations are presented. For example, by analyzing the infobox of *Second Sino-Japanese War*, we can extract the relation *partof(Second Sino-Japanese War; World War II)*. In total, we extracted 9,713 event-event relations from infoboxes.

In addition to infoboxes of events, many event-event relations are described by text. For example, the relation between *2011 Tohoku earthquake* and *869 Sanriku earthquake* is expressed by the following text in *2011 Tohoku earthquake*:

This megathrust earthquake was a recurrence of the mechanism of the earlier *869 Sanriku earthquake*, which has been estimated as having a magnitude of at least 8.4 Mw, which also created a large tsunami that inundated the Sendai plain.

However, it is difficult to automatically obtain such event-event relations in high accuracy from text. Fortunately, it is possible to exploit structure of a page at Wikipedia to help harvest event-event relations, as shown in Figure 2. In this way, we extracted additional 540 event-event relations.

## 2.3. Applications

We show that event information in *EventWiki* can be very useful for the following tasks and applications:

### 2.3.1. Event extraction

Event extraction is an important information extraction task. In this paper, we model event extraction as infobox generation problem for an event. Intuitively, an event infobox contains key information of an event. If an event infobox can be accurately generated, we can say that the information of an event is perfectly extracted. Usually, event

infobox generation problem can be divided into three sub-tasks: event classification, event schema extraction and slot filling. As mentioned before, *EventWiki* can provide rich information for event classification and event schema extraction. Moreover, it is extremely useful for training a slot filling model for event extraction.

Slot and value pairs in the infobox (intra-event information) in *EventWiki* can be used as weak (distant) supervision for training a slot filling model, as (Reschke et al., 2014) did. For example, for the slot value pair “*magnitude: 9.0*” in *2011 Tohoku earthquake*, we first find out the sentences which “9.0” appears in. The context information of “9.0” can be used as features and “*magnitude*” is used as the label of “9.0” for training a slot filling model.

### 2.3.2. Event-event relation extraction and inference

As slot filling for event extraction, we can also use inter-event information in *EventWiki* to train an event-event relation extraction model using distant supervision strategy. For an event relation triple, we can find out the sentences that mention both events in the triple. For example, for the triple *partof(Second Sino-Japanese War; World War II)*, the following sentence is found for generating training instance for relation extraction:

As the Western Allies entered the war against Japan, *the Sino-Japanese war* would become part of a greater conflict, the Pacific theatre of *World War II*. (label: *partof*)

To the best of our knowledge, even though relation extraction has been extensively studied in NLP, most of them focus on relations between static entities (e.g., people and locations) and there are very few resources for event-event relation extraction. *EventWiki* can plug the gap and make it possible to study event-event relation extraction problem. With event-event relations, it is also possible to apply knowledge inference approaches to *EventWiki* to infer event-event relations, which is a novel task and has not been well studied. For example, with *partof(Second Sino-Japanese War; World War II)* and *partof(Battle of Changsha, Second Sino-Japanese War)*, we can infer *partof(Battle of Changsha, World War II)*, which would be for useful for building event taxonomy.

### 2.3.3. Event summary generation

*EventWiki* is a good resource for event summary generation. As introduced before, each event entry in *EventWiki*

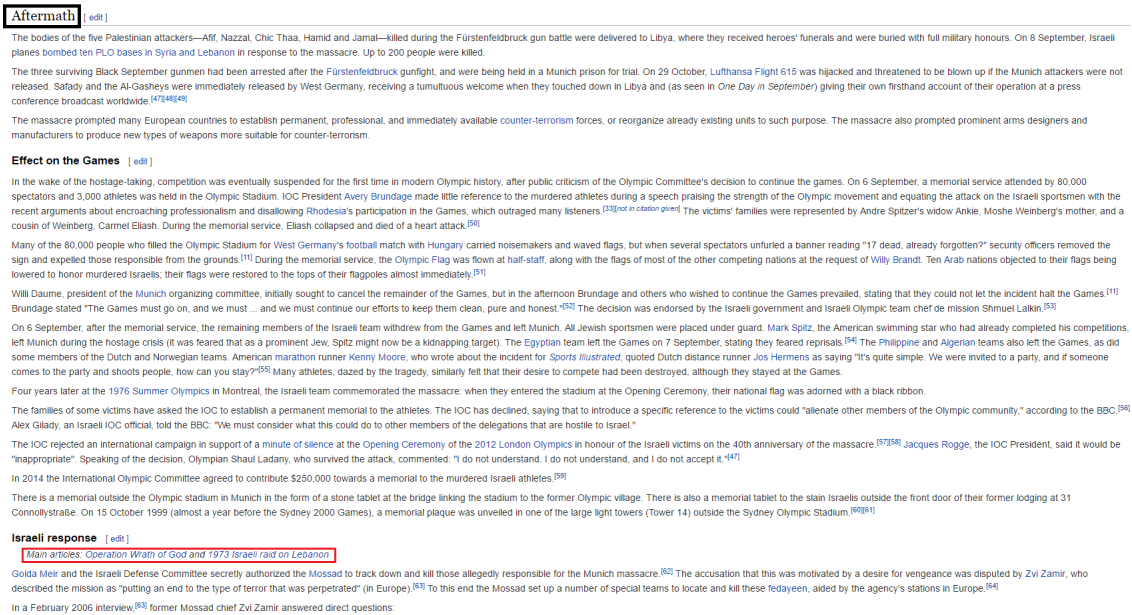


Figure 2: Part of “Munich massacre” article at Wikipedia. We use the structural information to harvest event-event relations such as *aftermath(Munich massacre, 1973 Israeli raid on Lebanon)*.

has a summary and full text description, which can be seen as naturally annotated data for training a summarization model that generates a summary from full text description. In addition, it can even be used as training data for generating a summary of an event from an event’s infobox or generating a text description of a (slot,value) pair of an event, which is very useful for question answering systems based on knowledge bases.

### 2.3.4. Event-centric knowledge base construction

As we introduced, *EventWiki* is a knowledge base of events which is collected from pages at Wikipedia that are manually constructed and edited. It is an example for event knowledge base and can be helpful for automatic event knowledge base construction.

Given a collection of documents about a major event, we can use *EventWiki* to predict the event’s type, extract the event’s schema, generate its infobox, identify the relations with other major events and generate a summary for it. When we finish all the steps above, the profile of the event is built, which can be used as an event entry for an event knowledge base. In other words, the above five steps can be seen as the subtasks of event knowledge base construction.

Among 21,275 events in *EventWiki*, we select 1,517 events happening during 1995-2010 and find<sup>3</sup> corresponding news articles of these events in English Gigaword (Parker et al., 2011), which can be used to test results of each step and evaluate end-to-end performance.

## 3. Related Resources

Knowledge bases are useful resources for various applications in many fields. Among the existing knowledge bases

<sup>3</sup>We use information retrieval package Lucene (Jakarta, 2004) to find relevant documents to these events and manually check the results.

such as WordNet<sup>4</sup> and Gene Ontology<sup>5</sup>, knowledge bases of general world knowledge such as Freebase<sup>6</sup>, DBpedia<sup>7</sup> and YAGO (Mahdisoltani et al., 2014) have received most attention. Most of the knowledge bases mainly focus on static entities and their related facts but pay little attention to influential events.

To the best of our knowledge, the most related resource to ours is the data released by (Reschke et al., 2014), which contains 193 air crash events and their corresponding news articles. In some sense, this data can be seen as a small knowledge base of air crashes. In contrast, our *EventWiki* contains much more major events of a variety of types and rich knowledge about events, which is the first complete event-centric knowledge base as far as we know.

## 4. Conclusion and Future Work

In this paper, we introduce a novel knowledge base – *EventWiki* which is the first knowledge base of major events. We introduce its structure and information contained in it, and show its potential value for NLP and automatic KB construction.

In future, we plan to enrich the event-event relation information by adopting a semi-automatic approach for analyzing relations between events in text. Additionally, we will regularly update the knowledge base to ensure that it contains the latest events and plan to provide an online interface and APIs for easier access.

## 5. Acknowledgments

The research work is supported by the National Science Foundation of China under Grant No. 61772040. The contact author is Zhifang Sui.

<sup>4</sup><https://wordnet.princeton.edu/>

<sup>5</sup><http://geneontology.org/>

<sup>6</sup><https://www.freebase.com/>

<sup>7</sup><http://wiki.dbpedia.org/>

## 6. Bibliographical References

- Jakarta, A. (2004). Apache lucene-a high-performance, full-featured text search engine library.
- Mahdisoltani, F., Biega, J., and Suchanek, F. (2014). Yago3: A knowledge base from multilingual wikipedias. In *7th Biennial Conference on Innovative Data Systems Research*. CIDR Conference.
- Reschke, K., Jankowiak, M., Surdeanu, M., Manning, C. D., and Jurafsky, D. (2014). Event extraction using distant supervision. In *LREC*, pages 4527–4531.

## 7. Language Resource References

- Parker, Robert and Graff, David and Kong, Junbo and Chen, Ke and Maeda, Kazuaki. (2011). *English Gigaword Fifth Edition*. Linguistic Data Consortium, 5.0, ISLRN 911-942-430-413-0.