OEKG: The Open Event Knowledge Graph

Simon Gottschalk[1], Endri Kacupaj[2], Sara Abdollahi[1], Diego Alves[3], Gabriel Amaral[4], Elisavet Koutsiana[4], Tin Kuculo[1], Daniela Major[5], Caio Mello[5], Gullal S. Cheema[6], Abdul Sittar[7], Swati[7], Golsa Tahmasebzadeh[6], and Gaurish Thakkar[3]

[1] L3S Research Center, Leibniz Universität Hannover, Germany
{gottschalk,abdollahi,kuculo}@L3S.de
[2] University of Bonn, Germany
kacupaj@cs.uni-bonn.de
[3] University of Zagreb, Croatia
dfvalio@ffzg.hr, gthakkar@m.ffzg.hr
[4] King's College London, United Kingdom
{gabriel.maia_rocha_amaral,elisavet.koutsiana}@kcl.ac.uk
[5] School of Advanced Study, University of London, United Kingdom
{Daniela.Major,caio.mello}@sas.ac.uk
[6] TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany
{gullal.cheema,golsa.tahmasebzadeh}@tib.eu
[7] Jožef Stefan Institute and Jožef Stefan International Postgraduate School, Slovenia
{abdul.sittar,swati}@ijs.si

Abstract. Accessing and understanding contemporary and historical events of global impact such as the US elections and the Olympic Games is a major prerequisite for cross-lingual event analytics that investigate event causes, perception and consequences across country borders. In this paper, we present the Open Event Knowledge Graph (*OEKG*), a multilingual, event-centric, temporal knowledge graph composed of seven different data sets from multiple application domains, including question answering, entity recommendation and named entity recognition. These data sets are all integrated through an easy-to-use and robust pipeline and by linking to the event-centric knowledge graph EventKG. We describe their common schema and demonstrate the use of the *OEKG* at the example of three use cases: type-specific image retrieval, hybrid question answering over knowledge graphs and news articles, as well as language-specific event recommendation. The *OEKG* and its query endpoint are publicly available.

1 Introduction

Contemporary and historical events such as the US presidential elections, the Olympic Games and major earthquakes change the world. Their media coverage, their varying perception by different communities, their historical evolution and

potentially global impact make cross-lingual event analytics a significant research topic in various fields of studies, including social science, computer science and digital humanities [4, 19].

When performing cross-lingual event analytics, the requirements towards event knowledge representation are manifold, given the heterogeneity, dynamicity and multilingualism of events [9]. Until now, there exists a large variety of event-related data sets [1, 4, 11, 20] that may help understand specific characteristics of events, but they are barely connected by now. This calls for new models and processes that enable intuitive access to the event-related knowledge spread across the world.

In this paper, we present the *OEKG*, the Open Event Knowledge Graph, which makes a step towards a holistic representation of event knowledge by the integration of event-related data sets from multiple and diverse application domains such as Question Answering, entity recommendation and Named Entity Recognition. Also, these data sets originate from different data collections, including knowledge graphs and news articles. One of these knowledge graphs is *EventKG_{light}*, a new version of the event-centric and multilingual knowledge graph *EventKG* [11]. The *OEKG* is built on top of *EventKG_{light}*, allowing for easier integration of additional data sets using RDF named graphs. We propose an efficient and robust pipeline facilitating this integration of several data sets in an easy-to-use manner.

Fig. 1 shows four example resources of *OEKG* and thus demonstrates its versatility resulting from the integration of several data sets:

- Events (Fig. 1a): Events are at the core of the *OEKG*. For example, the fire of the Notre-Dame in Paris is covered with its locations, labels in multiple languages, related events such as "The Notre Dame Cathedral holds its first mass since the April 15 fire", and more event characteristics.
- Places (Fig. 1b): Most events happen at specific event locations which are also part of the *OEKG*. Such places do not only hold labels and coordinates, but also images and further characteristics.
- News articles (Fig. 1c): Events are often reported in the media [5, 13]. Therefore, the *OEKG* provides access to annotated news articles. For example, the news article entitled "Boris Johnson takes charge of Olympic Park's future" is related to the Olympic Games.
- Questions and answers (Fig. 1d): Question Answering over knowledge graphs is an important natural language understanding task. The *OEKG* provides questions about events such as the Apollo 11 spaceflight, plus their answers (here, Neil Armstrong, Michael Collins and Buzz Aldrin).

Furthermore, the *OEKG* covers several other event-related aspects, including but not limited to (temporal) event relations, language-specific relevance scores and specialised class hierarchies. Put together, this makes the *OEKG* a versatile resource targeting a variety of potential information needs.

NOTRE-DAME DE PARIS FIRE
event__1053509
Download triples

Property	Object
rdfs:label	Пожар в „Света Богородица" (Париж) (bg)
rdfs:label	Пожар в Света Богородица (Париж) (bg)
rdfs:label	Branden i Notre Dame (da)
rdfs:label	Notre Dames brand, 2019 (da)
rdfs:label	Brand der Kathedrale Notre-Dame de Paris (de)

(a) Example event in the *OEKG*.

HANOVER
entity__4038877
Download triples

Property	Object
rdfs:label	Hanover (en)
rdfs:label	Hannover (de)
rdfs:label	Hannover (it)
rdfs:label	Hanover (pt)
rdfs:label	Ганновер (ru)
rdfs:label	Hannover (nl)
rdfs:label	Hanovre (fr)

(b) Example place in the *OEKG*.

Property	Object
rdf:type	so:Article
so:headline	Boris Johnson takes charge of Olympic Park's future (en)
so:mainEntity	oekg-r:event__981207 (2012 Summer Olympics)
so:datePublished	2012-09-12
so:inLanguage	en
so:publisher	theguardian
so:url	https://www.theguardian.com/politics/2012/sep/12/boris-johnson-olympic-park-legacy
so:mentions	oekg-r:event__640206 (Paralympic Games)
so:mentions	oekg-r:event__1635523 (Transport)
so:mentions	oekg-r:event__405245 (Olympic Games)
so:mentions	oekg-r:entity__1296070 (London)

(c) Example news article in the *OEKG*.

question1175

Property	Object
rdf:type	so:Question
so:mainEntity	oekg-r:event__991496 (Apŏllo 11)
so:suggestedAnswer	oekg-r:answer1175
so:text	Which astronaut from NASA was on the mission of Apollo 11? (en)

answer1175

Property	Object
rdf:type	so:Answer
so:mainEntity	oekg-r:event__991496 (Apollo 11)
so:text	The astronauts whose mission is Apollo 11 and are from NASA are Neil Armstrong, Buzz Aldrin, Michael Collins (astronaut). (en)
so:mentions	oekg-r:entity__12624230 (Neil Armstrong)
so:mentions	oekg-r:entity__11913301 (Michael Collins (astronaut))
so:mentions	oekg-r:entity__12989150 (Buzz Aldrin)

(d) Example question in the *OEKG*.

Fig. 1. Example resources in the *OEKG* (not all triples are shown).[8]

The *OEKG* contains more than 400 million triples from seven data sets and is publicly available: We provide the triple dumps for download, a SPARQL endpoint and access to all nodes on the *OEKG* website[9]. We also provide permanent access to the *OEKG* on Zenodo[10].

---

[8] The photos of the example place are taken from Wikimedia Commons, with the second photo being licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. The maps of the example event and the example place are licensed under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF).

[9] http://oekg.l3s.uni-hannover.de

[10] https://zenodo.org/record/4503163

The remainder of this paper is organised as follows: First, we present our integration pipeline (Section 2). Then, we describe the data sets integrated into the $OEKG$ (Section 3) and the $OEKG$ schema (Section 4). In Section 5, we provide two example use cases of the $OEKG$. Finally, we conclude in Section 6.

2   Creation of the $OEKG$

The creation of the $OEKG$ requires an integration pipeline where a set of data sets is transformed into a single, integrated knowledge graph that provides links between all the involved resources. $EventKG_{light}$ – a multilingual, event-centric knowledge graph later described in Section 3 – serves as the base data set of the $OEKG$ that contains nodes representing real-world entities and events.

Our integration pipeline is driven by the goal to make the inclusion of a new data set into the $OEKG$ as simple as possible, which allows a robust and efficient process. Only then, it is possible to integrate a large variety of data sets in an efficient and faultless way. To do so, we follow a strategy defined by Galkin et al. [8] where the data from different sources is stored under respective named graphs. Starting from $EventKG_{light}$, new data sets are added consecutively, each accompanied by a unique named graph. Fig. 2 exemplifies this integration process when adding the first new data set to $EventKG_{light}$, under the named graph new_graph.
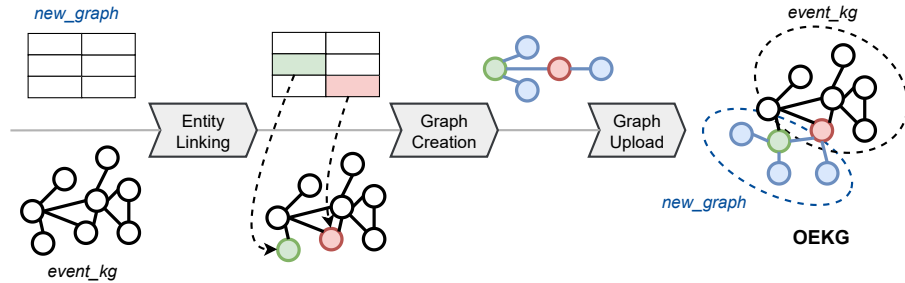

Fig. 2. Example of the $OEKG$ integration pipeline where a new, tabular, data set is added to the $OEKG$ under the named graph new_graph.

In detail, the integration process follows the following three steps:

1. Entity Linking: We require that each graph added to $OEKG$ is connected to $EventKG_{light}$. That means any resource representing a real-world entity or event is represented by an $OEKG$ resource URI. To facilitate this linking, we provide a web API that allows easy access to the $OEKG$ resource URIs given Wikidata or DBpedia URIs. In our example in Fig. 2, some input table cells are successfully linked to $EventKG_{light}$.

Algorithm 1 Example: Extension of the *OEKG* with a data set *news* that has an article about Barack Obama

---

1: **procedure** EXTENDOEKG(e)
2:     graphName $\leftarrow$ "news" .
                                                                    ▷ Entity Linking
3:     entityId $\leftarrow$ getId("en","Barack_Obama")
                                                                    ▷ Graph Creation
4:     $G \leftarrow$ new Graph(graphName)
5:     articleId $\leftarrow$ "article1"
6:     G.add(oekg-r:articleId, rdf:type, so:Article)
7:     G.add(oekg-r:articleId, so:mentions, oekg-r:entityId)
8:     fileName $\leftarrow$ storeGraphIntoFile(G)
                                                                    ▷ Graph Upload
9:     uploadGraph(fileName, graphName)

---

2. Graph Creation: After retrieval of the *OEKG* resource URIs, a set of triples is created for each data set and serialised as an N-Triples[11] file, using the RDFLib Python library[12]. In our example, a graph consisting of five nodes is created, two of them being already part of the *OEKG*.
3. Graph Upload: We provide another API method that allows uploading an N-Triples file together with the identifier of a named graph. The respective triples are then added to the *OEKG*. In our example, the resulting graph consists of two subgraphs that can be queried in isolation or together.

2.1   Example

Consider Algorithm 1 for an example of our integration pipeline. In this example, the new data set to be added to the *OEKG* under the named graph *news* contains one news article about Barack Obama. First, the *OEKG* URI of Barack Obama is retrieved via the provided API method using the English Wikipedia label (line 3). Second, a graph is created consisting of two triples and serialised into an RDF file (lines 4 - 8)[13]. Third, this file is uploaded via the provided API method (line 9). In this example, one new node is added to the *OEKG* (`oekg-r:articleId`) connected to an existing node (`oekg-r:entityId`).

2.2   Schema Extension

If possible, the data sets were transformed into triples using the *EventKG*$_{light}$ schema of the base graph. Otherwise, the use of standard vocabularies such as schema.org[14] was encouraged. In every other case, schema extensions were uploaded into the *OEKG* through separate schema files using the same procedure. We will present the resulting *OEKG* schema in Section 4.

---

[11] https://www.w3.org/TR/n-triples/
[12] https://rdflib.dev/
[13] Relevant prefixes used by the *OEKG* are later defined in Table 2.
[14] https://schema.org/

3   Data Sets

The *OEKG* integrates seven data sets which are described in this section. Table 1 provides an overview of these data sets, including the number of triples in the *OEKG* within their respective named graph. While some of these data sets are implicitly related to events, others add to the event knowledge from a different perspective, which will also prove useful as we will later show at the example of three use cases.

Table 1. Statistics of the different data sets contained in the *OEKG*.

Data Set	Short Description	Triples
$EventKG_{light}$ [11]	A light-weight version of EventKG, a multilingual, event-centric, knowledge graph.	$434,752,387$
EventKG+Click [1]	A data set of language-specific event-centric user interaction traces	$118,662$
$VQuAnDa$ [12]	A verbalization question answering dataset	$38,243$
MLM [3]	A benchmark dataset for multitask learning with multiple languages and modalities	$942,753$
InfoSpread [20]	A data set for information spreading over the news	$277,992$
$TIME$ [4]	Two collections of news articles related to the Olympic legacy and Euroscepticism	$70,754$
$UNER$ [2]	The universal named-entity recognition framework	$206,622$
$OEKG$	The Open Event Knowledge Graph	$436,407,413$

- **$EventKG_{light}$** [11]: The *EventKG* is a multilingual resource incorporating event-centric information extracted from several large-scale knowledge graphs such as Wikidata, DBpedia and YAGO, as well as less structured sources such as the Wikipedia Current Events Portal and Wikipedia event lists in 15 languages. It contains nodes representing real-world entities and events and (temporal) relations between them. For the *OEKG*, we have created *EventKG_light*, a light-weight version of EventKG that omits provenance information denoting the origin of relations, favouring an easier integration with other data sets.
  In the *OEKG*, *EventKG_light* serves as the base graph that other data sets are connected to. That is to establish an agreement concerning the identification of event-related real-world objects such as persons, places and events themselves.
- **$EventKG+Click$** [1]: *EventKG+Click* is a cross-lingual dataset that reflects the language-specific relevance of events and their relations and aims to

provide a reference source to train and evaluate novel models for event-centric cross-lingual user interaction. It directly builds upon *EventKG* and language-specific information on user interactions with events, entities, and their relations derived from the Wikipedia clickstream.

In the *OEKG*, *EventKG+Click* can be used for recommending events to users based on actual user interaction traces. Examples of particular relevant events from a language-specific view include the 2016 Berlin truck attack from the German perspective and the 2009 Russian Premier League from the Russian perspective [1].

- **VQuAnDa** [12]: The Verbalization QUestion ANswering DAtaset is a dataset for Question Answering (QA) over knowledge graphs that includes the verbalization of each answer. Through this verbalisation, *VQuAnDa* intends to completely hide any semantic technologies and provides a fluent experience between the users and the knowledge graph. *VQuAnDa* consists of $5,000$ questions accompanied by SPARQL queries and DBpedia entity links.

    QA over Knowledge Graphs is a common task in natural language processing [6]. Via the integration of question/answer pairs into the *OEKG*, both the question/answers pairs and the background knowledge are encapsulated into the same resource, enabling seamless training and application of QA systems.

- **MLM** [3]: The Multiple Languages and Modalities data set is a resource for training and evaluating multitask systems in multiple modalities, for example, cross-modal (text/image) retrieval and location estimation. *MLM* comprises text in three languages, images and location data, extracted from the Wikidata entries of $236,000$ human settlements.

    MLM is added to the *OEKG* for adding images as an additional modality to the knowledge graph. As locations are typical event characteristics, photos of locations are an immediate benefit to the representation of events.

- **InfoSpread** [20]: The data set for Information Spreading over the News provides news articles covering three contrasting events (Global Warming, FIFA world cups and earthquakes). Initially, the goal of this data set was to understand information spreading patterns over news articles. *InfoSpread* contains $7,773$ news articles related to these events in five languages.

    News articles are often used as a means to identify events [13] and oftentimes it is the media itself that makes events known to the public [5]. Therefore, the inclusion of news articles into the *OEKG* is an important step towards coverage of event-centric data from different viewpoints.

- **TIME** [4]: The temporal discourse analysis applied to media articles data set is a collection of Brazilian, British and Spanish news articles covering the concept of Olympic legacy and the concept of Euroscepticism.

    With the collection of news articles to specified events, the *OEKG* serves as an example for in-depth analysis of single events through knowledge graphs.

- **UNER** [2]: The Universal Named Entity Recognition framework proposes a 4-level class hierarchy for training and testing Named Entity Recognition tools. For example, *UNER* contains the class `Earthquake`, which is a leaf node of the following branch of superclasses: `Natural`, `NaturalPhenomenon`, `Event` and `Name`.

In the *OEKG*, *UNER* adds to the already given class hierarchy from the DBpedia ontology. Given how challenging it is to recognise named events in texts [16], we envision that the inclusion of *UNER* classes into the *OEKG* can help training and evaluating NER systems in the specific context of event-centric data.

Following the integration pipeline described in Section 2, the described data sets were added to the *OEKG*. For additional information or increased interlinkage with *EventKG$_{light}$*, some data sets were extended before:

- Via the Wikifier[15] and spaCy[16], entities and events mentioned in news articles (*TIME* and *InfoSpread*) were identified. This is to establish a connection between the news articles and *EventKG$_{light}$*: Given this connection, one may query for news articles about specific events or entities.
- Sentiment analysis, i.e., the computational study of people's opinions, sentiments, emotions, moods, and attitudes [14], contributes towards the understanding of natural-language texts and can, in particular, facilitate an analysis of news articles across languages [15]. In the *OEKG*, we enrich news articles by employing the sentiment detection system SentiStrength [21] on their headlines. That way, the *OEKG* enables queries for particularly positive or negative news articles, potentially initiating further event-centric analyses of the news articles in the context of specific events.
- To further increase the linkage between different sources, the *UNER* classes were aligned to the DBpedia ontology using the `skos` vocabulary[17] when possible.

4   Schema

Fig. 3 shows the *OEKG* schema. As described in Section 2.2, this schema is based on the *EventKG$_{light}$* schema and then extended by demand. Prefixes used in the *OEKG* schema and in the remainder of this paper are listed in Table 2[18].

In detail, the different data sets contribute to the following parts of the *OEKG* schema:

- **EventKG$_{light}$**: The *EventKG* schema is based on the Simple Event Model (**sem**)[19] and its three main classes `sem:Event`, `sem:Actor` and `sem:Place`, that are connected via `sem:hasPlace` and (temporal) relations modeled by `oekg-s:Relation` (omitted from Fig. 3 for brevity). *EventKG* further distinguishes between different types of events (`oekg-s:TextEvent`, `oekg-s:EventSeries` and `oekg-s:EventSeriesEdition`). In

---

[15] http://wikifier.org/

[16] https://spacy.io/

[17] https://www.w3.org/TR/swbp-skos-core-spec/

[18] For a full list of prefixes used in the *OEKG*, see oekg.l3s.uni-hannover.de/sparql.

[19] https://semanticweb.cs.vu.nl/2009/11/sem/

Table 2. Selected prefixes used by the *OEKG*.

Prefix	URI
oekg-r:	http://oekg.l3s.uni-hannover.de/resource/
oekg-s:	http://oekg.l3s.uni-hannover.de/schema/
oekg-g:	http://oekg.l3s.uni-hannover.de/graph/
uner:	http://oekg.l3s.uni-hannover.de/uner/
so:	http://schema.org/
rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs:	http://www.w3.org/2000/01/rdf-schema#
xs:	http://www.w3.org/2001/XMLSchema#
sem:	http://semanticweb.cs.vu.nl/2009/11/sem/
onyx:	http://www.gsi.dit.upm.es/ontologies/onyx/ns#
skos:	http://www.w3.org/2004/02/skos/core#

comparison to the *EventKG* schema, *EventKG$_{light}$* omits link count relations and adds the `skos:prefLabel` to entities for a more efficient access to their labels.

– ***EventKG+Click***: To model language-specific, weighted relations for the representation of event-centric cross-lingual user interaction, we have introduced two new classes: `oekg-s:LanguageSpecificRelation` that assigns one or more instances of `oekg-s:LanguageSpecificRelation-Score` to a source entity and a target entity. Such instances hold the score between the source and target entity in a specific language.

– ***VQuAnDa***: A question, its suggested answer and their verbalisation are represented using **schema**.**org**'s classes `so:Question` and `so:Answer`. Entities that appear in the question text are linked to *EventKG$_{light}$* instances via `so:mainEntity`, entities in the answer via `so:mentions`.

– ***MLM***: Images are assigned to places via `so:image`, descriptions via `so:description`.

– ***InfoSpread*** **and** ***TIME***: News articles are represented via `so:Article` and the respective properties denoting the headline (`so:headline`), for instance. News articles are connected to *EventKG$_{light}$* instances via `so:mentions`, which denote the appearance of an *OEKG* entity or event in the text. For the representation of news articles' sentiment, we follow the schema of the TweetsKB [7], using the **onyx** vocabulary and its classes `onyx:EmotionSet`, `onyx:Emotion` and `onyx:EmotionCategory` to assign a set of emotions of different strengths to a news article.

– ***UNER***: Entities are assigned *UNER* classes using `rdf:type`. Furthermore, the *UNER* class hierarchy and its connection to the DBpedia ontology are established using the **owl** and the **skos** vocabulary.

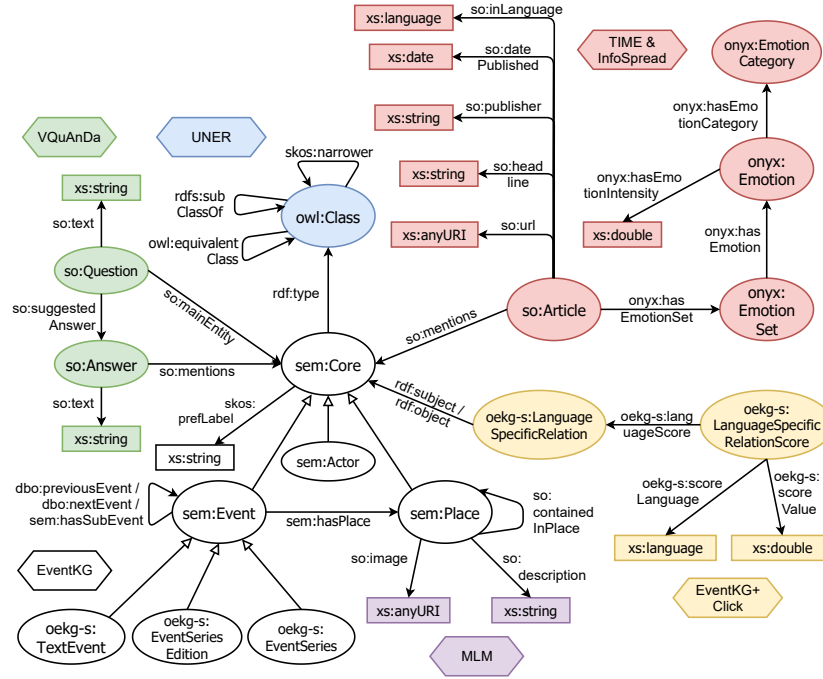**Fig. 3.** Excerpt of the *OEKG* schema. ⇢ marks `owl:subClassOf` relations. Regular arrows mark the `rdfs:domain` and `rdfs:range` restrictions on properties. Classes are coloured w.r.t. the data set for which they have been added. For brevity, we have omitted classes regarding relations between entities and events, as well as temporal attributes from the EventKG schema.

5   Example Use Cases

In this section, we demonstrate the *OEKG* and its ability to enable integrated access over multiple datasets via three example use cases.

5.1   Image Retrieval: *EventKG$_{light}$*, *MLM* & *UNER*

Event classification in images is an important task for various applications in the fields of computer vision, including geolocation estimation and place classification [17]. Such tasks typically rely on the existence of a well-defined class hierarchy and the availability of images. The *OEKG* facilitates queries both for the *UNER* type hierarchy specifically designed for Named Entity Recognition, and for images of locations, using the *MLM* data. In combination, event locations in *EventKG$_{light}$*, *MLM*'s image links, and the *UNER* type hierarchy enable retrieval of images relevant for specific event types.

We demonstrate the *OEKG*'s potential for image retrieval by an example query for images from earthquake regions shown in Listing 1.1: It queries for entities typed as earthquakes using the `uner:Earthquake` class, their locations (*EventKG$_{light}$*) and the images assigned to such locations (*MLM*). Table 3 presents selected results of this query, including a photo of the port of Messina and more.

```
SELECT DISTINCT ?Location ?Image WHERE {
  ?earthquake rdf:type uner:Earthquake ;
    sem:hasPlace ?Location  .
  ?Location so:image ?Image .
}
```

Listing 1.1. SPARQL query: Images of locations where earthquakes happened.

Table 3. Selected *OEKG* results of the SPARQL query in Listing 1.1.[20]

Location	Ferrara	Messina	Guaranda
Image			

5.2 Question Answering over News Articles: *EventKG$_{light}$*, *VQuAnDa, InfoSpread & TIME*

Question Answering (QA) is the task of supplying precise answers to questions, posed by users in natural language, and is typically divided into QA over free text and QA over knowledge graphs [6]. Through the integration of *EventKG$_{light}$*, *VQuAnDa, TIME* and *InfoSpread* into the *OEKG*, the *OEKG* facilitates a combination of these two tasks, i.e., hybrid approaches: We can query for news articles which specifically mention the entities part of the question/answer pair.

---

[20] These photos are taken from Wikimedia Commons. They are licensed under the following licenses. Ferrara: Creative Commons Attribution 2.5 Italy license. Messina: Creative Commons Attribution-Share Alike 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic. Guaranda: Creative Commons Attribution 2.0 Generic license.

This way, two sources for answering the question can be provided: the *OEKG* itself, as well as the news article potentially holding the answer to the initially posed question.

For example, the query in Listing 1.2 asks for a question in *VQuAnDa* (`?question`) that is about an event (`?questionEntity rdf:type sem:Event`). The query then searches for news articles (`?article`) mentioning both that event and one of the suggested answer entities. It returns the question "Whose wife is a presenter at WWE? (en)" and its verbalised answer "The people whose partners are presenters at WWE are John Cena, Dwayne Johnson." together with the Spanish news articles entitled "¿Qué luchador tiene el mayor porcentaje de victorias en la historia de WWE?" (*Which wrestler has the highest percentage of victories in in the history of WWE?*). The question entity "WCE (en)" is mentioned in the news article, as well as both answers: John Cena and Dwayne Johnson.

```
SELECT DISTINCT ?questionText ?answerText ?headline
                ?questionEntity ?answerEntity WHERE {

  ?question so:suggestedAnswer ?answer;
    so:mainEntity ?questionEntity ;
    so:text ?questionText .
  ?questionEntity rdf:type sem:Event .

  ?answer so:mentions ?answerEntity ;
    so:text ?answerText .

  ?article rdf:type so:Article ;
    so:mentions ?questionEntity, ?answerEntity ;
    so:headline ?headline .
}
```

Listing 1.2. SPARQL query: News articles that mention entities of a question/answer pair.

5.3   Event Recommendation: *EventKG_light* & *EventKG+Click*

As defined by Ni et. al, entity recommendation is the problem of suggesting a contextually-relevant list of entities in a particular context [18]. This task is particularly relevant in Web search. With the *OEKG*, we can specifically create language-specific recommendations for events and further enrich them with relevant event characteristics.

The query in Listing 1.3 asks for events relevant to the First World War, from the Russian point of view. We filter for the most relevant related events (`FILTER(?value >= 0.8)`) and retrieve *EventKG_light*'s event characteristics

```
SELECT ?Label ?StartDate WHERE {
  ?event owl:sameAs dbr:World_War_I.
  ?r oekg-s:source ?event ;
    oekg-s:target ?target ;
    oekg-s:hasLanguageSpecificRelationScore [
      oekg-s:scoreValue ?value ;
      oekg-s:scoreLanguage 'ru'^^xsd:language
    ] .
  ?target skos:prefLabel ?Label ;
    sem:hasBeginTimeStamp ?StartDate .
  FILTER(?value >= 0.8) .
}
ORDER BY ?StartDate
```

Listing 1.3. SPARQL query: Events related to the First World War from a Russian point of view.

to order the resulting list of events chronologically. Table 4 lists the results of this query, that clearly show a Russian focus. This result could be used for creating a language-specific event timeline similar to the link-based EventKG+TL system [10], but now inferred from actual user interaction traces in *EventKG+Click*.

Table 4. All OEKG results for the SPARQL query in Listing 1.3.

Label	StartDate
Brusilov Offensive (en)	1916-05-22
Russian Civil War (en)	1917-11-07
Treaty of Brest-Litovsk (en)	1918-03-03

6 Conclusion

In this paper, we have introduced the *OEKG* – the Open Event Knowledge Graph[21]. The *OEKG* comprises event-related knowledge from seven data sets of various application domains. We have presented an easy-to-use, efficient and robust pipeline that facilitated a seamless integration of seven data sets into the *OEKG*. At the examples of image retrieval, question answering over text and event recommendation, we have exemplified three use cases of the *OEKG*.

---

[21] http://oekg.l3s.uni-hannover.de/

References

1. Abdollahi, S., Gottschalk, S., Demidova, E.: EventKG+Click: A Dataset of Language-specific Event-centric User Interaction Traces. In: CLEOPATRA Workshop @ ISWC (2020)
2. Alves, D., Kuculo, T., Amaral, G., Thakkar, G., Tadic, M.: Uner: Universal named-entity recognitionframework. In: CLEOPATRA Workshop @ ISWC (2020)
3. Armitage, J., Kacupaj, E., Tahmasebzadeh, G., Maleshkova, M., Ewerth, R., Lehmann, J.: MLM: A Benchmark Dataset for Multitask Learning with Multiple Languages and Modalities. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. pp. 2967–2974 (2020)
4. Cheema, G.S., Major, D., Mello, C., Sittar, A.: TIME: Temporal Discourse Analysis applied to Media Articles. `http://cleopatra-project.eu/index.php/2020/06/01/time-temporal-discourse-analysis-applied-to-media-articles/` (2020)
5. Dayan, D., Katz, E.: Media Events. harvard university press (1994)
6. Dimitrakis, E., Sgontzos, K., Tzitzikas, Y.: A Survey on Question Answering Systems over Linked Data and Documents. Journal of Intelligent Information Systems pp. 1–27 (2019)
7. Fafalios, P., Iosifidis, V., Ntoutsi, E., Dietze, S.: TweetsKB: A Public and Large-scale RDF Corpus of Annotated Tweets. In: European Semantic Web Conference. pp. 177–190. Springer (2018)
8. Galkin, M., Auer, S., Kim, H., Scerri, S.: Integration Strategies for Enterprise Knowledge Graphs. In: 2016 IEEE Tenth International Conference on Semantic Computing (ICSC). pp. 242–245 (2016)
9. Gottschalk, S., Bernacchi, V., Rogers, R., Demidova, E.: Towards Better Understanding Researcher Strategies in Cross-lingual Event Analytics. In: International Conference on Theory and Practice of Digital Libraries. pp. 139–151. Springer (2018)
10. Gottschalk, S., Demidova, E.: EventKG+TL: Creating Cross-lingual Timelines from an Event-centric Knowledge Graph. In: European Semantic Web Conference. pp. 164–169. Springer (2018)
11. Gottschalk, S., Demidova, E.: EventKG–the Hub of Event Knowledge on the Web–and Biographical Timeline Generation. Semantic Web **10**(6), 1039–1070 (2019)
12. Kacupaj, E., Zafar, H., Lehmann, J., Maleshkova, M.: VQuAnDa: Verbalization QUestion ANswering DAtaset. In: European Semantic Web Conference. pp. 531–547. Springer (2020)
13. Leban, G., Fortuna, B., Brank, J., Grobelnik, M.: Event Registry: Learning about World Events from News. In: Proceedings of the 23rd International Conference on World Wide Web. pp. 107–110 (2014)
14. Liu, B.: Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press (2020)
15. Lo, S.L., Cambria, E., Chiong, R., Cornforth, D.: Multilingual Sentiment Analysis: from Formal to Informal and Scarce Resource Languages. Artificial Intelligence Review **48**(4), 499–527 (2017)
16. Marujo, L., Ling, W., Gershman, A., Carbonell, J.G., Neto, J.P., Matos, D.: Recognition of Named-Event Passages in News Articles. In: Proceedings of COLING 2012: Demonstration Papers. pp. 329–336 (2012)

17. Muller-Budack, E., Springstein, M., Hakimov, S., Mrutzek, K., Ewerth, R.: Ontology-driven Event Type Classification in Images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2928–2938 (2020)
18. Ni, C.C., Sum Liu, K., Torzec, N.: Layered Graph Embedding for Entity Recommendation using Wikipedia in the Yahoo! Knowledge Graph. In: Companion Proceedings of the Web Conference 2020. pp. 811–818 (2020)
19. Rogers, R.: Digital Methods. MIT press (2013)
20. Sittar, A., Mladenić, D., Erjavec, T.: A Dataset for Information Spreading over the News. In: Conference on Data Mining and Data Warehouses (SiKDD) (2020)
21. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment Strength Detection in Short Informal Text. Journal of the American society for information science and technology **61**(12), 2544–2558 (2010)