xLiMe
FP7-ICT-2013-10
ICT-2013.4.1
Contract no.: 611346
www.xlime.eu

Deliverable D4.3.1

Early Semantic Graph Construction Prototype

Editor:	Aljaz Kosmerlj, JSI
Author(s):	Aljaz Kosmerlj, JSI; Blaz Novak, JSI; Gregor Leban, JSI
Deliverable Nature:	Prototype (P)
Dissemination Level:	Public (PU)
Contractual Delivery Date:	M24 – 31 October 2015
Actual Delivery Date:	M24 – 31 October 2015
Suggested Readers:	xLiMe project partners
Version:	1.0
Keywords:	text mining; natural language processing; semantic graph extraction

Disclaimer

Full Project Title:	xLiMe – crossLingual crossMedia knowledge extraction
Short Project Title:	xLiMe
Number and Title of Work package:	WP4 Cross-media Semantic Integration
Document Title:	D4.3.1 - Early Semantic Graph Construction Prototype
Editor:	Aljaz Kosmerlj, JSI
Work package Leader:	Aditya Mogadala, KIT

Executive Summary

This deliverable covers the early semantic graph construction work done in the scope of the xLiMe project. We first outline the extensions made to the Event Registry ontology, which expand its vocabulary to include links to data in other modalities such as video segments and social media posts. We then present an approach for construction of semantic graphs from clusters of news articles about the same real-world event leveraging the information redundancy offered by such collections of documents. Encouraging results from a small manual evaluation of the method are presented.

Table of Contents

Abbreviations

WP	Work Package
ER	Event Registry
RDF	Resource Description Framework
URI	Unified Resource Identifier
XSD	XML Schema Definition
ISO	International Organization for Standardization

1 Introduction

The semantic web movement has been growing in importance since its inception in the early 00's. The idea of *"a web of data that can be processed directly and indirectly by machines"* as described by its inventor, Tim Berners-Lee [1], is ever more relevant in the time of increasingly powerful data mining and analysis systems. In reality the web is unfortunately still far from this ideal. Though pretty much all of the big-name web companies are currently engaged in the development of knowledge graphs (Google Knowledge Graph [2] and Facebook's Social Graph [3] perhaps being the most notable examples) most of the web as well as most of the new content produced for the web remains semantically unannotated.

Automatic extraction of structured data aims to remedy this problem. This document describes the vocabulary to link news article clusters to relevant media of different modalities and an approach for constructing semantic graphs from news article text by using machine learning methodology. We leverage the information redundancy offered by clusters of articles discussing the same real-world events to identify the importance of individual extracted semantic frames which enables us to discard unnecessary and erroneous links.

This deliverable follows and further extends the (meta-)data model defined in WP1, namely in deliverable D1.1 [7]. The goal of the work is to provide the vocabulary and means to link concepts annotated in media of different modalities (text, audio, video) into a sematic graph structure. As such its input comes from the tools developed in WP3 and other tasks in WP4 described in their respective deliverables. Produced graphs contain information about events detected across different sources and media useful as input in the analytical tools developed within WP5.

The rest of the document is organized as follows. In Section 2 we describe the extensions made to the Event Registry ontology to allow for addition of new, multi-modal data extracted by the xLiMe pipelines. Section 3 introduces a new approach for extraction of semantic graphs from events by merging semantic frames extracted from individual articles. Plans for future work are outlined in Section 4 and the concluding remarks are in Section 5.

2 Event Registry ontology extension

Event Registry [4] (ER) is a system for global news media monitoring. It follows and collects news articles from over 100.000 news sources from all over the world and in 14 languages through a sophisticated newsfeed service [5] as a data source. All articles are processed in a linguistic and semantic analysis pipeline where several types of entities and concepts are identified and annotated in the text. ER uses these annotations to cluster articles describing the same real-world happenings into events. This is why throughout this document we use the term event interchangeably for both the real-world event as the article cluster describing it. Once events are formed they are linked between each other based on content relatedness and form storylines. A comprehensive set of information is extracted about each event e.g. the time and location of the event as well as top related entities and concepts.

In order to ensure a way of describing, transmitting and searching the rich data structure constructed by ER an RDF ontology (`http://eventregistry.org/rdf/ontology/`) was developed in the scope of the PlanetData [6], a European network of excellence on large-scale data management. However the ontology was outdated as it did not follow development of ER and its data structure. It was also not equipped with vocabulary to link it to media of different modalities. In order to extend it and fix these shortcomings we follow the specification of the xLiMe meta-data model [7]. We list the namespaces used in the following definition of the extended ontology vocabulary in Table 1.

Prefix	URI
er	http://eventregistry.org/rdf/ontology/
xsd	http://www.w3.org/2001/XMLSchema#
ma-ont	http://www.w3.org/ns/ma-ont#
sioc	http://sioc-project.org/ontology/

Table 1: Namespaces used in the ontology extension

The full definition of the extended ontology is included in Annex A. Here we highlight the extensions made in the scope of this deliverable. We also list the newly Linked Data added in the extension of the ER RDF extraction service and the external vocabulary used.

- **Property: date**

URI	http://eventregistry.org/rdf/ontology/date
Label	date of event
Domain	er:Event
Range	xsd:date
Comment	Date when the event occurred. If endDate is specified, this is the date when the event started. Datatype is Date

- **Property: endDate**

URI	http://eventregistry.org/rdf/ontology/endDate
Label	ending date of event
Domain	er:Event
Range	xsd:date
Comment	Date when the event ended. Datatype is Date.

- Article language is now linked using the `http://dublincore.org/documents/dcmi-terms/language` property which links articles to their ISO639-3 codes (via `http://lexvo.org/id/iso639-3/<lang_code>` URIs).

- Related concepts are now linked to articles as well (previously they were only linked to events). Because each concept relation is assigned a score signifying how important the concept in the article is, the concepts are linked to the article via blank event factor nodes (`http://purl.org/NET/c4dm/event.owl#Factor`).

- Similarity between articles automatically computed by ER is linked using the vocabulary from the `http://purl.org/ontology/similarity/` ontology. Since similarity is computed as a numeric score this relation is also linked via a blank node.

- **Property: isDuplicateOf**

URI	http://eventregistry.org/rdf/ontology/isDuplicateOf
Label	original instance of the article
Domain	er:Article
Range	er:Article
Comment	Specifies the article of which this one is a duplicate.

- **Property: hasRelatedMedia**

URI	http://eventregistry.org/rdf/ontology/hasRelatedMedia
Label	has related media
Domain	er:Article
Range	ma-ont:MediaResource, sioc:Post
Comment	This property relates an event with an instance of media in a different modality like a social media post, an audio clip or a video segment.

The software service extracting ER RDF data has been updated to extract the listed data structure and vocabulary extensions. It then annotates the data with the proper provenance information as specified in deliverable D1.1 [7] and pushes it to the Kafka data collection service [8].

3 Event semantic graph construction

Semantic graphs are a distillation of the meaning of a text in a relational structure suitable for further machine processing. The graph structure is capable of expressing information which is too complex to be efficiently encoded in propositional form (e.g. vectors of values). Nodes in such a graph represent entities and concepts mentioned in the graph and edges encode relations between them. Relations are commonly constructed by identifying subject-verb-object patterns in the parsed text.

A more general approach is to construct *semantic frames.* A frame is a representation of a situation involving several participants (entities, concepts and modifiers). Each participant plays a *semantic role* in the frame. For example, a semantic frame constructed for the sentence: “John called Mary yesterday.” would be:

F: call
A0: John
A1: Mary
AM-TMP: yesterday

The frame represents the predicate ‘call’ with two agent entities, ‘John’ and ’Mary’, and the word ‘yesterday’ filling the temporal role (denoted by AM-TMP). Note that frames can fill roles in other frames building a nested structure. By constructing nodes from frame names and roll fillers and constructing labelled edges from roles we can build a semantic graph. An example for the frame above is presented in Figure 1.

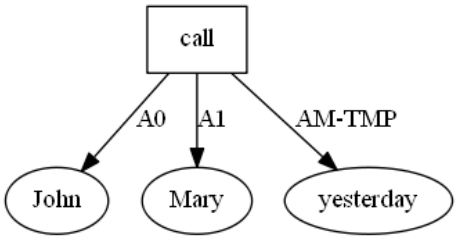

Figure 1: Example frame semantic graph

An approach for extracting semantic frames from news article text was developed in the XLike project [9]. The method uses dependency parses of article sentences to identify frames. Unfortunately, it often extracts spurious frames with little relation to the article subject matter. This is especially common in longer articles which discuss a wider context or include verbatim quotes. The aim of the work presented in this deliverable is to exploit the redundancy of information offered by clusters of articles in ER events to construct semantic graphs on the level of events by combining article semantic graphs.

3.1 Technical approach

In general, subgraph matching is computationally a very hard problem. Finding common subgraphs verbatim alone is exponentially complex with respect to the size of the graph. Since we want to take into account node generalizations this makes the problem even worse. In order to make the solution tractable, we simplify the problem by reducing it to searching for common frames only. This significantly reduces the computational complexity, while still retaining a good level of structural information as frames are subgraphs in the semantic graph by definition.

The frame extraction method we use matches frame names (i.e. verbs) to Wordnet synsets [11] but does not match other nodes (role fillers) to any ontology concepts. Since the language of different news articles differs this can result in several different frames with essentially the same meaning. To improve this and reduce the number of overall frames we match the terms in frame roles as well as frame names using the distance between their synsets in the Wordnet synset graph. If the hypernym/hyponym path is short enough the frames are merged. An example of frames suitable for merging from an article describing Apple Inc. achieving record profits is:

F: describe.01|report.01 A0:Agent: apple A1:Topic: income	F: describe.01|report.01 A0:Agent: apple A1:Topic: earnings	F: describe.01|report.01 A0:Agent: apple A1:Topic: gain.02|earn.01

Note that the topic role in the last frame is filled by another frame name (extracted most likely from a subordinate clause) and is still matched by our approach, which means we are able to effectively merge verbs with nouns when semantically appropriate.

Having merged semantically equivalent frames, we perform some additional cleaning of the resulting frame set. We remove all frames with none or just one argument as initial experiments have shown they are of little value and are typically overly general. We also remove all frames that involve pronouns. Taking the most frequent frames (or experimentally determined cutoff value was 10), we obtain the final semantic graph. An example subgraph for the event[1] describing the fatal car crash of mathematician John Nash, whose life was an inspiration for the Oscar-winning movie A Beautiful Mind, is presented in Figure 2.

[1] http://eventregistry.org/event/2905512

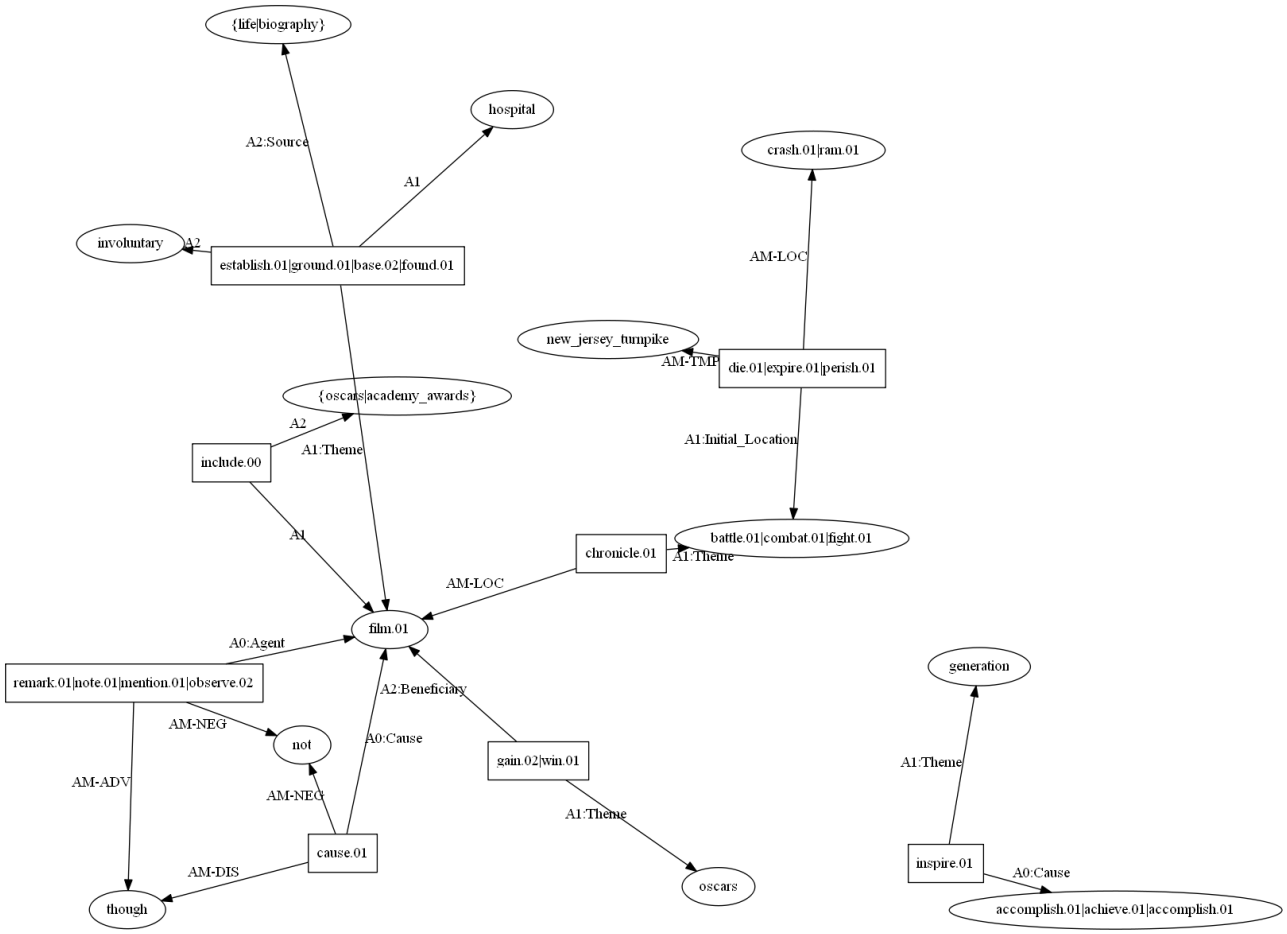

Figure 2: A subgraph of the semantic graph extracted from the event describing the fatal car accident of John Nash

3.2 Evaluation

Evaluation of extracted semantic graphs is difficult. Even more so in our case as to our knowledge there is no golden standard for semantic graphs extracted from events (i.e. article clusters). In order to obtain insight into the quality of the extracted graphs, we performed a manual evaluation of extracted frames. Since such evaluation is very labour intensive, we were only able to perform a small test.

For the test dataset we collected from Event Registry the top 15 events ranked by social score (a measure of how intensely the events were discussed in social media) from May to August 2015 that had at least 20 articles. Full listed of used events can be found in Annex B. We collected up to 200 most related articles (i.e. those closest to article centroid) from each event and used our semantic graph extraction method on them, resulting in 962 extracted frames over all events. Exact number of event articles and extracted frames is listed in Table 2.

event URI	No. of articles	No. of frames
2869218	117	85
2900400	167	61
2905512	190	50
2911484	188	66
2939388	188	143
2941988	75	12
2942974	137	117
2952242	107	80
2952931	141	9
2974565	183	47
2979180	175	165
2994206	84	37
3001562	190	48
3031380	124	13
3042322	133	29

Table 2: Number of articles and extracted frames per event in the test dataset

A human annotator reviewed all extracted frames and scored each using one of three possible scores:

1 - **Ok**: sound meaning, relevant to event content,

2 - **Partial**: partially ok, but some part of the frame either off in meaning or missing; as a rule of thumb this score was used when a single role in the frame was off,

3 - **Irrelevant or wrong**: the frame is to general or nebulous, does not convey information about the event or meaning is wrong with respect to the event.

The distribution of frame scores over all events is shown in the chart in Figure 3. The results show that most of the data is at least partially correct with roughly a third of the frames being scored as completely unsatisfactory.

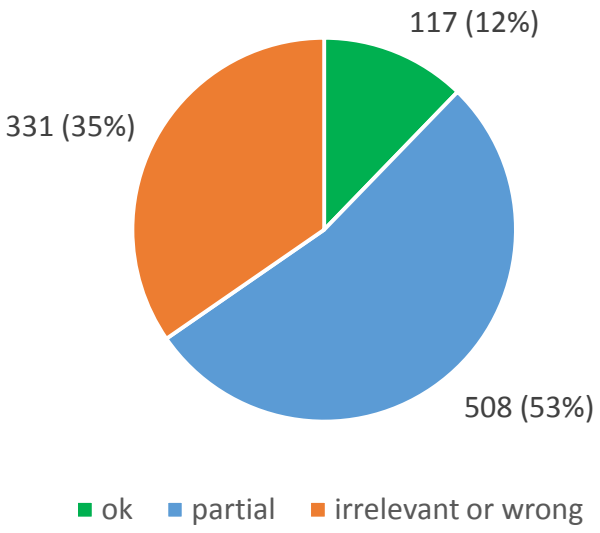

Figure 3: Distribution of scores over all events

The distribution of frame scores per event is shown in the chart in Figure 4.

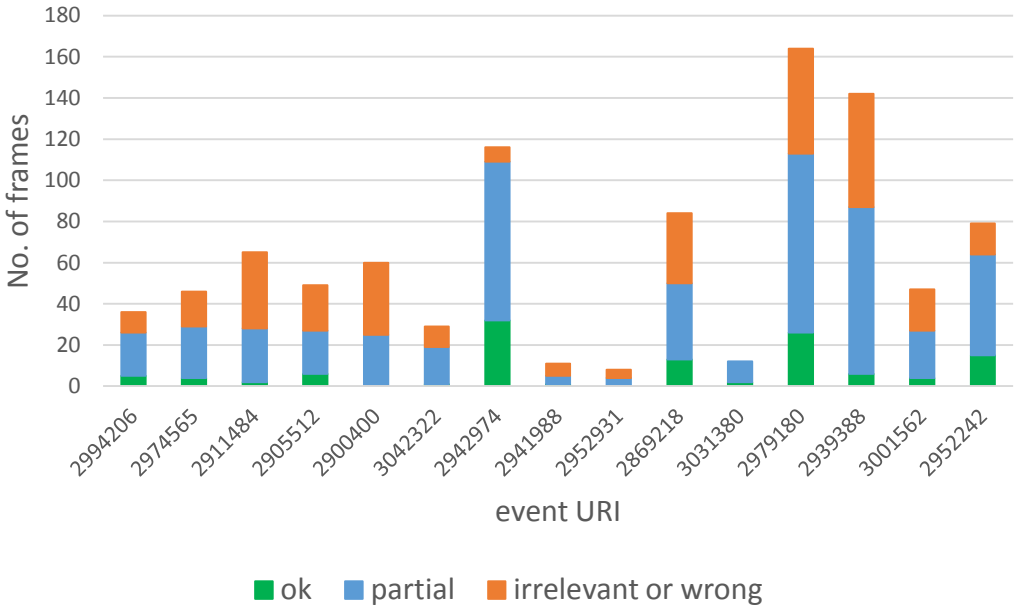

Figure 4: Distribution of frame scores per event

4 Future work

Results of our (relatively small) experiment, though encouraging, show that there is still plenty of room for improvement. By going over the extracted frames manually, we concluded that frame merging could still be improved. Wordnet alone does not cover many cases, especially if they are too specific (e.g. 'Apple sells *smartphones*' and 'Apple sells *iPhones*'). We experimented using xling [12], a Wikipedia-based semantic similarity measure, but so far it turned out to be prohibitively slow. We are also exploring other semantic similarity measures based on other knowledge bases, such as BabelNet [13] and Wiktionary [14]. A separate line of work is to link the concepts annotated by the tools from deliverable D4.1 [10] into the extracted frames taking advantage of the disambiguated values produced. We also plan to enhance the semantic graph by including links extracted from data of other modalities such as linked video segments and related social media.

5 Conclusion

This deliverable describes the Early Semantic Graph construction Prototype. The initial prototype focuses on extraction of structured semantic data from events – clusters of news articles describing the same real-world event. This is achieved through aggregation of semantic frames extracted from individual articles. We devised a computationally tractable approach which allows us to merge semantically equivalent frames, which both makes the resulting graph cleaner as well as accumulates the support of semantically important frames otherwise fragmented over several different literal frames. We performed a small evaluation of the extracted frames by manually scoring the extractions for a dataset of recent trending events. The results are encouraging but indicate several possible avenues for improvement. We also report on the extensions made to the Event Registry ontology to allow for inclusion of new data added by the xLiMe project.

References

[1] Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." *Scientific American* 284, no. 5 (2001): 28-37.

[2] https://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html

[3] https://developers.facebook.com/docs/graph-api

[4] http://eventregistry.org/

[5] http://newsfeed.ijs.si/

[6] http://www.planet-data.eu/

[7] xLiMe deliverable D1.1 – Prototype of (meta) Data Model

[8] xLiMe deliverable D1.2 – Prototype of Data Processing Infrastructure

[9] XLike deliverable D2.2.2 – Final Deep Linguistic Processing Prototype

[10] xLiMe deliverable D4.1 – Statistical Content Linking Prototype

[11] Miller, George A., “WordNet: A Lexical Database for English.” Communications of the ACM Vol. 38, No. 11: 39-41. 1995.

[12] http://aidemo.ijs.si/xling/wikipedia.html

[13] Navigli, Roberto, and Simone Paolo Ponzetto. "BabelRelate! A Joint Multilingual Approach to Computing Semantic Relatedness." *AAAI*. 2012.

[14] Zesch, Torsten, Christof Müller, and Iryna Gurevych. "Using Wiktionary for Computing Semantic Relatedness." *AAAI*. Vol. 8. 2008.

Annex A Event Registry ontology

```
<rdf:RDF
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:nsl="http://purl.org/ontology/storyline/"
  xmlns:rnews="http://iptc.org/std/rNews/2011-10-07#"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:event="http://purl.org/NET/c4dm/event.owl#"
  xmlns:er="http://eventregistry.org/rdf/ontology/"
>
    <owl:Ontology rdf:about="http://eventregistry.org/rdf/ontology/">
        <dcterms:title xml:lang="en">Event Registry Ontology</dcterms:title>
        <dcterms:created
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2013-11-
18</dcterms:created>
        <rdfs:comment xml:lang="en">The Event Registry Ontology models
information on global events and storylines across languages, domains and
granularities.</rdfs:comment>
        <foaf:maker>
            <foaf:Person rdf:about="http://www.linkedin.com/pub/alja%C5%BE-
ko%C5%A1merlj/5a/a7a/170">
                <foaf:name>Aljaž Košmerlj</foaf:name>
            </foaf:Person>
        </foaf:maker>
        <foaf:maker>
            <foaf:Person rdf:about="http://harth.org/andreas/foaf#ah">
                <foaf:name>Andreas Harth</foaf:name>
            </foaf:Person>
        </foaf:maker>
        <owl:versionInfo
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">1.0</owl:versionInfo>
    </owl:Ontology>
    <owl:Class rdf:about="http://eventregistry.org/rdf/ontology/Event">
        <rdfs:subClassOf rdf:resource="http://purl.org/ontology/storyline/Event"
/>
        <rdfs:label xml:lang="en">Event Registry Event Class</rdfs:label>
        <rdfs:comment xml:lang="en">Event Registry Event Class, a subclass of
nsl:Event. It models information about global events as it is collected by the
Event Registry application.</rdfs:comment>
        <rdfs:isDefinedBy rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:Class>
```

```
    <owl:Class rdf:about="http://eventregistry.org/rdf/ontology/Storyline">
        <rdfs:subClassOf
rdf:resource="http://purl.org/ontology/storyline/Storyline" />
        <rdfs:label xml:lang="en">Event Registry Storyline Class</rdfs:label>
        <rdfs:comment xml:lang="en">Event Registry Storyline Class, a subclass
of nsl:Storyline. It connects events found to be similar by the Event Registry
application into a common storyline. The events are ordered by
date.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:Class>
    <owl:Class rdf:about="http://eventregistry.org/rdf/ontology/Article">
        <rdfs:subClassOf        rdf:resource="http://iptc.org/std/rNews/2011-10-
07#NewsItem" />
        <rdfs:label xml:lang="en">Event Registry Article Class</rdfs:label>
        <rdfs:comment xml:lang="en">Event Registry Article Class, a subclass of
rnews:NewsItem. Models news items collected by the Event Registry application
news feed.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:Class>
    <owl:ObjectProperty
rdf:about="http://eventregistry.org/rdf/ontology/hasNewsArticle">
            <rdfs:domain
rdf:resource="http://eventregistry.org/rdf/ontology/Event" />
            <rdfs:range
rdf:resource="http://eventregistry.org/rdf/ontology/Article" />
            <rdfs:label xml:lang="en">has news article</rdfs:label>
            <rdfs:comment   xml:lang="en">Lists   news   articles   about   the
event.</rdfs:comment>
            <rdfs:isDefinedBy
rdf:resource="http://eventregistry.org/rdf/ontology/" />
      </owl:ObjectProperty>
    <owl:ObjectProperty
rdf:about="http://eventregistry.org/rdf/ontology/factorEntity">
            <rdfs:domain
rdf:resource="http://purl.org/NET/c4dm/event.owl#Factor" />
            <rdfs:range rdf:resource="http://www.w3.org/2002/07/owl#Thing" />
            <rdfs:label xml:lang="en">factor entity</rdfs:label>
            <rdfs:comment xml:lang="en">Specifies the entity that is a factor in
the event. Primarily the range are DBpedia resources.</rdfs:comment>
            <rdfs:isDefinedBy
rdf:resource="http://eventregistry.org/rdf/ontology/" />
      </owl:ObjectProperty>
    <owl:DatatypeProperty
rdf:about="http://eventregistry.org/rdf/ontology/factorScore">
        <rdfs:domain rdf:resource="http://purl.org/NET/c4dm/event.owl#Factor" />
        <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer" />
```



```
        <rdfs:label xml:lang="en">a factor score</rdfs:label>

        <rdfs:comment xml:lang="en">Property for expressing a factor relevance.
Datatype is Integer.</rdfs:comment>

        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>

    </owl:DatatypeProperty>

    <owl:DatatypeProperty
rdf:about="http://eventregistry.org/rdf/ontology/date">

        <rdfs:domain  rdf:resource="http://eventregistry.org/rdf/ontology/Event"
/>

        <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#date" />

        <rdfs:label xml:lang="en">date of event</rdfs:label>

        <rdfs:comment xml:lang="en">Date when the event occured. If endDate is
specified, this is the date when the event started. Datatype is
Date.</rdfs:comment>

        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>

    </owl:DatatypeProperty>

    <owl:DatatypeProperty
rdf:about="http://eventregistry.org/rdf/ontology/endDate">

        <rdfs:domain  rdf:resource="http://eventregistry.org/rdf/ontology/Event"
/>

        <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#date" />

        <rdfs:label xml:lang="en">ending date of event</rdfs:label>

        <rdfs:comment xml:lang="en">Date when the event ended. Datatype is
Date.</rdfs:comment>

        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>

    </owl:DatatypeProperty>

    <owl:DatatypeProperty
rdf:about="http://eventregistry.org/rdf/ontology/numberOfAllArticles">

        <rdfs:domain  rdf:resource="http://eventregistry.org/rdf/ontology/Event"
/>

        <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer" />

        <rdfs:label xml:lang="en">number of all articles</rdfs:label>

        <rdfs:comment xml:lang="en">Number of all collected news articles about
the event. Datatype is Integer.</rdfs:comment>

        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>

    </owl:DatatypeProperty>

    <owl:ObjectProperty
rdf:about="http://eventregistry.org/rdf/ontology/relatedImage">

            <rdfs:domain
rdf:resource="http://eventregistry.org/rdf/ontology/Event" />

            <rdfs:range
rdf:resource="http://eventregistry.org/rdf/ontology/Image" />

            <rdfs:label xml:lang="en">has related image</rdfs:label>

            <rdfs:comment   xml:lang="en">Property   for   relating   images   to
```

```
events.</rdfs:comment>
            <rdfs:isDefinedBy
rdf:resource="http://eventregistry.org/rdf/ontology/" />
      </owl:ObjectProperty>
    <owl:Class
rdf:about="http://eventregistry.org/rdf/ontology/NumberOfArticlesPerLanguage">
        <rdfs:label xml:lang="en">Event Registry Number Of Articles Per Language
Class</rdfs:label>
        <rdfs:comment xml:lang="en">Event Registry Number Of Articles Per
Language Class holds information about numbers of collected articles per
language about the event.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:Class>
    <owl:ObjectProperty
rdf:about="http://eventregistry.org/rdf/ontology/numberOfArticlesPerLanguage">
            <rdfs:domain
rdf:resource="http://eventregistry.org/rdf/ontology/Event" />
            <rdfs:range
rdf:resource="http://eventregistry.org/rdf/ontology/NumberOfArticlesPerLanguage"
/>
            <rdfs:label      xml:lang="en">number      of      articles      per
language</rdfs:label>
            <rdfs:comment xml:lang="en">Specifies the number of articles about
the event per language.</rdfs:comment>
            <rdfs:isDefinedBy
rdf:resource="http://eventregistry.org/rdf/ontology/" />
      </owl:ObjectProperty>
    <owl:DatatypeProperty
rdf:about="http://eventregistry.org/rdf/ontology/numberOfArticles">
        <rdfs:domain
rdf:resource="http://eventregistry.org/rdf/ontology/ArticleNumbers" />
        <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer" />
        <rdfs:label xml:lang="en">number of articles</rdfs:label>
        <rdfs:comment xml:lang="en">Number of collected news articles about the
event. Datatype is Integer.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:DatatypeProperty>
    <owl:DatatypeProperty
rdf:about="http://eventregistry.org/rdf/ontology/language">
        <rdfs:domain
rdf:resource="http://eventregistry.org/rdf/ontology/NumberOfArticlesPerLanguage"
/>
        <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string" />
        <rdfs:label xml:lang="en">language of articles</rdfs:label>
        <rdfs:comment xml:lang="en">Language of articles specified as a ISO 639-
1 two letter code. Datatype is String.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
```



```
/>
    </owl:DatatypeProperty>
    <owl:DatatypeProperty
rdf:about="http://eventregistry.org/rdf/ontology/similarity">
        <rdfs:domain
rdf:resource="http://eventregistry.org/rdf/ontology/Article" />
        <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer" />
        <rdfs:label xml:lang="en">similarity of article</rdfs:label>
        <rdfs:comment xml:lang="en">Property measuring article similarity to the
event as computed by the Event Registry application. Datatype is
Integer.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:DatatypeProperty>
    <owl:Class rdf:about="http://eventregistry.org/rdf/ontology/Category">
        <rdfs:label xml:lang="en">Event Registry Event Category</rdfs:label>
        <rdfs:comment xml:lang="en">Event Registry Event Source holds a dmoz
category the event and its weight.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:Class>
    <owl:ObjectProperty
rdf:about="http://eventregistry.org/rdf/ontology/inCategory">
            <rdfs:domain
rdf:resource="http://eventregistry.org/rdf/ontology/Event" />
            <rdfs:range
rdf:resource="http://eventregistry.org/rdf/ontology/Category" />
            <rdfs:label xml:lang="en">category of the event</rdfs:label>
            <rdfs:comment xml:lang="en">Specifies one of the dmoz categories the
event is in.</rdfs:comment>
            <rdfs:isDefinedBy
rdf:resource="http://eventregistry.org/rdf/ontology/" />
      </owl:ObjectProperty>
    <owl:DatatypeProperty
rdf:about="http://eventregistry.org/rdf/ontology/weight">
        <rdfs:domain
rdf:resource="http://eventregistry.org/rdf/ontology/Category" />
        <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer" />
        <rdfs:label xml:lang="en">weight of category</rdfs:label>
        <rdfs:comment xml:lang="en">Property measuring how strongly the event
belongs in a category. Datatype is Integer.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:DatatypeProperty>
    <owl:Class rdf:about="http://eventregistry.org/rdf/ontology/ArticleSource">
        <rdfs:label xml:lang="en">Event Registry Article Source</rdfs:label>
        <rdfs:comment xml:lang="en">Event Registry Article Source specifies the
```

```
title of the article publisher and the publisher's website.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:Class>
    <owl:ObjectProperty
rdf:about="http://eventregistry.org/rdf/ontology/isDuplicateOf">
        <rdfs:domain
rdf:resource="http://eventregistry.org/rdf/ontology/Article" />
        <rdfs:range rdf:resource="http://eventregistry.org/rdf/ontology/Article"
/>
        <rdfs:label xml:lang="en">original instance of the article</rdfs:label>
        <rdfs:comment xml:lang="en">Specifies the article of which this one is a
duplicate.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:ObjectProperty>
    <owl:ObjectProperty
rdf:about="http://eventregistry.org/rdf/ontology/hasRelatedMedia">
        <rdfs:domain  rdf:resource="http://eventregistry.org/rdf/ontology/Event"
/>
        <rdfs:range rdf:resource="http://rdfs.org/sioc/ns#Post" />
        <rdfs:range rdf:resource="http://www.w3.org/ns/ma-ont#MediaResource" />
        <rdfs:label xml:lang="en">has related media</rdfs:label>
        <rdfs:comment  xml:lang="en">This  property  relates  an  event  with  an
instance of media in a different modality like a social media post, an audio
clip or a video segment.</rdfs:comment>
        <rdfs:isDefinedBy  rdf:resource="http://eventregistry.org/rdf/ontology/"
/>
    </owl:ObjectProperty>
</rdf:RDF>
```

Annex B Events from the manual evaluation dataset

- http://eventregistry.org/event/2999085

  title: Rare blue moon comes Friday

- http://eventregistry.org/event/3001132

  title: Wife of dead officer says RCMP made her husband a 'scapegoat' in Dziekanski death

- http://eventregistry.org/event/2913728

  title: Kung Fury

- http://eventregistry.org/event/3031380

  title: Inside Amazon: Wrestling Big Ideas in a Bruising Workplace

- http://eventregistry.org/event/2939388

  title: Pastor among 9 killed at Charleston AME church

- http://eventregistry.org/event/2894624

  title: BB King, Defining Bluesman for Generations, Dies at 89 - New York Times

- http://eventregistry.org/event/2952242

  title: The Latest: Gay-marriage ruling another blow to GOP beliefs

- http://eventregistry.org/event/2971027

  title: Mexican drug lord Joaquin 'El Chapo' Guzman taunts the world after his escape from prison

- http://eventregistry.org/event/2900400

  title: In Early Vote Count, Ireland Appears Headed Toward Legalizing Same-Sex Marriage

- http://eventregistry.org/event/3032147

  title: History made: Army Ranger School to graduate its first female students ever - Washington Post

- http://eventregistry.org/event/2929714

  title: Kalief Browder, 1993-2015

- http://eventregistry.org/event/2974565

  title: Boehner Wants Investigation On Whether Planned Parenthood Is Selling Organs From Aborted Fetuses

- http://eventregistry.org/event/2905512

  title: Famed 'A Beautiful Mind' mathematician John Nash, wife killed in taxi crash, police say

- http://eventregistry.org/event/2911300

  title: Fifa crisis live: Visa threatens position as major sponsors start to turn

- http://eventregistry.org/event/3043420

  title: France train shooting: Attack 'was well prepared' - BBC News