FRED as an Event Extraction Tool

Aldo Gangemi[1,2], Ehab Hassan[1], Valentina Presutti[2], Diego Reforgiato Recupero[2]

[1] LIPN, Université Paris13-CNRS-SorbonneCité, France
[2] STLab, ISTC-CNR, Rome-Catania, Italy.

Events are elusive entities; as the authors of [7] argue, even human annotators do not agree on what is an event and what is its boundary in terms of the extension of its participants, temporal and geospatial extent, etc.

More aspects of events appear when trying to recognize or extract them from text: polarity of speaker's judgment on events, negation, modality, relations (temporal, causal, declarative, etc.) to other events, etc.

For example, the text:

The Black Hand might not have decided to barbarously assassinate Franz Ferdinand after he arrived in Sarajevo on June 28th, 1914.

expresses three events (*decide, assassinate, arrive*), with *Black Hand* being a participant in two of them, *Franz Ferdinand* in the third (*arrive*), a temporal extent for the third (*June 28th, 1914*), and a relative temporal extent for the other two (given the third's extent and the past tense suffixes in the first and third), a geo-spatial extent (*Sarajevo*), a judgment with negative polarity on the second event (*barbarously*), a negation (*not*) over the modality (*might*) modifying the first event, and an explicit temporal relation between the second and third event (*after*).

Extracting, logically representing, and connecting elements from a sentence is crucial to create semantic applications that are event-aware. In addition, it's important to disambiguate as much as possible the entities and concepts expressed, in order to make the extracted model *linked*, and to exploit the full power of the Semantic Web and Linked Data.

FRED[1] [5] is a tool to automatically transform knowledge extracted from text into RDF and OWL, i.e. it is a *machine reader* [2] for the Semantic Web. It is event-centric, therefore it natively supports event extraction. In a recent landscape analysis of knowledge extraction tools [3], FRED has got .73 precision, .93 recall, and .87 accuracy, largely better than the other tools attempting event extraction.

FRED is available as a RESTful API and as a web application. In its current form, it relies upon several NLP components: Boxer[2] for the extraction of the basic logical form of text and for disambiguation of events to VerbNet, UKB[3] or IMS[4] or BabelNet API[5] for word sense disambiguation, and Apache Stanbol[6] for named entity resolution.

---

[1] `http://wit.istc.cnr.it/stlab-tools/fred`
[2] `http://svn.ask.it.usyd.edu.au/trac/candc/wiki/boxer`
[3] http://ixa2.si.ehu.es/ukb/
[4] http://www.comp.nus.edu.sg/~nlp/sw/
[5] http://lcl.uniroma1.it/babelnet/
[6] http://stanbol.apache.org

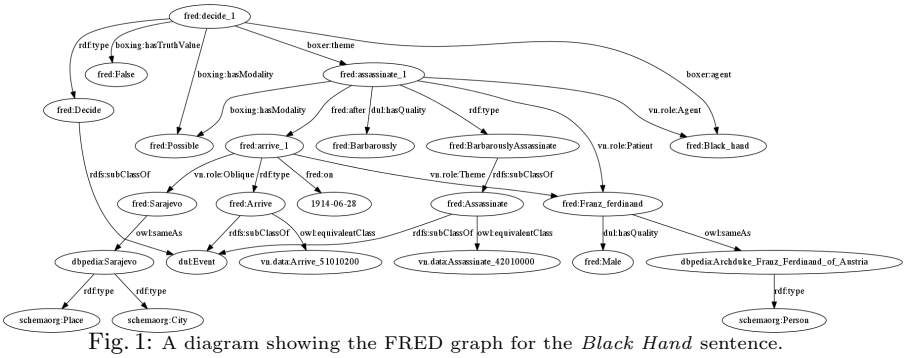

Fig. 1: A diagram showing the FRED graph for the *Black Hand* sentence.

FRED contains several functionalities for event extraction, which can be summarized according to typical subtasks:

- Event identity: FRED focuses on events expressed by *verbs*, *propositions*, *common nouns*, and *named entities* (typically proper nouns).
- Event classification: FRED uses Linked Data-oriented induction of types for the identified events, reusing e.g. VerbNet[7], WordNet[8], DBpedia[9], schema.org, and DOLCE[10] as reference ontologies.
- Event unity: FRED applies *semantic role labeling* [4] to verbs and propositions in order to detect event boundaries, and *frame detection* [1] for resolving roles against a shared event ontology.
- Event modifiers: FRED extracts *logical negation*, *basic modalities*, and *adverbial qualities*, applied to verbs and propositions, which can then be used as event judgment indicators.
- Event relations: FRED relates events via the role structure of verbs and propositions, and extracts *tense relations* between them.

The beginning and the following sentences are used as a lead example for showing FRED's functionalities:

The Renaissance was a cultural movement that spanned in Italy from the 14th to the 17th century. Some sources report that the Renaissance might have been started by Greek scholars from Constantinople.

In the diagram from Figure 2, the following events are recognized, extracted, classified, and aligned to WordNet, VerbNet, and/or DOLCE: `Renaissance` (classified as a `Movement`, and aligned to the WordNet `Motion` synset, and to the DOLCE `Situation` class), `span_1`, `report_1`, and `start_1` (classified as occurrences of the `Span`, `Report` and `Start` frames respectively, and aligned to VerbNet).

Furthermore, the events have participants (e.g. `Italy`, `scholar_1`, `source_1`, etc., also classified and linked appropriately) through some roles labelled with properties derived from VerbNet(e.g. `vn:Agent`), or from the lexicon used in the sentence (e.g. `ren:from`) In one case, a modal modifier (`Possible`) to the event `start_1` is added.

[7] http://verbs.colorado.edu/~mpalmer/projects/verbnet.html?

[8] http://wordnet.princeton.edu

[9] http://dbpedia.org

[10] http://www.ontologydesignpatterns.org/ont/dul/DUL.owl

Finally, some relations between events are detected: `report_1 vn:Theme start_1`, and `span_1 before report_1` (through the `now_1` interval).

See also Figure 1 for the graph obtained from the beginning sentence.

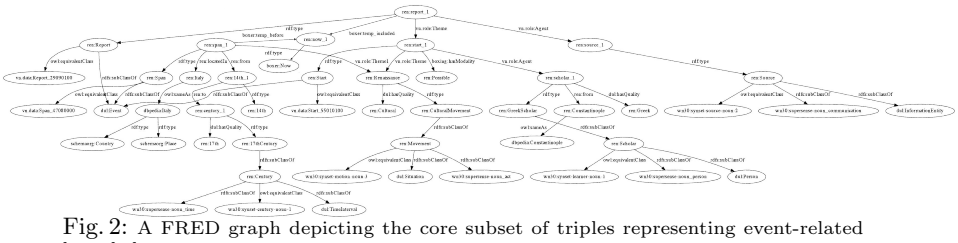

Fig. 2: A FRED graph depicting the core subset of triples representing event-related knowledge.

The triples given as output by FRED are more than those visualized, for example they include text spans and their reference to the semantic annotations, through the Earmark vocabulary [6].

FRED is therefore an intermediate component for event extraction and representation, which can be augmented with background knowledge, and whose graphs can be combined e.g. in time series for historical tasks.

FRED will be demoed as an event extractor by showing event-intensive sentences, and examples of views that focus on relevant event knowledge. RDF models can be morphed to concentrate on specific features. For example, Figure 3 semantically summarizes the model from the *Black Hand* sentence by only showing events with their relations, and their main participant, obtained by means of the following SPARQL query:

```
PREFIX dul: <http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#>
PREFIX vnrole: <http://www.ontologydesignpatterns.org/ont/vn/abox/role/>
PREFIX boxing: <http://www.ontologydesignpatterns.org/ont/boxer/boxing.owl#>
PREFIX boxer: <http://www.ontologydesignpatterns.org/ont/boxer/boxer.owl#>
PREFIX : <http://www.ontologydesignpatterns.org/ont/boxer/test.owl#>
CONSTRUCT {?e :agent ?x . ?e ?r ?e1}
WHERE {
 {{?e a boxing:Situation} UNION {?e a ?class . ?class rdfs:subClassOf+ dul:Event}}
 ?e ?p ?x
 FILTER (?p = vnrole:Agent || ?p = boxer:agent || ?p = vnrole:Experiencer || ?p = vnrole:Actor
     || ?p = vnrole:Actor1 || ?p = vnrole:Actor2 || ?p = vnrole:Theme)
 FILTER NOT EXISTS {?e vnrole:Theme ?x . ?e vnrole:Agent ?y
 FILTER (?x != ?y)}
 OPTIONAL {{{?e ?r ?e1} UNION {?e ?s ?z . ?z ?t ?e1}} {{?e1 a boxing:Situation} UNION
    {?e1 a ?class1 . ?class1 rdfs:subClassOf+ dul:Event}} FILTER (?e != ?e1)}}
```

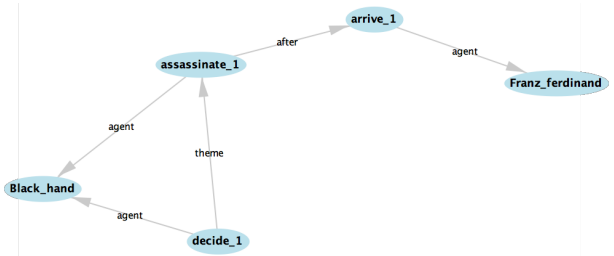

Fig. 3: A summarized FRED graph showing only event relations and agentive participants for the Black Hand sentence.

References

1. Bonaventura Coppola, Aldo Gangemi, Alfio Massimiliano Gliozzo, Davide Picca, and Valentina Presutti. Frame detection over the semantic web. In Lora Aroyo et al., editor, *ESWC*, volume 5554 of *LNCS*, pages 126–142. Springer, 2009.
2. Oren Etzioni, Michele Banko, and Michael Cafarella. Machine reading. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, 2006.
3. Aldo Gangemi. A comparison of knowledge extraction tools for the semantic web. In *Proceedings of ESWC2013*. Springer, 2013.
4. A. Moschitti, D. Pighin, and R. Basili. Tree kernels for semantic role labeling. *Computational Linguistics*, 34(2):193224, 2008.
5. Valentina Presutti, Francesco Draicchio, and Aldo Gangemi. Knowledge extraction based on discourse representation theory and linguistic frames. In *EKAW: Knowledge Engineering and Knowledge Management that matters*. Springer, 2012.
6. Peroni S., Gangemi A., and Vitali F. Dealing with markup semantics. In *Proceedings of the 7th International Conference on Semantic Systems, Graz, Austria (i-Semantics2011)*. ACM, 2011.
7. Chris Welty and Lora Aroyo. Harnessing disagreement for event semantics. In *Proc. of DERIVE Wks at ISWC2012*, 2012.