Identifying motifs for evaluating open knowledge extraction on the Web

Aldo Gangemi [a,b,*], Diego Reforgiato Recupero [b,c], Misael Mongiovì [b], Andrea Giovanni Nuzzolese [b], Valentina Presutti [b]

[a] Universitè Paris 13, Sorbonne Cité, CNRS, Paris, France
[b] ISTC-CNR, via Gaifami, 18, 95126 Catania, Italy
[c] Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy

ARTICLE INFO

ABSTRACT

Open Knowledge Extraction (OKE) is the process of extracting knowledge from text and representing it in formalized machine readable format, by means of unsupervised, open-domain and abstractive techniques. Despite the growing presence of tools for reusing NLP results as linked data (LD), there is still lack of established practices and benchmarks for the evaluation of OKE results tailored to LD. In this paper, we propose to address this issue by constructing RDF graph banks, based on the definition of logical patterns called *OKE Motifs*. We demonstrate the usage and extraction techniques of motifs using a broad-coverage OKE tool for the Semantic Web called FRED. Finally, we use identified motifs as empirical data for assessing the quality of OKE results, and show how they can be extended trough a use case represented by an application within the Semantic Sentiment Analysis domain.

1. Introduction

Translating natural language text to formal data that can be used or integrated into knowledge bases is an important research task due to its applications in intelligent systems and data science, and therefore it is central to the Semantic Web (SW) community. One of the main open challenges is to establish shared practices and benchmarks for evaluating its results.

In recent years, the production of structured data from text has become scalable. Machine reading is a good example. In [1], the machine reading paradigm is defined as a procedure for extracting knowledge from text by relying on bootstrapped, self-supervised Natural Language Processing (NLP) performed on basic tasks. Machine reading can process massive amounts of text in reasonable time, can detect regularities hardly noticeable by humans, and its results can be reused by machines for applied tasks. The same techniques can be combined with logic-oriented approaches in order to produce formal knowledge from text, i.e., to perform OKE, which can be defined as *the extraction of knowledge from text, and its representation in formalized machine readable form* (see [2] for

a survey on web data extraction and [3] for the work that introduced OKE). OKE is unsupervised, open-domain, and abstractive.[1] A key problem that has not been solved yet is how machine reading tools can be evaluated and compared without available benchmarks or best practices. How can we measure the precision and recall of a method that structures unstructured text or produces formal knowledge? When machine reading tools need to be compared, data-sets are built and annotated according to some guidelines and gold standards are thus created for the underlying domain of application. As an example, authors in [5] show the annotation efforts for an entire year of NewsReader, a European project related to financial and economic data for decision making. They defined the guidelines and several features (such as entity type, factuality, certainty, polarity and time attribute, etc.) without using any formal framework that could help them with the formalization process. The NLP community is also moving towards similar objectives, e.g., with the AMR initiative [6]. AMR implements a simplified, standard neo-Davidsonian semantics using standard feature structure representation where predicates senses and core semantic roles are drawn from the OntoNotes project.[2]

---

* Corresponding author.
   E-mail addresses: aldo.gangemi@istc.cnr.it (A. Gangemi), diego.reforgiato@unica.it (D.R. Recupero), misael.mongiovi@istc.cnr.it (M. Mongiovì), andrea.nuzzolese@istc.cnr.it (A.G. Nuzzolese), valentina.presutti@istc.cnr.it (V. Presutti).

---

[1] *Abstractive* means that the result of text analysis is not a (set of) text segment(s), but rather a *representation* of a text in a knowledge representation language, cf. [4] for a definition of *abstractive* techniques in NLP.
[2] https://catalog.ldc.upenn.edu/LDC2013T19.

Which formal semantics should be employed when reusing machine reading output is not yet agreed. For example, knowledge extraction for the SW is mostly evaluated based on NLP benchmarks (cf. the discussion in [7] and the recent work in [8]). Although they provide solutions for a wide set of SW methods, there are still many metrics, quality measures and problems left uncovered. Moreover, there is nothing there that allows an organization of the tree banks in a structural way with respect to OKE and SW tasks. We argue the urgency and opportunity to define OKE tasks and their associated benchmarks, which would provide the SW community with a native platform to assess research advancement. We think that the right direction is creating *RDF graph banks*, which would radically change OKE research similarly to how syntactic tree-banks [9] did in computational linguistics. A tree-bank is a text corpus annotated with a syntactic sentence structure, providing large-scale empirical data for NLP tasks evaluation. An RDF graph bank is a text corpus annotated with an RDF graph structure. Extending the approach of treebanks, Ontonotes [10], Groeningen Meaning Bank [11], and the semantic banks expressed in Abstract Meaning Representation [6], RDF graph banks can be validated by experts in both SW and linguistics, and can be used as benchmarks for evaluating tools, for learning machine reading models, and to design new tools.

In this paper we identify RDF "motifs" that are as close as possible to good practices in SW and LD. Some elementary patterns (motifs) are defined in order to partition any graph bank into semantically homogeneous subsets. Such motifs correspond to typical logical patterns used in SW and LD. Then we build two sample RDF graph banks (extracted from 100 text sentences and 151 text sentences), and show how they can be validated and refined by RDF experts.

The paper is organized as follows. Section 2 presents the background context of the problem. In Section 3 we describe FRED, the machine reader we have developed and whose graphs we have used as sources for motif identification and for the production of the sample RDF graph bank. In Section 4 we formally define motifs. In Section 5 we present a list of relevant motifs and show how to identify them. Section 6 describes two examples of graph banks created by using motifs, and show how it can be used to evaluate SW tasks. Section 7 shows how we have extended, derived and identified the identified motifs to create a successful application [12,13] of Semantic Sentiment Analysis showing the sentiment motifs. Section 8 ends the paper with conclusions, challenges and possible directions for SW machine reading and graph banks.

2. Background

NLP and SW. The integration between Natural Language Processing (NLP) and SW, under the hat of "semantic technologies", is progressing fast. Most work has been opportunistic: on the one hand exploiting NLP algorithms and applications (typically named-entity recognizers and sense taggers) to populate SW data-sets or ontologies, or for creating NL query interfaces; on the other hand exploiting large SW data-sets and ontologies (e.g., DBpedia,[3] YAGO,[4] Freebase,[5] etc.) to improve NLP algorithms. For example, large text analytics and NLP projects such as Open Information Extraction (OIE, [14]), Alchemy API,[6] and Never Ending Language Learning (NELL, [15]), perform grounding of extracted named entities in publicly available identities such as Wikipedia, DBpedia and Freebase. The links between the two areas are becoming tighter,

and clearer practices are evidently needed. Standardization attempts have been introduced with reference to linguistic resources (WordNet,[7] FrameNet,[8] and the growing linguistic linked open data cloud), and the recent proposal of Ontolex-Lemon by the Ontolex W3C Community Group[9] will possibly improve resource reuse. Recently, platforms such as Apache Stanbol,[10] NIF [16] and the NLP2RDF project [17], NERD [18], FOX,[11] FRED [19][12] made it simpler to reuse NLP components as LD, as well as to evaluate them on reference benchmarks, as with GERBIL [8].[13]

Semantic interoperability issues. Interoperability efforts so far mainly focused on the direct transformation of NLP data models into RDF. When this is apparently simple, as in named entity resolution (a.k.a., entity linking), semantic interoperability problems are not so evident. On the contrary, with more advanced tasks, such as relation extraction, compositional analysis of terms, taxonomy induction, frame detection, etc., those problems become evident, and when different results should be combined in order to form a formally and pragmatically reliable ontology, advanced solutions are needed. In practice, even in entity linking, semantics is not as trivial as expected, and explicit assumptions have to be taken about what it means to represent e.g., both `dbpedia:Barack_Obama` and `dbpedia:African_American` as OWL individuals, or to create an `owl:sameAs` link between two resources. Classical work on ontology learning such as [20] takes the integration problem from a formal viewpoint, and uses linguistic features to extract occurrences of logical axioms, such as subclass of, disjointness, etc. Some work from NLP followed a similar direction [21], e.g., NELL relation properties and ontology [22], and "formal semantics" applied to NL (e.g., [23], [24]). These works assume some axiomatic forms, and make the extraction process converge to that form. This is good in principle, but the current state of the art does not really help with establishing clear cut criteria on how to convert NL extractions to RDF or OWL.

From the perspective of NLP, there are a (few) approaches from natural language formal semantics which output formal data structures, but they are not easily interpretable into SW languages. For example, Discourse Representation Structure (DRS), as shown in the output of Boxer [23], is a first-order logic data structure that heavily uses *discourse referents* as variables to anchor the predicates into extensional interpretations, and a *boxing* representation that contextualises the scope of logical (boolean, modal, inferential) operators. Both issues need non-trivial decisions on the side of RDF and OWL design, such as (i) what variables should be accommodated in a SW representation, or ignored? (ii) What logical operators can be safely represented in the formal semantics supported by SW languages? (iii) What predicates should be represented, and in which form, in RDF or OWL?

From the perspective of LD, even porting the original NLP tools data structures into RDF can be beneficial (cf. e.g., the LODifier method [25]), but the reuse of those data will require some intelligence to be integrated. Our stance is that LD are better served if NLP results are reused with a shared semantics that is ready to be integrated with existing RDF data. For example, if a NLP tool outputs data about *Barack Obama* (i.e., its roles, types, relations to other entities), we should be ready to integrate those data to e.g., http://dbpedia.org/resource/Barack_Obama , so that the integrated data preserve their semantics, modulo updates or

---

[7] https://wordnet.princeton.edu/.
[8] https://framenet.icsi.berkeley.edu/fndrupal/.
[9] http://www.w3.org/community/ontolex/wiki/Main_Page.
[10] http://stanbol.apache.org.
[11] http://aksw.org/Projects/FOX.html.
[12] http://wit.istc.cnr.it/stlab-tools/fred.
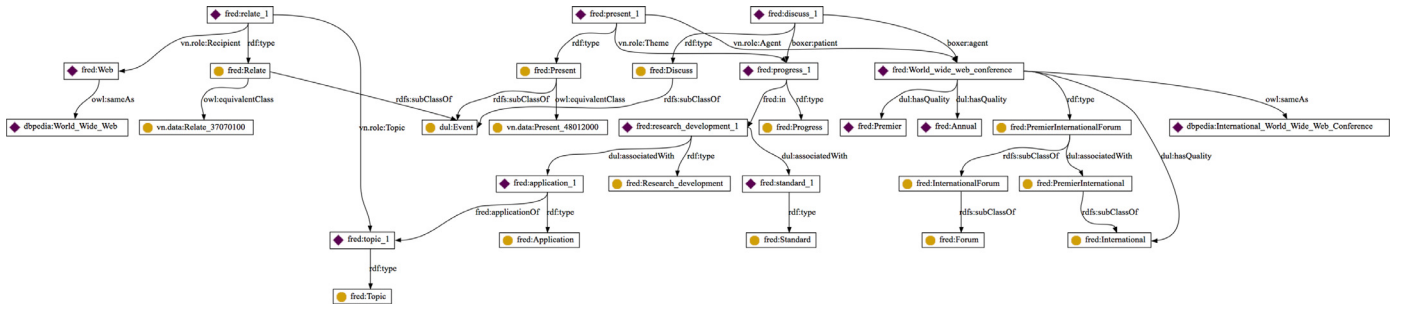[13] http://aksw.org/Projects/GERBIL.html.

Fig. 1. An example RDF graph produced by FRED for the sentence: *The annual World Wide Web Conference is the premier international forum to present and discuss progress in research, development, standards, and applications of the topics related to the Web*. Nodes with a yellow circle represent classes. Purple diamonds represent individuals. Literals are identified by green rectangles (not shown). Links represent properties.

non-monotonic conflicts. Currently this is made mostly in the context of *distantly supervised* approaches, which interpret NLP results in terms of existing LD or vocabularies, but do not attempt to generate full-fledged LD out of text. Some unsupervised approaches exist which attempt to produce LD from text for specific extraction tasks (e.g., entity linking [18] or relation extraction [26]), while the main tool that produces LD for broad-coverage and integrated extraction tasks is FRED [19].

3. FRED

FRED [19] is a machine reader for the SW that coherently integrates and represents multiple semantic parsing results such as frame detection, semantic role labeling, entity linking, taxonomy induction, etc. It has been successfully applied in semantic sentiment analysis, relation augmentation and discovery [7,12,13,19,27,28]. FRED generates RDF graph representations out of the knowledge extracted by components dedicated to basic NLP tasks. An example of a FRED graph is depicted in Fig. 1, whose caption includes an explanation on how to read the graph.

FRED formally represents, integrates, improves, and links the output of several NLP tools, which can be "plugged in", and are partly offered as options in the FRED API. FRED is also accessible by means of a Python API, *fredlib*,[14] which exposes features for retrieving FRED graphs from user-specified sentences, and managing them. The Python API hides details related to the communication with the FRED service, and returns a FRED graph object that is easily manageable. FRED graph objects expose methods for retrieving useful information, including the set of individual and class nodes, equivalences and type information, categories of FRED nodes (events, situations, qualities, general concepts) and categories of edges (roles and non-roles). It also includes a function to retrieve motifs as defined below. *fredlib* uses *curl* to send FRED each sentence, and to extract the semantic triples from the graph of each sentence. Moreover, *fredlib* reuses *rdflib* (for managing RDF graphs)[15] and *networkx*[16] (for managing complex networks) libraries.

Following an approach common in linguistic treebanks, FRED has also been used for bootstrapping an RDF graph bank annotated with explicit motifs. Annotation grounding is available directly from FRED by means of the Earmark vocabulary [29] and the NLP Interchange Format (NIF) [16], while *fredlib* provides the motif extraction and annotation.

4. OKE motifs

Evaluating against a SW graph bank. There is not much clarity about what should be evaluated as a good NLP-based knowledge extraction tool for the SW. Recently, we have presented the results of a first landscape analysis of existing tools [7], which proposes some mappings between traditional NLP tasks/data structures, and SW tasks/logical patterns. That was a landscape analysis, but proved useful to identify the main areas of vagueness in the interface. In this paper, we try to make that proposal more precise, in a *programmatical* way by providing the community with methods for creating OKE graph banks.

We have started building a *OKE graph bank* containing RDF named graphs extracted from individual sentences. These named graphs can be modularized into *motif occurrences* that match customizable *motifs*. Motifs are query patterns corresponding to logical RDF/OWL patterns found in extracted graphs. Named graphs are eventually evaluated by experts in order to produce gold standard graph banks.

Formally, a motif is a subgraph class M whose occurrences $m \in M$ occur in some graph $g \in G$, G being a graph class, so that $m \preceq g$, with $\preceq$ defined as the usual subgraph relation: with V, V_m being sets of vertices, E, E_m being sets of edges, $g = (V, E)$ and $m = (V_m, E_m)$, $V_m \subseteq V$, $E_m \subseteq E$.

Following practices in the semi-automatic production of linguistic tree banks [9,11], existing tools can be used to automatically produce the graphs to be evaluated. In the closest case to ours [11], the only large existing NLP meaning bank [11], the authors have used the tool that provides the output for the largest set of NLP semantic tasks (Boxer). In the previous landscape analysis [7], we had followed a similar procedure, but using the whole set of compared tools, and unifying their outputs. However, as noticed in that experience, the more tools are used with non-standardized output, the more effort has to be carried out by evaluators to make sense of them for SW tasks, even before starting to assess the ground truth for those tasks.

Therefore, since (again based on [7]), FRED appears to have the broadest coverage of SW tasks among the currently available tools, we have used it alone to produce a draft graph bank, and three independent expert annotators have established the ground truth to evolve those graphs into the first core of the machine reading graph bank we are envisaging.

Notice that a graph bank is not bound to fixed motifs; on the contrary, it can be designed according to own needs, and then used in order to evaluate specific tasks. Some motifs are intuitively basic for the SW, e.g., types, subclasses, equivalences, identity, property assertions, etc. A set of nine such basic motifs has been used in the sample evaluation presented in this paper.

Table 1

A summary of nine OKE motifs. The $\cdot^I$ symbol represents the extensional interpretation of a constant. *Event* and *Situation* are primitives in the built-in ontology of FRED formal representation (cf. [19]). The notation "|*X*|" denotes the class of predicate *X*. Also, please note that when a verb is negated, (e.g., *I did not like the movie*), FRED introduces the `boxing:hasTruthValue` relation between the verb instance and the `False` value. The examples are in RDF-Turtle (https://www.w3.org/TR/2014/REC-turtle-20140225/) syntax.

MOTIF	TYPE	DESCRIPTION	SEMANTICS
IDENTITY	Edge	`owl:sameAs` ex.: `local:Obama owl:sameAs dbr:Barack_Obama .`	$a^I = b^I$
TYPE	Edge	`rdf:type` ex.: `local:Obama rdf:type dbo:Person .`	$a^I \in C^I, C^I \in \texttt{owl:Class}$
SUBCLASS	Edge	`rdfs:subClassOf` ex.: `local:SkiResort rdfs:subClassOF local:Resort .`	$C^I \subseteq D^I$

EQUIVALENCE	Edge	`owl:equivalentClass`
		ex.: `local:Arrive owl:equivalentClass`
		`  verbnet:Arrive_51010200 .`
			$C^I \equiv D^I$

ROLE	Edge	A triple that defines the role of an entity. in an event or situation. ex.: `local:Arrive verbnet:Theme` `  local:ArchdukeFranzFerdinand .`	$(e^I, \tilde{a}^I) \in \|ThematicRole\|^I,$ $\texttt{where } e^I \in E^I \subseteq Event^I$
PROPERTY	Edge	Binary facts, i.e. any domain property assertion, (hence excludes ROLE, MODALITY, and NEGATION) ex.: `local:Paris local:capitalOf local:France .`	$(a^I, b^I) \in \|P\|^I,$ $\texttt{where } \|P\|^I \cap \|ThematicRole\|^I \subseteq \emptyset$
MODALITY	Edge	a modality mark applied to either events or situations. ex.: `local:Survive mod:hasModality mod:Possible .`	$(e^I, m^I) \in \|ModRel\|^I,$ $\texttt{where } e^I \in E^I \subseteq (Event \cup Situation)^I$
NEGATION	Edge	a negation mark applied to either events or situations. ex.: `local:Survive boxing:hasTruthValue neg:False .`	$(e^I, False^I) \in \|TruthRel\|^I,$ $\texttt{where } e^I \in E^I \subseteq (Event \cup Situation)^I$

FACT	N-ARY	A N-ARY subgraph. It may include:
		EVENT: An event (frame occurrence)
		connected to associated entities
		
		SITUATION: A situation (boolean-based)
		connected to associated entities
		
		ex.: `local:Arrive verbnet:Theme`
		`  local:ArchdukeFranzFerdinand ;`
		`  verbnet:Location local:Sarajevo .`
			$e^I \in E^I \subseteq Event^I$
			$(e^I, a^I) \in \|role_i\|^I, ..., (e^I, z^I) \in \|role_n\|^I$
			$\mathbf{s^I \in S^I \subseteq Situation^I}$
			$(s^I, a^I) \in \|role_i\|^I, ..., (s^I, z^I) \in \|role_n\|^I$

BINARIZED	Inferred	A binary relation inferred (projected) out of an N-ARY one ex.: `local:ArchdukeFranzFerdinand local:arrivesIn` `  local:Sarajevo .`	$(a^I, b^I) \in \|R\|^I,$ $\texttt{where } e^I \in E^I \subseteq (Event \cup Situation)^I,$ $(e^I, a^I) \in \|ThematicRole_i\|^I,$

COMPOSED	Inferred	A binary relation inferred
		by following a property path
		ex.: `local:ArchdukeFranzFerdinand`
		`  local:arrivesIn/local:capitalOf`
		`local:Bosnia .`
			$(a^I, b^I) \in \|R\|^I,$
			$\texttt{with } (a^I, c^I) \in \|R_i\|^I, (c^I, b^I) \in \|R_j\|^I$

5. Identifying motifs

Motifs are defined here as basic building blocks (subgraph patterns) of RDF graphs. In other words, an RDF graph can be seen as a combination of multiple applications of motifs (*motif occurrences*). The motifs we present in this paper include basic logical patterns such as class membership, sub-classing, binary relationship, negation, etc., as well as more "design-oriented" patterns such as N-ARY or path-based relationships. The OKE motifs presented here have been extracted from an empirical analysis of the output of FRED.

Motif classes. In Table 1 we report nine motifs we have identified within FRED's graphs.[17] We identified two classes of motifs from FRED's graphs, depending on their shape. A first class includes *"edge motifs"*, i.e., motifs that correspond to single triples. Those motifs correspond most easily to the output of semantic parsers from NLP, and are so basic that they cannot be reduced to more elementary ones. A second class, *"N-ARY motifs"*, refers to star-shaped motifs, where a "root" node is connected to a number of related nodes. An example of N-ARY motif is EVENT, correspond-

ing to a Neo-Davidsonian [30] event linked to multiple participants by means of thematic roles such as agent, patient, theme, experiencer, location, instrument, etc. Event types are *semantic frames* as defined in FrameNet or VerbNet. Other N-ARY motifs identify subgraphs that are not based on the detection of a traditional semantic frame. E.g., SITUATION holds together logical structures such as intersections or unions. Although N-ARY motifs can be reduced into multiple single edge motifs, they are still considered basic non-reducible blocks, since an occurrence of an N-ARY motif is fully determined by all its parts together.

The two classes include the nine motifs that we have used within the study in this paper. At the bottom of Table 1 we have also included two more motifs, *"inferred motifs"*, that contain the binary relation patterns that are produced by means of an inference, either by projecting binary relations out of N-ARY ones (BINARIZED motif), or by composing property paths (COMPOSED motif). Inferred motifs have emerged from the output of a new OKE tool, called Legalo [31], which materializes inferred relations by using either projection or path composition.

It remains intended that any other tool or multiple tools – as in the landscape analysis provided in [7] – can be used to create graph banks. The set of motifs presented here can also be

---

[17] The reader is invited to play with FRED to see occurrences of those motifs.

extended according to the particular output of alternative tools and resources.

Some tools, as e.g., analyzed in [7], indeed hint at e.g., more special cases of elementary triple-based motifs. For example, Re-Verb [32][18] is an extractive Open Information Extraction tool, and as such it extracts text segments when remarkable triplet segments are recognized, e.g., the segment It --- has some similarities to --- Simon from the sentence *It has some similarities to Simon, which is a wildly popular electronic game that was introduced in 1978.*

In this case, a special kind of binary relation is recognized, which, after appropriate representation into RDF, would not fit into any of the motifs included in the two classes described above, being then classified as a generic PROPERTY motif occurrence. In particular, a PROPERTY motif can be considered as a domain binary relation, distinct from a ROLE binary relation within an N-ARY motif, and from either MODALITY and NEGATION triples. Cases like these point at additional motifs to be discovered when new tools and tasks are proposed or analysed.

Finally, there are other possible classifications, e.g., TBox vs. ABox axioms as inspired by Description Logic (DL) semantics [33] used in OWL,[19] which are orthogonal to ours, and easily derivable.

6. Gold standard creation

In this section we describe the creation of two textual corpora (that we call here "Balanced Corpus" and "Evolutionary Corpus"), converted into OKE graph banks, which can be used to test SW tasks.

Balanced Corpus. For the first graph bank we have selected a corpus of 100 sentences,[20] balanced between online news, scientific reports, Twitter, and Wikipedia definitions. The four domains (25 sentences each) are supposed to demonstrate the open-domain assumption of machine reading. News sentences are related to New York Times articles about the civil war in Syria, scientific report sentences are from computer science papers citing each other (cf. [27] for more detail on this collection), tweets are extracted from the Twitter subject "Soccer fan" (only tweets with no hashtags, url or mentions have been considered in order to avoid any cleaning bias), and Wikipedia sentences are definitions of interlinked resources whose linking path starts from the Miles Davis (musician) page. The average sentence length is 17.3 words per sentence, with tweets being the shortest (11.7) and articles the longest (24.6).

Each of the 100 sentences from the first corpus has been sent to the FRED REST service[21] to obtain semantic and lexical RDF triples. *fredlib* has been used for such a purpose. The output triples have been organized in 100 named graphs, each identifying a sentence with a proper URI, and including the list of extracted triples for that sentence. Named graphs have been handled using the TriG syntax.[22] The total number of generated triples is 23,017, 4799 of which are purely semantic triples, 13,547 are annotation triples that annotate text segments according to the Earmark and NIF vocabularies, 4371 are metadata triples that annotate text segments with parts of speech, and 300 are metadata triples annotating named graphs with their identifiers, related sentences, and sentence topics. Several evaluation strategies can be devised con this

graph bank. In a first experiment, we asked independent experts in the domain of SW to evaluate the triples from this graph bank. The annotators could perform the following operations:

- delete a given triple obtained by FRED;
- modify some of its values (subject, relationship, object);
- add new triples to the generated ones.

To ease the work of the annotators, we provided guidelines that include the ontologies to be used, and some examples that provide explanations for the annotation task. This first evaluation experiment has been successfully applied to 40 graphs, leading to the graph changes,[23] that eventually lead to a gold standard. Examples of missing triples are: incomplete roles in the N-ARY motif, missing disambiguation in the EQUIVALENCE motif, missing resolution in the IDENTITY motif, missing NEGATION or MODALITY when a negation or modality element is present in the sentence, missing binary relation that is clearly expressed in the sentence (either PROPERTY or ROLE motifs), missing type or sub-classes that are clearly expressed in the sentence (TYPE or SUBCLASS motifs). An example of the validation task is a very bad graph produced by FRED, shown in Fig. 2, for the tweet sentence *David Moyes shares Manchester United fans' frustration.* The errors in this graph are due to a failure of the part-of-speech tagging component, which tags *share* as a noun instead of a verb.[24] Fig. 3 shows the corrected graph where users would edit and add missing triples to the corresponding motifs of such a sentence.

In particular, the users identify the missing EVENT motif that involves fred:share_1, fred:David_moyes and fred:frustration_1 that replaces the incorrect PROPERTY motifs dul:associatedWith, dul:hasQuality and fred:manchester outgoing from fred:share_1. They would also identify the missing PROPERTY motifs fred:frustration_1 fred:frustationOf fred:fan1 and fred:fan_1 fred:fanOf fred:manchester1, plus a few incorrect PROPERTY motifs.

Regarding OKE quality measures, precision can be obtained by checking the percentage of motif occurrences produced by the tool that are included in the gold standard, while recall can be computed as the percentage of motif occurrences in the gold standard that map motif occurrences from the tool. Note that this evaluation may consider a single motif, a group, or all motifs, e.g., assessing overall performance.

This experiment, however, showed also that careful expert evaluation requires a lot of time (9 human-days to evaluate 40 graphs), and we started considering alternative approaches. The first is the one adopted by Meaning Bank, which, however, seems to pose the same problems. Another approach is to involve generic crowdsourced evaluation, by removing any reference to formal or linguistic expertise, and simply making appropriate questions to be administered as crowdsourced "jobs". This last approach is being attempted in ongoing work in order to reach a ground truth. We are using CrowdFlower,[25] a crowdsourcing platform where tasks can be defined and assigned to online workforce of millions of people, and that we have already used for other applications of FRED (cf. [34]).

Evolutionary Corpus. While the Balanced Corpus has been used to test the feasibility of a scalable and rigorous evaluation, we also provide an actual gold standard graph bank starting from a corpus of sentences that evolved in the last three years as a

---

[18] http://reverb.cs.washington.edu/.

[19] http://www.w3.org/TR/owl2-overview/.

[20] It can be publicly downloaded from http://wit.istc.cnr.it/motifs/sentencebase.csv.

[21] http://wit.istc.cnr.it/stlab-tools/fred/api.

[22] http://www.w3.org/2004/03/trig/.

[23] http://wit.istc.cnr.it/motifs/graphbankevaluation.csv.

[24] Part-of-speech tagging has been found to be the major source of errors in FRED graphs, probably because it crashes syntactic ordering to the point that the interpretation of the sentence gets corrupted.

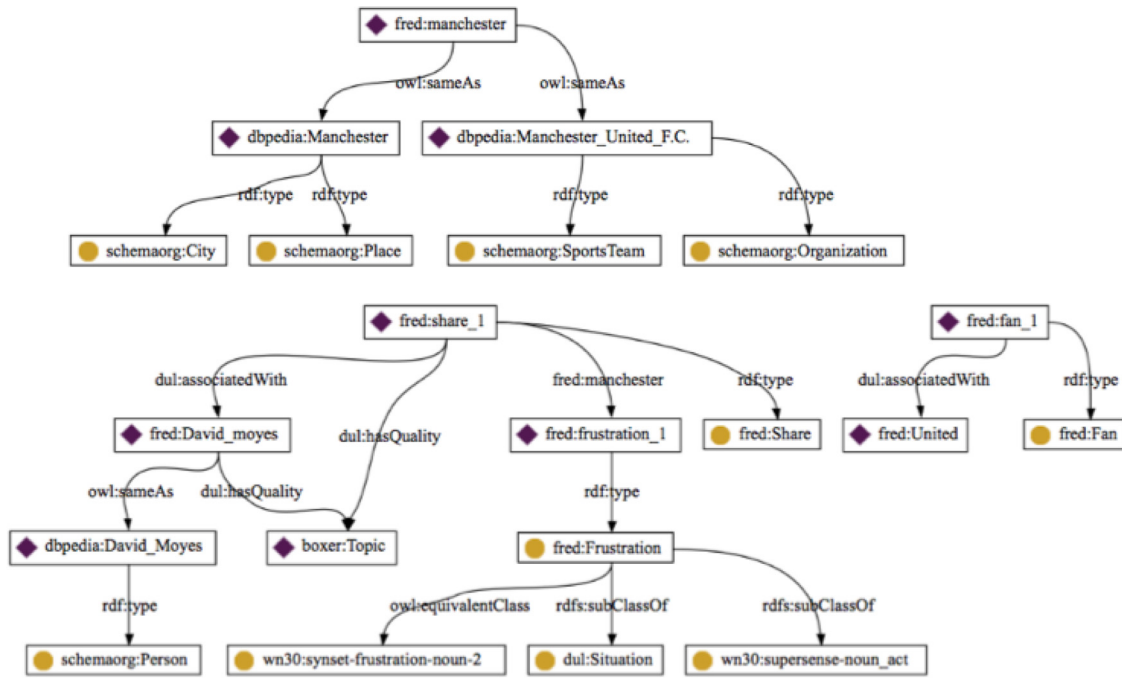[25] http://www.crowdflower.com/.

Fig. 2. The RDF graph produced for the tweet *David Moyes shares Manchester United fans' frustration*. The error of the part-of-speech tagger for the verb *shares* is propagated to FRED, which generates three non-connected components as a representation of the tweet.
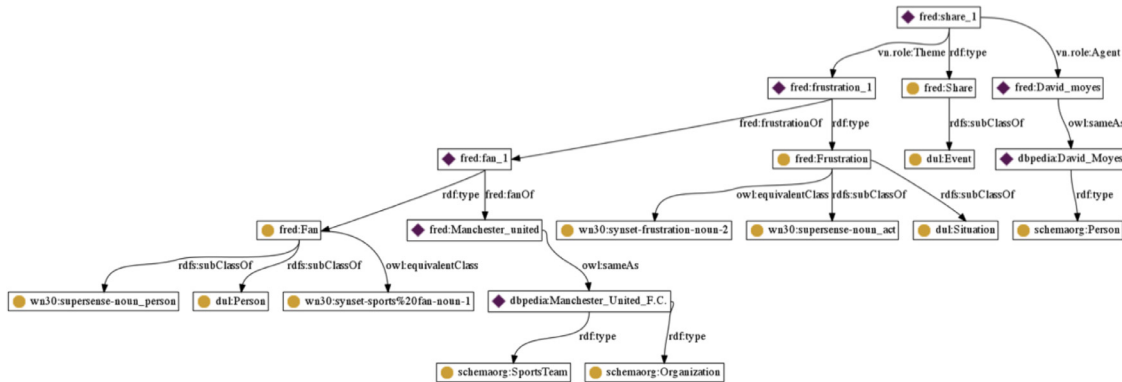

Fig. 3. The gold RDF graph for the same sentence, created using raters' corrections.

benchmark for OKE functionalities. Instead of an a-priori design, this second corpus emerged incrementally based on the collective observations, needs, and debugging challenges from the developers of OKE tools in our laboratory. The graph bank derived from this corpus with the current version of FRED (the sentences are used as unit tests on a daily basis, and have been proof-checked for optimizing FRED's output), and provides a ground truth in a pragmatic and operational way, distinct from the principled approaches described before. This evolutionary graph bank was produced from a corpus of 151 sentences[26] (10 from blogs, 22 from Wikipedia, 18 general sentences, 5 from legal documents, 41 from news, 4 from reviews, 3 from advertisement, 51 test sentences). The average sentence length is 16.7 words per sentence. The same process seen for the first gold standard has been applied. The total number of generated triples is 23,305, 9801 of which are purely semantic triples, 10,650 are annotation triples, 2552 are metadata triples that annotate text segments with parts of speech, 302 are metadata triples annotating named graphs with their identifiers, related sentences, and sentence topics.

To give an idea of the computational time required by a user to build the first graph bank, the process of extracting triples from FRED and building the named graphs for 100 sentences took 156 seconds (this time is the average of 5 different measurements in order to minimise differences in Internet latency for accessing FRED's RESTful service).[27] The two graph banks we have generated (for the 100 sentences and for the 150 sentences), the above aforementioned guidelines, and the examples can be publicly downloaded from http://wit.istc.cnr.it/motifs/graphbank.zip (in TriG syntax).

fredlib has also been used to run SPARQL queries to the generated named graphs, to classify triples according to the motifs presented above, and to include motif occurrences in a CSV file together with their reference named graph, sentence, and motif ready for a user annotation task. The following line shows a

Table 2
Number of extracted triples for each motif for each text type for the first gold standard.

Text type	Role	Equ	Neg	Sub	Fact	Prop	Type	Mod	Identity
News	85	123	3	209	189	204	709	2	38
Scientific reports	78	143	5	246	147	176	877	12	17
Tweets	76	105	13	207	148	204	575	6	22
Wikipedia definition	77	321	0	466	231	281	1911	1	24

Table 3
Number of extracted triples for each motif for each text type for the second gold standard.

Text type	Role	Equ	Neg	Sub	Fact	Prop	Type	Mod	Identity
Conversational	11	49	3	62	31	344	342	2	3
Wikipedia	30	172	1	264	69	1317	2269	3	40
Legal	11	49	1	87	16	439	938	5	3
Reviews	8	33	0	51	42	272	432	0	0
Advertisement	0	22	0	22	10	96	96	1	1
News	72	310	7	480	127	2558	4584	13	70
Test	21	109	9	136	74	959	691	5	26
General	20	70	6	98	49	544	618	3	14

SPARQL query to extract the SUBCLASS motif from the extracted named graphs.[28]

```
SELECT DISTINCT ?graph ?class ?super WHERE {
GRAPH ?graph { ?class rdfs:subClassOf ?super}}
```

The above mentioned motifs cover the totality of semantic triples generated by FRED with the minimal extraction parameters setup (i.e. only VerbNet roles, no tense representation). Table 2 shows the number of extracted triples for the first gold standard for each motif and for each domain whereas Table 3 shows that with respect to the second graph bank (i.e., the gold standard).

7. Use case - a Semantic sentiment analysis application

The motifs that we have identified from FRED's graphs are as elementary as possible, and apparently cover FRED's OKE expressivity. In this section we want to show how we extended them to come up with a novel method of Sentic Computing [35][29] that we have successfully applied [12,13] to the Semantic Sentiment Analysis domain. Sentiment Analysis aims at identifying and extracting the attitude of a subject (an opinion holder) towards a topic or the overall tonality of a document. When it aims at identifying aspects or features of subjects or products the problem is known as aspect-based opinion mining [36]. In the past, it has been typically tackled by means of statistical techniques. With the introduction of the SW, researchers have a new set of resources and background knowledge to improve existing methods for Sentiment Analysis, or coming up with new ones. As a consequence, sentiment lexicons have been extended using concept knowledge [37,38]. Authors in [39] analyse tweets related to Germany elections and calculate the polarity of emerging topics showing how information like relations between topics and the polarity can be used to extend existing knowledge bases to improve concept-level sentiment analysis methods. The reader is referred to [40] for a comprehensive, state of the art review on the work performed within sentiment analysis between 2012 and 2014. In [13] we have developed a model for opinion sentences by first identifying motifs that occur when expressing opinions. Let us consider the example sentence *Anna says the weather will become beautiful* whose FRED graph representation is shown in Fig. 4.

For sentences with such a structure, the ROLE, PROPERTY and MODALITY motifs are those we have identified, and specialized, in order to extract opinion features (denoting new motifs) such as HOLDER, TOPIC or SUBTOPIC, triggering verbs and opinion words. In [13] the reader can find details on how we come up with *Anna*, agent of the TRIGGERING event *say*, being the HOLDER of the opinion, *weather*, agent of the verb *become*, being the TOPIC, and *beautiful*, quality of *weather*, being the OPINION CONTENT.

Formally, we can extensionally define a specialization of a motif as $M_1 \subseteq M$. Note that this is a subset relation, not a subgraph relation (as in other works), i.e., a specialization restricts the set of allowed occurrences for a motif. For example, the HOLDER motif is a specialization of the ROLE motif. This move allows the OKE modeler to map existing subgraphs in terms of specialized motifs that only hold from a certain viewpoint, in this case sentiment analysis.

By using a large set of rules for sentences with different structures, we were able to cover a wide spectrum of expressions containing opinions and sentiments, and identify holders, topics and opinions related to topics. The adoption of existing lexical and sentiment resources (such as SentiWordNet[30] and SenticNet[31]) allowed us to come up with a numeric score for each opinion we retrieved (a continuous number between the -1 and +1 scale where -1 corresponds to extremely negative and +1 to extremely positive). The Semantic Sentiment Analysis tool we have developed on top of FRED has been called Sentilo.[32] Fig. 5 shows the new ontology model we have developed to represent the opinion context of a given sentence using Sentilo. The reader can notice the opinion features (resulting from the definition of rules that also took into account a subset of the motifs defined above) that Sentilo correctly identifies: *weather* as topic, *say* as triggering verb, *Anna* as opinion holder, *beautiful* as opinion quality corresponding to 0.449 and correctly propagated to the referred topic *weather*. By formally specializing the motifs we have presented in this paper, we were able to develop rules for extracting opinion features from English opinion sentences.

8. Conclusions

We have presented a practical model to the creation of graph banks from Semantic Web-oriented Open Knowledge Extraction, which can eventually evolve into gold standards. We have introduced the formal notion of *OKE motif* as a special kind of subgraph pattern useful when RDF and description logics are employed to

---

[28] A composite query for full motif classification is at http://wit.istc.cnr.it/motifs/sparql.txt.

[29] http://sentic.net/computing/.

[30] http://sentiwordnet.isti.cnr.it/.

[31] http://www.sentic.net/.

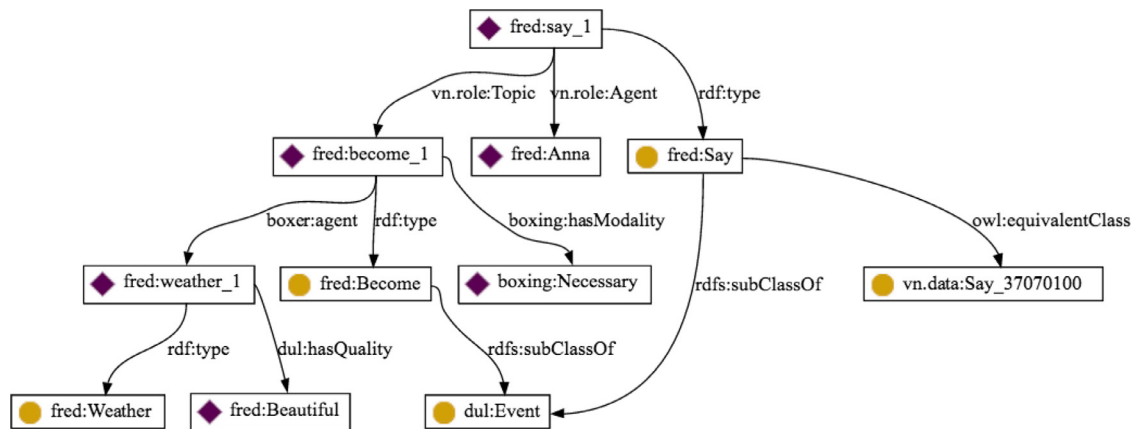[32] http://wit.istc.cnr.it/sentilo-release/sentilo.

Fig. 4. FRED's graph for the example sentence *Anna says the weather will become beautiful.*
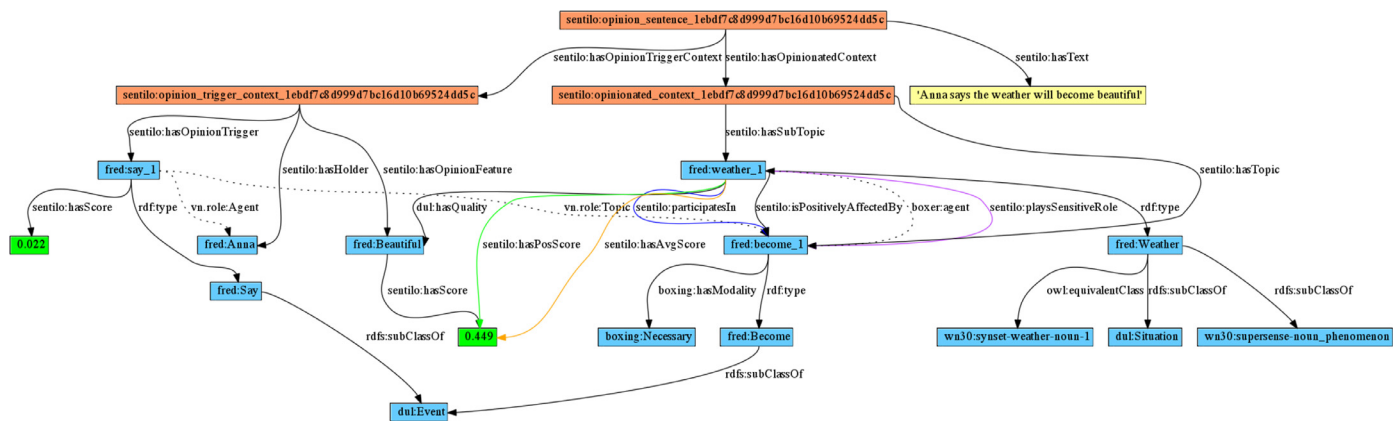

Fig. 5. Sentilo's graph for the example sentence *Anna says the weather will become beautiful.*

formalize NLP results. Some OKE motifs have been proposed based on an extraction conducted on a broad-range OKE tool. OKE motifs can be used to annotate OKE graph banks, to build OKE benchmarks, to evaluate OKE tools, to compare heterogeneous tools, and to perform on-demand OKE graph transformations. They can also be specialized for the sake of application tasks, such as sentiment analysis, question answering, similarity detection, etc. Motifs are implemented as SPARQL query patterns to be submitted to a graph bank. Similarly as with linguistic treebanks, we have used a broad-coverage tool (FRED) to identify motifs, and to generate draft named graphs that can be used to evaluate specific SW tasks after a crowd-sourced correction. We have described that motifs ease the detection of rules and the development of algorithms for sentiment analysis feature detection. In the domain of sentiment analysis, we are currently investigating whether language patterns exist for sarcasm and irony that can be matched to some of the motifs we have introduced in this paper. This is another step toward a new way of thinking about the formal merging of results from OKE, and NLP tools in general. As started in [7], an open-minded attitude towards the class of algorithms and tools to be considered should be accompanied by a serious reflection on which formal representation can work in practice. The reward is quite high, since a common set of motifs would facilitate evaluation, integration, comparison, reusability, all natural tasks in the context of the creation of knowledge bases for the SW and Linked (Open) Data.

Acknowledgments

References

[1] O. Etzioni, M. Banko, M.J. Cafarella, Machine reading, in: Y. Gil, R.J. Mooney (Eds.), Proceedings of the Twenty-first Conference on Artificial Intelligence (AAAI), AAAI Press, Boston, Massachusetts, 2006, pp. 1517–1519.

[2] E. Ferrara, P.D. Meo, G. Fiumara, R. Baumgartner, Web data extraction, applications and techniques: A survey, Knowl. Based Syst. 70 (2014) 301–323, doi:10.1016/j.knosys.2014.07.007.

[3] B.V. Durme, L. Schubert, Open knowledge extraction through compositional language processing, in: Proceedings of the 2008 Conference on Semantics in Text Processing, in: STEP '08, Association for Computational Linguistics, Stroudsburg, PA, USA, 2008, pp. 239–254.

[4] D.R. Radev, E. Hovy, K. McKeown, Introduction to the special issue on summarization, Comput. Linguist. 28 (4) (2002) 399–408.

[5] M. van Erp, P. Vossen, R. Agerri, A.-L. Minard, M. Speranza, R. Urizar, E.L.I. Aldabe, Annotated Data, Committee Specification, VU University of Amsterdam, Fondazione Bruno Kessler, Universidad Del Pais Vasco, 2015.

[6] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffit, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, N. Scheider, Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse, Association for Computational Linguistics, Sofia, Bulgaria, 2013, pp. 178–186.

[7] A. Gangemi, A Comparison of Knowledge Extraction Tools for the Semantic Web, in: Proceedings of the Tenth Extended Semantic Web Conference (ESWC), Montpellier, France, The Semantic Web: Semantics and Big Data, 2013, pp. 351–366.

[8] R. Usbeck, M. Röder, A.-C.N. Ngomo, al., GERBIL – General Entity Annotator Benchmark, in: Proceedings of the Twenty-Fourth International World Wide Web Conference (WWW2015), ACM, Florence, Italy, 2015.

[9] A. Abeillé, Treebanks: Building and using parsed corpora, 20, Springer, Dordrecht, Netherlands, 2003.

[10] S. Pradhan, E. Hovy, M. Marcus, M. Palmer, L. Ramshaw, R. Weischedel, Ontonotes: a unified relational semantic representation, in: Proceedings of the IEEE ICSC 2007, Irvine, CA, USA, 2007.

[11] V. Basile, J. Bos, K. Evang, N. Venhuizen, Developing a large semantically annotated corpus, in: Proceedings of LREC 2012, Istanbul, Turkey, 2012, pp. 3196–3200.

[12] D.R. Recupero, V. Presutti, S. Consoli, A. Gangemi, A. Nuzzolese, Sentilo: Frame-based Sentiment Analysis, Cognit. Comput. 7 (2) (2014) 211–225.

[13] A. Gangemi, V. Presutti, D.R. Recupero, Frame-based detection of opinion holders and topics: a model and a tool, IEEE Comput. Intell. 9 (1) (2014) 20–30.

[14] O. Etzioni, A. Fader, J. Christensen, S. Soderland, Mausam, Open information extraction: The second generation, in: T. Walsh (Ed.), Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI), AAAI Press/IJCAI, Barcelona, Spain, 2011, pp. 3–10.

[15] M. Gardner, P.P. Talukdar, B. Kisiel, T. Mitchell, Improving learning and inference in a large knowledge-base using latent syntactic cues, in: Proc. of the EMNLP 2013, Seattle, USA, 2013.

[16] S. Hellmann, J. Lehmann, S. Auer, Linked-Data Aware URI Schemes for Referencing Text Fragments, EKAW 2012, LNCS 7603, Springer, Galway, Ireland, 2012.

[17] S. Hellmann, J. Lehmann, S. Auer, M. Brümmer, Integrating nlp using linked data, in: H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, J.X. Parreira, N.N. Lora Aroyo, C. Welty, K. Janowicz (Eds.), Proceedings of the Twelth International Semantic Web Conference (ISWC), Part II, Springer, Sydney, Australia, 2013, pp. 98–113.

[18] G. Rizzo, R. Troncy, S. Hellmann, M. Bruemmer, NERD meets NIF: Lifting NLP extraction results to the linked data cloud, in: C. Bizer, T. Heath, T. Berners-Lee, M. Hausenblas (Eds.), Proceedings of the Fifth Workshop on Linked Data on the Web (LDOW) co-located with the International World Wide Web Conference (WWW), CEUR Workshop Proceedings, 937, CEUR-WS.org, Lyon, France, 2012.

[19] V. Presutti, F. Draicchio, A. Gangemi, Knowledge extraction based on Discourse Representation Theory and Linguistic Frames, in: A. ten Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Aquin, A. Nikolov, N. Aussenac-Gilles, N. Hernandez (Eds.), Knowledge Engineering and Knowledge Management - Proceedings of the Eighteenth International Conference (EKAW), Lecture Notes in Computer Science, 7603, Springer, Galway City, Ireland, 2012, pp. 114–129.

[20] P. Cimiano, J. Völker, Text2onto - a framework for ontology learning and data–driven change discovery, in: A. Montoyo, R. Munoz, E. Metais (Eds.), Proceedings of the Tenth International Conference on Applications of Natural Language to Information Systems (NLDB), Springer, Alicante, Spain, 2005, pp. 227–238.

[21] J. Davis, P. Domingos, Deep transfer: A markov logic approach, AI Mag. 32(1) (2011) 51–53.

[22] T. Mohamed, J. Hruschka, R. Estevam, T.M. Mitchell, Discovering relations between noun categories, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, in: EMNLP '11, Association for Computational Linguistics, Stroudsburg, PA, USA, 2011, pp. 1447–1455.

[23] J. Bos, Wide-Coverage Semantic Analysis with Boxer, in: J. Bos, R. Delmonte (Eds.), Proceedings of Conference on Semantics in Text Processing (STEP), College Publications, Venice, Italy, 2008, pp. 277–286.

[24] R. Moot, C. Retoré, The logic of categorial grammars: A deductive account of natural language syntax and semantics, Springer, Berlin Heidelberg, 2012.

[25] I. Augenstein, S. Padó, S. Rudolph, Lodifier: Generating linked data from unstructured text, in: Proceedings of the Ninth International Conference on The Semantic Web: Research and Applications, Heraklion, Crete, 2012, pp. 210–224.

[26] A. Dutta, C. Meilicke, H. Stuckenschmidt, Enriching structured knowledge with open information, in: Proceedings of the Twenty-Fourth International Conference on World Wide Web, in: WWW '15, ACM, New York, NY, USA, 2015, pp. 267–277, doi:10.1145/2736277.2741139.

[27] A.D. Iorio, A.G. Nuzzolese, S. Peroni, Towards the automatic identification of the nature of citations, in: Proceedings of the Third Workshop on Semantic Publishing (SePublica), CEUR, Montpellier, France, 2013, pp. 63–74.

[28] A. Gangemi, F. Draicchio, V. Presutti, A.G. Nuzzolese, D.R. Recupero, A Machine Reader for the Semantic Web, in: E. Blomqvist, T. Groza (Eds.), International Semantic Web Conference (Posters & Demos), CEUR Workshop Proceedings, 1035, CEUR-WS.org, Sydney, Australia, 2013.

[29] S. Peroni, A. Gangemi, F. Vitali, Dealing with markup semantics, in: C. Ghidini, A.N. Ngomo, S.N. Lindstaedt, T. Pellegrini (Eds.), Proceedings of the Seventh International Conference on Semantic Systems (I-SEMANTICS), ACM International Conference Proceeding Series, ACM, Graz, Austria, 2011, pp. 111–118.

[30] D. Davidson, The logical form of action sentences, in: N. Rescher (Ed.), The Logic of Decision and Action, University of Pittsburgh Press, Pittsburgh, 1967.

[31] V. Presutti, S. Consoli, A.G. Nuzzolese, D.R. Recupero, A. Gangemi, I. Bannour, H. Zargayouna, Uncovering the semantics of wikipedia pagelinks, in: K. Janowicz, S. Schlobach, P. Lambrix, E. Hyvönen (Eds.), Knowledge Engineering and Knowledge Management - Proceedings of the Nineteenth International Conference (EKAW), Lecture Notes in Computer Science, 8876, Springer, Linköping, Sweden, 2014, pp. 413–428.

[32] A. Fader, S. Soderland, O. Etzioni, Identifying relations for open information extraction, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, in: EMNLP '11, Association for Computational Linguistics, Stroudsburg, PA, USA, 2011, pp. 1535–1545.

[33] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider, The Description Logic Handbook, Cambridge University Press, New York, USA, 2007.

[34] V. Presutti, S. Consoli, A.G. Nuzzolese, D. Reforgiato Recupero, A. Gangemi, I. Bannour, H. Zargayouna, Uncovering the semantics of wikipedia wikilinks, in: Nineteenth International Conference on Knowledge Engineering and Knowledge Management (EKAW 2014), Linköping, Sweden, 2014.

[35] E. Cambria, A. Hussain, Sentic Computing. A Common-Sense-Based Framework for Concept-Level Sentiment Analysis, Springer, Cham, Switzerland, 2015.

[36] T. Wang, H. fung Leung, R.Y. Lau, Q. Li, H. Min, Product aspect extraction supervised with online domain knowledge, Knowl. Based Syst. 71 (2014) 86–100, doi:10.1016/j.knosys.2014.05.018.

[37] A. Weichselbraun, S. Gindl, A. Scharl, Enriching semantic knowledge bases for opinion mining in big data applications, Knowl. Based Syst. 69 (2014) 78–85, doi:10.1016/j.knosys.2014.04.039.

[38] E. Cambria, D. Olsher, D. Rajagopal, Senticnet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis, in: C.E. Brodley, P. Stone (Eds.), Proceedings of the Twenty-Eight AAAI Conference on Artificial Intelligence, AAAI Press, Québec, Canada, 2014, pp. 1515–1521.

[39] S. Rill, D. Reinel, J. Scheidt, R.V. Zicari, Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis, Knowl. Based Syst. 69 (2014) 24–33, doi:10.1016/j.knosys.2014.05.008.

[40] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, Knowl. Based Syst. 89 (2015) 14–46, doi:10.1016/j.knosys.2015.06.015.