Text2AMR2FRED, a Tool for Transforming Text into RDF/OWL Knowledge Graphs via Abstract Meaning Representation

Aldo Gangemi1, Arianna Graciotti2, Antonello Meloni3, Andrea Nuzzolese4, Valentina Presutti2, Diego Reforgiato Recupero3, Alessandro Russo4 and Rocco Tripodi2

1Department of Philosophy and Communication Studies, University of Bologna, 40126 Bologna, Italy
2Department of Modern Languages, Literatures, and Cultures, University of Bologna, 40126 Bologna, Italy
3Department of Mathematics and Computer Science, University of Cagliari. Via Ospedale 72, 09124 Cagliari, Italy
4Institute of Cognitive Sciences and Technologies, CNR, via San Martino della Battaglia 44, 00185, Rome, Italy

Abstract
This paper presents Text2AMR2FRED, a text-to-Knowledge Graph (KG) pipeline that transforms multilingual natural language text into logically sound KGs. It enables at-scale information retrieval and knowledge extraction. This pipeline overcomes the lax logic and interoperability challenges faced by existing semantic parsers and machine readers. Adhering to Semantic Web standards, it transforms text content into structured knowledge and enriches it by exploiting external knowledge.

Keywords
Knowledge Graphs, Abstract Meaning Representation, Natural Language Processing, Semantic Frames

1. Introduction
Transforming natural language text into logically sound Knowledge Graphs (KGs) supports at-scale information retrieval from collections of texts.

Natural Language Processing (NLP) and Semantic Web (SW) communities dedicated significant effort to text-to-KG pipelines. The NLP community exploited the progress of Machine Learning (ML) and Neural Networks (NN) to improve semantic parsing. Graph-based semantic parsing has gained attention due to the potential of general-purpose representations, such as Abstract Meaning Representation (AMR) [1]. Text-to-AMR transduction based on neural machine translation and sequence-to-sequence (seq2seq) models achieved promising results both in scenarios limited to English and AMR parsing (SPRING, [2], Transition-based AMR parsing [17]) and in multilingual [3] and multi-formalisms scenarios (SGL, [14]). However, neural semantic
parsers struggle with making the extracted knowledge interoperable and exploitable due to its formalisms’ balkanisation [11] and lax logic.

The SW provides means to formally represent the extracted knowledge according to interoperable ontologies, therefore favouring knowledge augmentation with heterogeneous Knowledge Bases (KBs) and alignment with other ontologies. The SW machine reader FRED [6] encodes the extracted information using Semantic Web (SW) standards. The resulting KGs enable the exploration and retrieval of facts extracted from heterogeneous text corpora through structured queries, as well as their augmentation through alignment to other KGs. This alignment supports the disclosure of explicit knowledge that would otherwise remain hidden in texts. However, FRED relies on cumbersome NLP pipelines, hard to maintain and unsuitable to scale in multilingual scenarios.

To overcome such limitations, this paper presents Text2AMR2FRED\(^1\), a revised architecture of FRED’s text-to-KG pipeline. It exploits pre-trained end-to-end text-to-AMR parsers, which mitigate error propagation typical of component-based pipelines. In fact, thanks to AMR’s generalization of lexical and syntactic variations, it allows a more abstract and robust representation of text, without employing ad hoc data augmentation strategies, such as lexical substitution [9, 8]. Furthermore, the SotA AMR parser’s multilingual capabilities enhance the scalability of our application, expanding its reach beyond the original English-only input restriction of FRED’s NLP pipeline.

2. Text2AMR2FRED at work

![Graph resulting from the text2AMR parsing powered by SPRING [2] of the sentence Apple unveils revolutionary watch obtained via Text2AMR2FRED WebApp\(^1\).](image)

Text2AMR2FRED implements a pipeline to produce KGs automatically from unstructured text. These KGs are event-centric, as they rely on PropBank Frames\(^2\) [13]. The process for generating a KG from an input text relies on two modules: (1) the text-to-AMR parsing module, which takes natural language sentences as input and transforms them into AMR graphs. Sentences in English

\(^1\)https://arco.istc.cnr.it/txt-amr-fred/

\(^2\)PropBank Frames are the core lexicon of the PropBank paradigm and consist of predicate-argument structures named “rolesets”. A complete list of PropBank frames can be found at http://propbank.github.io/v3.4.0/frames/
are parsed by SPRING\(^3\) [2]. Sentences in other languages are parsed by USeA\(^4\) [12], which takes input in 100 languages. (2) the AMR-to-FRED translation, which extends the AMR2FRED\(^5\) tool [10] to transform AMR graphs into OWL-compliant RDF KGs, following FRED’s theoretical model [6]. The integration of the two modules is eased and streamlined by the APIs provided by both tools. Specifically, the AMR graph produced by the text-to-AMR parsers from the input text can be directly used as input for the AMR2FRED tool to get a corresponding KG. This makes it possible for tools such as the Machine Reading suite\(^6\) to query both components through the Text-to-AMR-to-FRED APIs\(^7\) and generate RDF named graphs from input text sentences or paragraphs in batches. Text2AMR2FRED is also released to the public via a user-friendly web app\(^1\).

The AMR-to-FRED translation facilitates KGs enrichment, which can be achieved by employing Framester\(^5\). Thanks to Framester, additional relevant knowledge missing in the text (e.g., common sense knowledge) can be recovered from other KBs such as WordNet\(^8\), DBPedia\(^9\), DOLCE-Zero\(^10\). For example, the output KGs are enriched through Word Sense Disambiguation (WSD) based on the RDF version of WordNet, included in Framester. The WSD process is applied to AMR elements (usually nouns and adjectives) that miss links to lexical resources. Figure 1 shows the AMR graph corresponding to the sentence “Apple unveils revolutionary watch”. The reader may notice that the text-to-AMR parser associates predicates in AMR graphs with PropBank word senses and Named Entities with their corresponding entities in Wikipedia. The node z\(3\) / watch instead is missing a link to lexical resources. Therefore, we disambiguate it against Framester. The WSD process consists of submitting the original sentence to EWISER\(^11\), a WSD system well-suited for multilingual scenarios due to its SotA performance in both all-words English WSD and multilingual WSD tasks. As Figure 2 shows, we associate the result of WSD (WordNet’s synsets) with the AMR nodes missing links to any external source and whose label

---

\(^3\)http://nlp.uniroma1.it/spring/
\(^4\)https://github.com/SapienzaNLP/usea
\(^5\)https://github.com/infovillasimius/amr2Fred/tree/master
\(^6\)https://github.com/anuzzolese/machine-reading
\(^7\)http://framester.istc.cnr.it/txt-amr-fred/api/docs
\(^8\)https://wordnet.princeton.edu
\(^9\)https://www.dbpedia.org
\(^10\)http://www.ontologydesignpatterns.org/ont/d0.owl
\(^11\)https://github.com/SapienzaNLP/ewiser
corresponds to the lemma of the input sentence. This association is implemented through the `owl:equivalentClass` property between the identified node and the selected WordNet’s synset in Framester. For the example above, we use EWI$\text{SER}$ and keep the information for the lemma “watch”. For the same entities (those not linked with external information sources), we further exploit Framester to generate alignments to two top-level ontologies: WordNet “supersenses” (through the `rdfs:subClassOf` property) and a subset of DOLCE+DnS Ultra Lite (DUL) classes.

3. Conclusions and Future Work

Text2AMR2FRED is a tool that mitigates the issues of existing NLP semantic parsers and machine readers, adhering to Semantic Web standards to ensure interoperable knowledge extraction. It enhances the informativeness of KGS by aligning them with domain-specific ontologies, enabling interrogation through structured queries. This approach uncovers implicit knowledge from text, enabling the output of KGS with external KBs.

Future work will focus on creating resources for the evaluation of the tool. The AMR parsers employed in our tool can be leveraged to perform AMR-to-text tasks and compare the original textual excerpts to the automatically generated ones via a back-translation [16] approach. This allows the calculation of similarity metrics, such as BLEURT [15] or others, between the original and generated texts. Under the hypothesis that generated sentences with (relatively) high similarity scores correspond to high-quality AMR graphs, automatic filters can be designed and applied to prevent lower-quality AMR graphs’ transformation into RDF/OWL KGS. Our evaluation method will be completed by the analysis of Motifs, basic logical patterns employed in SW, defined in [7], in the output KGS. The Motifs-based validation will permit a cross-tools knowledge extraction tasks comparative evaluation, following the method outlined in [4].

References


