Text2AMR2FRED, a tool for transforming text into RDF/OWL Knowledge Graphs via Abstract Meaning Representation[*]

Aldo Gangemi[1], Arianna Graciotti[2,*], Antonello Meloni[3], Andrea Nuzzolese[4], Valentina Presutti[2], Diego Reforgiato Recupero[3], Alessandro Russo[4] and Rocco Tripodi[2]

[1]*Department of Philosophy and Communication Studies, University of Bologna, 40126 Bologna, Italy*

[2]*Department of Modern Languages, Literatures, and Cultures, University of Bologna, 40126 Bologna, Italy*

[3]*Department of Mathematics and Computer Science, University of Cagliari. Via Ospedale 72, 09124 Cagliari, Italy*

[4]*Institute 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.
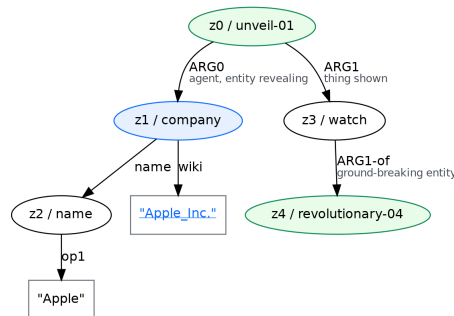
2. Text2AMR2FRED at work

Figure 1: Graph resulting from the text2AMR parsing powered by SPRING [2] of the sentence *Apple unveils revolutionary watch* obtained via Text2AMR2FRED WebApp[1].

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/
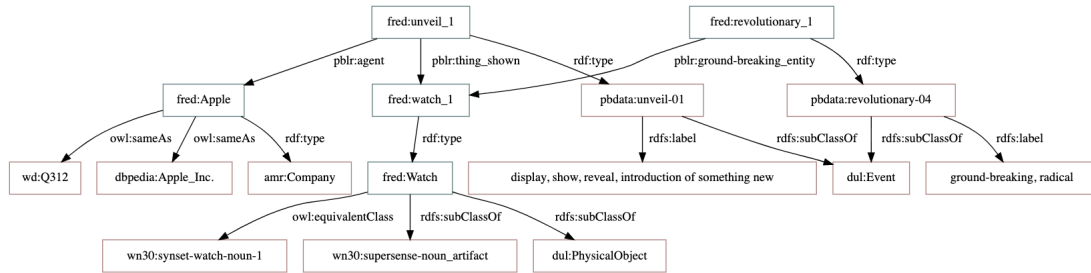
Figure 2: Graph resulting from the AMR2FRED translation powered by AMR2FRED of the sentence *Apple unveils revolutionary watch* obtained via Text2AMR2FRED WebApp[1].

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 z3 / 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

---

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 EWISER 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

[1] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, and N. Schneider. Abstract Meaning Representation for Sembanking. In *Proc. of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria, August 2013. ACL.

[2] M. Bevilacqua, R. Blloshmi, and R. Navigli. One SPRING to Rule Them Both: Symmetric AMR Semantic Parsing and Generation without a Complex Pipeline. *Proc. of the AAAI Conference on Artificial Intelligence*, 35(14):12564–12573, May 2021.

[3] R. Blloshmi, R. Tripodi, and R. Navigli. XL-AMR: Enabling Cross-Lingual AMR Parsing with Transfer Learning Techniques. In *Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2487–2500, Online, November 2020. ACL.

[4] A. Gangemi. A Comparison of Knowledge Extraction Tools for the Semantic Web. In *The Semantic Web: Semantics and Big Data*, pages 351–366, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[5] A. Gangemi, M. Alam, L. Asprino, V. Presutti, and D. R. Recupero. Framester: A Wide Coverage Linguistic Linked Data Hub. In *EKAW 2016*, pages 239–254, Bologna, Italy, 2016. Springer International Publishing.

[6] A. Gangemi, V. Presutti, D. R. Recupero, A. G. Nuzzolese, F. Draicchio, and M. Mongiovì. Semantic Web Machine Reading with FRED. *Semantic Web*, 8(6):873–893, 2017.

[7] A. Gangemi, D. Reforgiato Recupero, M. Mongiovì, A. Nuzzolese, and V. Presutti. Identifying Motifs for Evaluating Open Knowledge Extraction on the Web. *Knowledge-Based Systems*, 108:33–41, May 2016.

[8] C. Lacerra, T. Pasini, R. Tripodi, and R. Navigli. Alasca: an automated approach for large-scale lexical substitution. In *Proc. of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 3836–3842. ijcai.org, 2021.

[9] C. Lacerra, R. Tripodi, and R. Navigli. Genesis: A Generative Approach to Substitutes in Context. In *Proc. of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10810–10823, Online and Punta Cana, Dominican Republic, November 2021. ACL.

[10] A. Meloni, D. Reforgiato Recupero, and A. Gangemi. AMR2FRED, A Tool for Translating Abstract Meaning Representation to Motif-Based Linguistic Knowledge Graphs. In *The Semantic Web: ESWC 2017 Satellite Events*, pages 43–47, Portorož, Slovenia, 2017. Springer International Publishing.

[11] S. Oepen, O. Abend, L. Abzianidze, J. Bos, J. Hajic, D. Hershcovich, B. Li, T. O'Gorman, N. Xue, and D. Zeman. MRP 2020: The Second Shared Task on Cross-Framework and Cross-Lingual Meaning Representation Parsing. In *Proc. of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*, pages 1–22, Online, November 2020. ACL.

[12] R. Orlando, S. Conia, S. Faralli, and R. Navigli. Universal Semantic Annotator: the First Unified API for WSD, SRL and Semantic Parsing. In *Proc. of LREC 2022*, pages 2634–2641, Marseille, France, June 2022. European Language Resources Association.

[13] S. Pradhan, J. Bonn, S. Myers, K. Conger, T. O'gorman, J. Gung, K. Wright-bettner, and M. Palmer. PropBank Comes of Age—Larger, Smarter, and more Diverse. In *Proc. of SEM 2022*, pages 278–288, Seattle, Washington, 2022. ACL.

[14] L. Procopio, R. Tripodi, and R. Navigli. SGL: Speaking the Graph Languages of Semantic Parsing via Multilingual Translation. In *Proc. of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 325–337, Online, June 2021. ACL.

[15] T. Sellam, D. Das, and A. Parikh. BLEURT: Learning robust metrics for text generation. In *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7881–7892, Online, July 2020. ACL.

[16] R. Sennrich, B. Haddow, and A. Birch. Improving Neural Machine Translation Models with Monolingual Data. In *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany, August 2016. ACL.

[17] J. Zhou, T. Naseem, R. Fernandez Astudillo, Y.-S. Lee, R. Florian, and S. Roukos. Structure-aware Fine-tuning of Sequence-to-sequence Transformers for Transition-based AMR Parsing. In *Proc. of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6279–6290, Online and Punta Cana, Dominican Republic, November 2021. ACL.