

## MUSICBO, A MUSICAL HERITAGE KNOWLEDGE GRAPH

A. Gangemi, A. Graciotti, E. Marzi, **A. Meloni**,  
A. Nuzzolese, V. Presutti, D. Reforgiato Recupero,  
A. Russo, R. Tripodi  
Department of Mathematics and Computer Science,  
University of Cagliari. Via Ospedale 72, 09124 Cagliari, Italy  
`diego.reforgiato@unica.it`

The MusicBO Knowledge Graph (KG), documenting Bologna’s pivotal role in the European Musical Heritage, is created using an automated text-to-Knowledge Graph pipeline on a multilingual and diachronic corpus. The corpus contains 137 documents in English, French, Italian, and Spanish. The documents’ publication dates span from 1700 to the current era. This corpus includes an extensive variety of textual genres, such as historical dissertations, critical analyses, correspondences, and business-oriented documents. The KG has been created automatically by leveraging text2AMR2FRED<sup>1</sup> [2], a revised architecture of FRED’s text-to-KG pipeline [3]. For creating MusicBO KG, we applied the mentioned pipeline to a subset of the mentioned corpus, made of 47 documents in English and 51 documents in Italian. The documents were originally in .pdf, images, or .docx formats. We extrapolated plain text from them through ad hoc Optical Character Recognition (OCR) technologies. We then performed co-reference resolution, rule-based minimal post-OCR corrections, and sentence splitting on the extrapolated plain texts. We submitted the pre-processed sentences to State-of-the-Art (SotA) neural models for text2AMR parsing. For sentences in English, we used SPRING<sup>2</sup> [1]. For sentences in Italian, we used USEA<sup>3</sup> [5]. AMR graphs, grounded in PropBank Frames<sup>4</sup>, serve as an intermediate event-centric representation, well-suited for retrieving ‘who-did-what-to-whom’ in a text. The produced AMR graphs are filtered based on automatic metrics. The remaining AMR graphs undergo AMR-to-FRED translation, which exploits the AMR2FRED tool<sup>5</sup> [4] to transform AMR graphs into OWL-compliant RDF KGs, following FRED’s theoretical model<sup>6</sup>. The resulting KG provides scholars, particularly musicologists, with access to structured information extracted from historical and multilingual sources, poised to potentially enhance their research and understanding of the musical heritage domain. In particular, statistics about triples and data stories can be extrapolated from the KG. Some examples are provided. The proposed KG is freely accessible at <https://polifonia.disi.unibo.it/musicbopilot/query>.

---

<sup>1</sup><https://arco.istc.cnr.it/txt-amr-fred/>

<sup>2</sup><https://nlp.uniroma1.it/spring/>

<sup>3</sup><https://github.com/SapienzaNLP/usea>

<sup>4</sup><http://propank.github.io/v3.4.0/frames/>

<sup>5</sup><http://framester.istc.cnr.it/amr-2-fred>

<sup>6</sup><http://wit.istc.cnr.it/stlab-tools/fred/demo/>

## References

- [1] Michele Bevilacqua et al. 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, 2021.
- [2] Aldo Gangemi, Arianna Graciotti, Antonello Meloni, Andrea Nuzzolese, Valentina Presutti, Diego Reforgiato Recupero, Alessandro Russo, and Rocco Tripodi. Text2AMR2FRED, a tool for transforming text into RDF/OWL Knowledge Graphs via Abstract Meaning Representation. In *22nd International Semantic Web Conference*, Athens, Greece, November 2023. CEUR Workshop Proceedings.
- [3] Aldo Gangemi, Valentina Presutti, Diego Reforgiato Recupero, Andrea Giovanni Nuzzolese, Francesco Draicchio, and Misael Mongiovì. Semantic Web Machine Reading with FRED. *Semantic Web*, 8(6):873–893, 2017.
- [4] Antonello Meloni, Diego Reforgiato Recupero, and Aldo 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.
- [5] Riccardo Orlando, Simone Conia, Stefano Faralli, and Roberto Navigli. Universal Semantic Annotator: the first unified API for WSD, SRL and Semantic Parsing. In *Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022)*, Marseille, France, June 2022. European Language Resources Association.