Tell Me the Truth. Validating the Semantic Alignment between the Annotation User Interface and the Knowledge Base

Bonora, Paolo

paolo.bonora@unibo.it
Department of Classical Philology and Italian Studies,
University of Bologna, Italy

Dello Buono, Martina

martina.dellobuono2@unibo.it
Department of Classical Philology and Italian Studies,
University of Bologna, Italy

Giovannetti, Francesca

francesc.giovannett6@unibo.it
Department of Classical Philology and Italian Studies,
University of Bologna, Italy

Tomasi, Francesca

francesca.tomasi@unibo.it
Department of Classical Philology and Italian Studies,
University of Bologna, Italy

Problem Description and Research Question

Our proposal focuses on user annotation of existing knowledge bases (KB). When proposing new semantic relationships through an interface (UI), users expect that the annotation will extend the KB as well as that the annotation semantics will comply with the conceptual model underlying the KB. The annotation application should act as an intermediary between the user and the formal representation of the knowledge that is being inserted becoming, together with the UI, a critical component of the knowledge management chain. However, how can we ensure that the knowledge being added to the knowledge base is correctly aligned with the semantics of the user annotation? Leveraging two use cases from archival and textual scholarship, we outline a methodology and a testing approach for assessing the semantic neutrality of UIs in annotation applications.

Use Cases

The first use case is the archive of the Bolognese writer Giuseppe Raimondi (1895-1985), which comprises the author's library (ca. 6,200 literature and art publications, 265 journal issues, ca. 50 exemplars of his published works) and papers (over 10,000 items including letters and postcards, drawings, photographs; 460 manuscript notebooks; 480 manuscript and typescript articles). The original finding aid for Giuseppe Raimondi's archive, dating back to the 1990s, was converted to linked open data to facilitate research. Over time, several scholars have investigated the collection, revealing significant connections that help contextualize Raimondi's work. An example is Rossi and Wenzlawski's reconstruction of some of the links between the library and the papers showing Valéry's reception in Raimondi's literary production (Rossi / Wenzlawski 2020).

The second is the *Repertorio della Poesia Italiana in Musica, 1500-1700* (RePIM) [1], the incipit index of Italian poetry in music composed between the 16 th and the 17 th centuries. Through the systematic examination of poetry books, RePIM aims to discover the identity of the authors of poetic texts that have been set to music. RePIM contains approximately 53.000 texts in music, with the authorship of 20.700 of them successfully identified (Pompilio / Bonora 2021). The scholarly community's contribution is required for the number of identified authors to grow. The web application will be expanded to allow users to annotate newly identified links between texts [2], along with scholarly notes and the provenance of the annotation.

UI Design

The methodology followed for designing the UI is based on Garret's user-centred design model (Garrett 2012) and Tomasi's project workflow best practices (Tomasi 2022). The result is a UI set on panels. The first panel allows users to set the subject of a triple; the central panel allows users to establish semantic relationships by specifying the predicate and the object. Relationships can be viewed in three modes: the "Relations" panel displays data as a graph; the "Interpretation" panel shows the RDF source and the assertions expressed in natural language. The following figures depict the UIs for UC1 (Fig. 1) and UC2 (Fig.2).
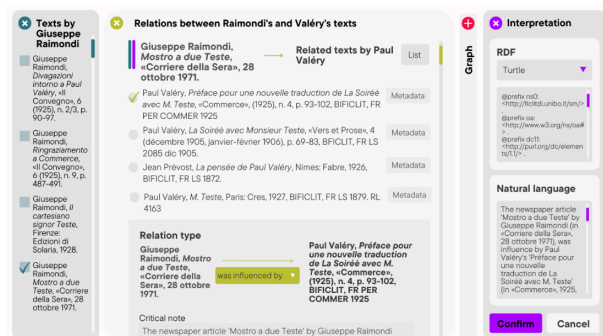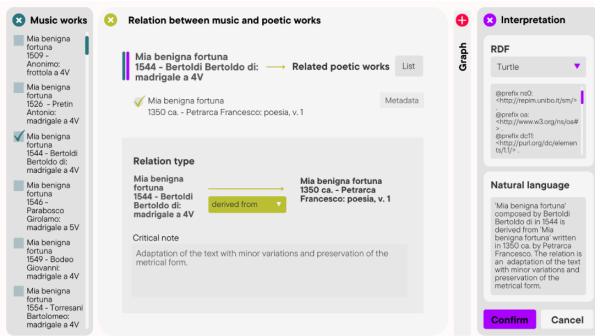

Fig. 1 UI of UC1.

Fig. 2 UI of UC2.

Methodology and Validation

Ensuring the semantic alignment between the annotation UI and the produced knowledge graph is a twofold task. The user should validate the triples representing the annotation and the system should guarantee that any further representation of the newly added triples preserves the annotation's original meaning. To perform the initial validation, we will use RDF-to-Text techniques to provide the user with the natural language translation of the triples (see Bernstein et al. 2005; Affolter et al. 2019; Ngomo et al. 2019) along with the RDF representation. Once the user confirms the assertions, these will be incorporated into the KB.

Second, we will need to test the bi-univocity of the UI. To do so, we will assess a selected set of triples resulting from a read-and-write (RW) cycle. The test case (TC) reads the triples from the KB prints the triples on the UI and, finally, saves the triples once again without any modifications. The TC is passed if the resulting knowledge graph matches the original one. To implement the testbed and assess the consistency of the data presentation layers and knowledge base, we will create a prototype, which will be available at https://tell-me-the-truth.github.io/web-app/.

Conclusion

Our study aims to define a methodology to assess UI's semantic transparency in annotation applications. Any layer interposed between the KB and the user acts inevitably as a "semantic lens" (Peroni et al. 2014) that may introduce a semantic distortion. Using the selected UCs, the proposed approach attempts to investigate whether the phenomenon is detectable and, if so, how to measure it using user feedback.

Notes

1. RePIM is also published as RDF dataset at https://doi.org/10.5281/ZENODO.5692109.
2. The most frequent relationships between texts are: "derived from", "related work", "paraphrase of", "revision of" and "citation".

Bibliography

Affolter, Katrin / **Stockinger, Kurt** / **Bernstein, Abraham.** (2019) 'A Comparative Survey of Recent Natural Language Interfaces for Databases', *The VLDB Journal*, 28(5), pp. 793–819. DOI: 10.1007/s00778-019-00567-8.

Bernstein, Abraham / **Kaufmann, Esther** / **Göhring, Anne.** (2005) 'Querying Ontologies: A Controlled English Interface for End-Users', in Y. Gil et al. (eds) *The Semantic Web – ISWC 2005*. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Computer Science), pp. 112–126. DOI: 10.1007/11574620_11.

Garrett, Jesse J. (2011) *The Elements of User Experience: User-Centered Design for the Web and Beyond*. 2nd ed. Berkeley, Calif: New Riders.

Ngomo, Axel-Cyrille N. / **Moussallem, Diego** / **Bühmann, Lorenz.** (2019) 'A Holistic Natural Language Generation Framework for the Semantic Web'. DOI: 10.48550/AR-XIV.1911.01248.

Peroni, Silvio / **Tomasi, Francesca** / **Vitali, Fabio** / **Zingoni, Jacopo.** (2014) 'Semantic Lenses as Exploration Method for Scholarly Articles', in T. Catarci, N. Ferro, and A. Poggi (eds) *Bridging Between Cultural Heritage Institutions*. Berlin, Heidelberg: Springer Berlin Heidelberg (Communications in Computer and Information Science), pp. 118–129. DOI: 10.1007/978-3-642-54347-0_13.

Pompilio, Angelo / **Bonora, Paolo.** (2021) 'RePIM in LOD: Semantic Technologies to Preserve Knowledge about Italian Secular Music and Lyric Poetry from the 16th -17th Centuries', in *AIUCD 2022 - Proceedings*, pp. 193–195. DOI: 10.6092/unibo/amsacta/6848.

Rossi, Federica / **Wenzlawski, Alina.** (2020) 'Nello scrittoio di Giuseppe Raimondi: carte e libri di un letterato bolognese su Paul Valéry', in G. Di Domenico and F. Sabba (eds) *Il privilegio della parola scritta: gestione, conservazione e valorizzazione di carte e libri di persona*. Roma: Associazione italiana biblioteche, pp. 177–194.

Tomasi, Francesca. (2022) *Organizzare la conoscenza: digital humanities e web semantico: un percorso tra archivi, biblioteche e musei*. Milano: Editrice Bibliografica (Biblioteconomia e scienza dell'informazione, 39).