

Towards a Semantic Network of Dante's Works and Their Contextual Knowledge

Tavoni, Mirko

Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa

mirko.tavoni@gmail.com

Andriani, Paola

Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa

paola.andriani@gmail.com

Bartalesi, Valentina

ISTI-CNR

valentina.bartalesi@isti.cnr.it

Locuratolo, Elvira

ISTI-CNR

elvira.locuratolo@isti.cnr.it

Meghini, Carlo

ISTI-CNR

carlo.meghini@isti.cnr.it

Versienti, Loredana

ISTI-CNR

loredana.versienti@isti.cnr.it

Short Paper

URL: dhwriter.org/paper/1394113965.344.417.25.html

Towards a Semantic Network of Dante's Works and Their Contextual Knowledge

- Tavoni, Mirko
Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa
mirko.tavoni@gmail.com
- Andriani, Paola
Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa
paola.andriani@gmail.com
- Bartalesi, Valentina
ISTI-CNR
valentina.bartalesi@isti.cnr.it
- Locuratolo, Elvira
ISTI-CNR
elvira.locuratolo@isti.cnr.it
- Meghini, Carlo
ISTI-CNR
carlo.meghini@isti.cnr.it
- Versienti, Loredana
ISTI-CNR
loredana.versienti@isti.cnr.it

Summary

We present the “Towards a Digital Dante Encyclopedia” project, a three years Italian National Research Project, that aims at building a prototypical digital library endowed with services supporting scholars in creating, evolving and consulting a digital encyclopedia of Dante Alighieri and of his works.

1. Introduction

In the field of digital humanities, scholars are increasingly producing digital editions of texts and manuscripts. The representation of knowledge included in literary texts is a complex issue, requiring rich vocabularies, also called ontologies, for representing the many different aspects that are investigated by scholars. In literature, there are many ontologies that focus on different aspects of textual information but one single ontology representing all these aspects does not exist.

The “Towards a Digital Dante Encyclopedia” project is a three years Italian National Research Project, started in 2012, that aims at building a prototypical digital library endowed with services supporting scholars in creating, evolving and consulting a digital encyclopedia of Dante Alighieri and of his works. The digital library is based on a semantic representation of Dante's works and of the knowledge embedded in them in RDF language [1], a language recommended by the Web Consortium for the representation of knowledge. In RDF, every piece of knowledge is represented as a triple (subject predicate object), and a set of triples form an RDF graph, generally called *semantic network*, in order to highlight the formal linguistic nature of the representation. The services being developed address several tasks carried out by the scholars building the encyclopedia, starting with the visualization of references to primary sources (*i.e.*, other authors' works which Dante referred to his own works), their types and their distribution both in time and in the works of Dante. The overall goal is to shed light into the cultural context in which Dante wrote his works and into the development of Dante's reference library over time.

This part of the project is divided in several phases. The first phase regards the creation of an ontology for the knowledge embedded in scholarly commentaries to *Convivio* [2], the philosophical treatise which we choose as initial case study. In the second phase, the ontological model is generalized to represent the knowledge embedded in the scholarly commentaries to other Dante's works. In the third phase, Dante's works along with their attached commentaries are inserted into the digital library, as part of the semantic network being built. In the fourth phase, the primary sources referenced by Dante in his works, as reported by the commentaries, are inserted into the digital library, following the same semantic approach. In the last phase, services are developed, as web applications that allow scholars to browse the semantic network of Dante's work, of primary sources, or of references linking the former to the latter. The references will be visualized in an intuitive way through tables and charts, highlighting their distribution in Dante's work and over time.

We present the structure of the semantic network, as it currently stands and indicate how it will be further developed. Furthermore, we highlight the benefits brought by the visualization service of primary sources to scholars.

2. ONTOLOGY FOR THE REPRESENTATION OF *CONVIVIO*

In order to detect the primary sources used by Dante to write his *Convivio*, we relied on the most recent and updated commentary to the text, that of Gianfranco Fioravanti (Mondadori, in the press), and created an ontology for representing the relevant knowledge carried by this commentary. In particular, our ontology represents:

- the passage of Dante's text (e.g., "Sì come dice lo Filosofo nel principio della Prima Filosofia") to which a quotation from a source refers;
- the correspondent book, chapter and paragraph of Dante's text
- the author of the work referenced in the commentary (e.g., Aristotle);
- the title of the work referenced in the commentary (e.g., *Metaphysics*);
- the thematic area of the work referenced in the commentary (e.g., Aristotelianism).

In order to create an ontology for the semantic representation of the above information, we investigated several existing ontologies (e.g. CIDOC-CRM [3], FRBR [4], FaBiO [5], SKOS [6]), and we chose the classes and properties that we considered the most appropriate to represent the above information. Furthermore, we added our own classes and properties for the representation of the categories of knowledge that were not addressed by the existing ontologies. Then, we transformed the initial commentary into an RDF graph structured according to the ontology [7].

On the basis of our ontology, we are approaching the remaining phases of the "Towards a Digital Dante Encyclopedia" project. To such aim, we are using the ontology developed so far in order to represent other works of Dante (e.g. *De Vulgari Eloquentia*, *Monarchia*) as well as the knowledge carried by commentaries to them. At the same time, we are collecting the primary sources of Dante's work in a digital format, for insertion into the semantic network underlying the digital library. Our diachronic analysis, in fact, aims at representing the evolution of Dante's knowledge about primary sources.

3. THE MODEL POPULATION

In order to enrich our RDF graph, as we have done for *Convivio*, we are collecting information for other Dante's works. In particular we are focusing on (i) the text of the work along with the attached commentaries; (ii) the primary sources referenced in the notes. We are currently storing the RDF triples generated according to our ontology both for the notes and the primary sources. We are relying on the Virtuoso [8] technology for storing and accessing large RDF graphs.

It is important to note that the works of Dante, as well as most of their primary sources, exist in some digital format. However, to the best of our knowledge, there is no semantic representation that integrates this information into a unique body of knowledge, expressed through a formal ontology. We do not expect the knowledge base that we build to give a coherent view of Dante's works. The knowledge in it may, and in general will be incoherent and incomplete, and our ontology is flexible enough to allow both.

The creation of the semantic network is a very time-consuming and knowledge-intensive process. It requires researching the most appropriate ontologies for representing all aspects, and in several cases it requires developing a new ontology to fill the gaps of existing ones. Once the ontology is created, the works of Dante, the primary sources, and the knowledge embedded in them will have to be expressed in this ontology, and this is also a technically demanding task. But the benefits are enormous. The digital representation of the knowledge can support scholars in several conceptually simple but time-consuming tasks, allowing them to focus on the more intellectual aspects of their work. The semantic network will be usable for a wide variety of purposes, which go well beyond the specific services built by our project. It will constitute a backbone that can be enriched with other knowledge about Dante and the historical events, people, artistic movements, etc. that have come across Dante and as such contribute to form the context in which Dante's life and art took place. In this sense, creating the semantic network is the most important achievement of our project. Our project will build only one part of this

network, but will also lay the bases for the extensions and enrichments that will complete what we have started.

4. WHY THIS UNIFIED ARCHIVE WILL BE IMPORTANT TO STUDY THE CULTURE OF DANTE

The importance of the archive and tools described above in order to study how the culture of Dante developed in time is obvious. The fact of gathering the current information on the primary sources used by Dante in his works, and the fact of having this information available in digital format, will improve and make more efficient the research of primary sources by the scholars. Having all the information dispersed on paper books, in fact, makes impossible a systematic overview of the culture of Dante and a well-ordered perception of how it was gradually set up in time. On the contrary, the automatic visualization of data about primary resources, according to different parameters (in chronological order, or by type of source, or by author, by work, etc.), will allow to explore the dynamics of the multi-faceted culture of Dante in relation to the diverse and often conflicting stages of his biography and to study the evolution in time of Dante's cultural background.

References

1. Manola F., Miller, E. (2004). RDF Primer. W3C Recommendation 10 February 2004. Available at: <http://www.w3.org/TR/rdf-primer/>
2. ALIGHIERI, D. 1987. *Convivio*, a c. di P. Cudini, Milano, Garzanti
3. Doerr, M. 2003. The CIDOC CRM - an ontological approach to semantic interoperability of metadata. *AI Magazine*, vol. 24(3), 75-92
4. Tillett, B. 2004. What is FRBR? A conceptual model for the bibliographic universe. *Library of Congress Cataloging Distribution Service*, vol. 25(5), 1-8.
5. Peroni, S., and Shotton, D. 2012. FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. *J. Web Semantics: Science, Services and Agents on the World Wide Web*.
6. Alistair, M., Matthews, B., Wilson M., and Brickley, D. 2005. SKOS Core: Simple knowledge organisation for the Web. In *Proceedings of Dublin Core Conference*.
7. Bartalesi, V., Meghini, C., Locuratolo, E., Versienti, L. 2013. A preliminary study on the semantic representation of the notes to Dante Alighieri's *Convivio*. In *Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities (DH-CASE '13)*. ACM, New York, NY, USA.