

Knowledge Acquisition on Dante Alighieri's Works

K

Elvira Immacolata Locuratolo

ISTI-CNR, Italy

Valentina Bartalesi Lenzi

ISTI-CNR, Italy

INTRODUCTION

Towards a Digital Dante Encyclopedia is a three-year Italian National Project that aims to build a digital library endowed with services to support scholars in creating, evolving and consulting a digital encyclopedia of Dante Alighieri and of his works (Alighieri, 2011, 2014). Studies concerning the semantic representation of the notes to Dante Alighieri's works can be found in Bartalesi, Locuratolo, Meghini and Versienti (2013). A semantic network of Dante's works and their contextual knowledge is described in Andriani, Bartalesi, Locuratolo, Versienti, Meghini, Tavoni and Versienti (2013).

This article is a research paper related with one of the project objectives, i.e. the identification and the implementation of visualization diagrams to support the of knowledge acquisition on Dante's works. The collaboration between humanists and computer scientists was essential for the purposes of this activity. Moreover, knowledge from humanists was advantageously exploited to individuate the right diagrams; knowledge from computer scientists was advantageously exploited to solve methodological problems of research, implementation and graphical representation.

The approach followed for the purposes of this activity was based on three phases, respectively called identification, implementation and graphical representation. With regard to the identification

phase, two high level structures were introduced: the former, called Dante work's sources, is concerned with the cited primary sources; the latter, called Dante work's structure, is concerned with the structure of Dante's works. For each of them, significant types of histograms were identified. The structure Dante work's sources is applicable to both all the Dante's works/a selected Dante's work. In the former case, the histograms that can be derived from the structure are sufficient to fulfill the project purposes; in the latter case, the histograms derived from both the structures, Dante work's sources and Dante work's structure, need to be related.

The implementation phase is concerned with a table summarizing the structures and the knowledge from humanists. The significant types of citation histograms are represented in a model that does not reduce the completeness of the graphical information. The model is formed by a table containing four columns identified by the 4-tupla (Domain, Range, Graphical Type, Refinement).

With respect to the graphical representation, the histogram is a function depending from the parameters Domain and Range. The domain is represented on the horizontal axis of a cartesian diagram, whereas the citation distributions are represented on the vertical axis. The possibility to refine the acquired knowledge, as well as the possibility to increase it is provided. An examples of citation histograms is given.

BACKGROUND

The study concerning the semantic representation of the knowledge contained in commentaries to Dante Alighieri's works, especially focused on primary sources which are the works of other authors cited by Dante in his texts, is summarized. The approach is composed of several steps: the first step of them started from the analysis of an Excel style sheet, where some pieces of knowledge included in a commentary of *Convivio* (Alighieri, 2014), a philosophical essay composed by Dante Alighieri in the years between 1304 and 1307, were organized and reported by an Italian scholar. In the Excel sheet, every note is given in a row, and is composed of the following pieces of knowledge:

- Number of chapter of the annotated text, represented as a pair of the form book chapter (e.g., 1.01 indicates the first chapter of the first book of *Convivio*);
- Number of the paragraph;
- The text fragment to which the note applies (e.g., “Sì come dice lo filosofo nel principio della Prima Filosofia”);
- The text of the note;
- The last three columns give information about a primary source reported into the note, structured as:
 - Author (e.g. Aristotele);
 - Title (e.g. Metafisica);
 - Thematic area (e.g. Aristotelismo).

In the second step, in order to create an ontology in RDF (Manola & Miller, 2004) for the semantic representation of the previous pieces of knowledge, the terms belonging to vocabularies used in the Digital Libraries domain were identified. For this purpose, several existing ontologies, e.g. CIDOC-CRM (Doerr, 2003), FRBR (Tillet, 2005), FaBiO (Peroni & Shotton, 2012) SKOS (Miles, Matthews, Wilson & Brickley, 2005) were investigated, and the terms that we considered useful to represent

our knowledge were chosen. In order to describe all the knowledge embedded in the commentary, we added classes and properties (Bartalesi, Locuratolo, Meghini & Versienti, 2013).

In the third phase, we built a RDF graph that represents the semantics of the commentary structured according to the ontology. After this preliminary study on commentary to *Convivio*, we used the developed ontology to represent other works of Dante as well as the knowledge carried by commentaries to them. Since the structure of knowledge reported in the Excel file is valid also for other Dante's works (i.e. *Monarchia* (Alighieri, 2014), *De Vulgari Eloquentia* (Alighieri, 2011), and *Vita Nova* (Alighieri, 2011)), the ontology was populated with other knowledge. Three experts extracted the previous reported pieces of knowledge from commentaries to these Dante's works, using a semi-automatic tool we developed. Finally, the RDF graph was stored into a Virtuoso triple store (Erling & Mikhailov, 2009).

The technical advantages of our approach are essentially three: (1) researchers can add classes and relationships to our ontology, thereby refining it, (2) our ontology can be linked to other ontology to extend the represented domain, (3) any user can download and use our model freely, using the paradigm of Linked Data.

Web services were developed to visualize the knowledge stored in the graph in form of charts and tables and CSV format. Using SPARQL query language to extract knowledge from Virtuoso, the services make it possible to address several tasks carried out by the scholars building an encyclopedia of Dante's works, starting with the visualization of the distribution of primary sources both in time and in the works of Dante. The overall goal is to shed light on the cultural context in which Dante wrote his works and on the development of Dante's reference library over time. Our diachronic analysis, in fact, aims at representing the evolution of Dante's knowledge about primary sources.

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/knowledge-acquisition-on-dante-alighieris-works/184209

Related Content

Enhancement of TOPSIS for Evaluating the Web-Sources to Select as External Source for Web-Warehousing

Hariom Sharan Sinha (2018). *International Journal of Rough Sets and Data Analysis* (pp. 117-130).
www.irma-international.org/article/enhancement-of-topsis-for-evaluating-the-web-sources-to-select-as-external-source-for-web-warehousing/190894

I-Schools and the Present Worldwide Trend and the Indian Scenario

Prantosh Kr. Paul and D. Chatterjee (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 2525-2534).
www.irma-international.org/chapter/i-schools-and-the-present-worldwide-trend-and-the-indian-scenario/112669

Fuzzy Decision Support System for Coronary Artery Disease Diagnosis Based on Rough Set Theory

Noor Akhmad Setiawan (2014). *International Journal of Rough Sets and Data Analysis* (pp. 65-80).
www.irma-international.org/article/fuzzy-decision-support-system-for-coronary-artery-disease-diagnosis-based-on-rough-set-theory/111313

Metaheuristic Algorithms for Detect Communities in Social Networks: A Comparative Analysis Study

About Ella Hassanien and Ramadan Babers (2018). *International Journal of Rough Sets and Data Analysis* (pp. 25-45).
www.irma-international.org/article/metaheuristic-algorithms-for-detect-communities-in-social-networks-a-comparative-analysis-study/197379

A Framework for Understanding Human Use of Computers

Andrew Basden (2008). *Philosophical Frameworks for Understanding Information Systems* (pp. 120-173).
www.irma-international.org/chapter/framework-understanding-human-use-computers/28082