

A web application for exploring primary sources: The DanteSources case study

Valentina Bartalesi, Carlo Meghini and Daniele Metilli
ISTI-CNR, Italy

Mirko Tavoni and Paola Andriani

Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa, Italy

Abstract

We present the methodological and technical process we adopted to develop DanteSources, a Web application that allows free access to the knowledge about Dante Alighieri's primary sources, i.e. the works of other authors that Dante cites in his texts. Up to now, this knowledge has been collected in many paper books, making it difficult for the scholars to retrieve it and to produce a complete overview of these data. Using Semantic Web technologies, we developed an ontology expressed in the Resource Description Framework Schema vocabulary providing the terms to represent this knowledge in a machine-readable form. A semi-automatic tool helps the scholars to populate the ontology with the data included in authoritative paper commentaries to Dante's works. Then, the tool automatically saves the resulting Resource Description Framework graph in a triple store. On top of this graph, we developed DanteSources, a Web application that allows users to extract and display the information stored in the knowledge base in the form of charts and tables. Finally, we report the results of a survey to collect suggestions from end-users on their interactions with DanteSources. The methodology and the tools we developed are easily reusable, e.g. to represent the knowledge about primary sources of other authors of the Italian and the international literature.

Correspondence:

Valentina Bartalesi, via G.
Moruzzi 1, 56124 Pisa, Italy.

E-mail:

valentina.bartalesi@isti.cnr.it

1 Introduction

One important kind of knowledge studied by scholars in the Humanities field is represented by the primary sources of literary texts, which are crucial in order to reconstruct the history of the creation of a text. Primary sources are the literary works of other authors that an author refers to in her/his texts (Dalton and Charnigo, 2004; Hellqvist, 2010; Malkmus, 2010). In particular, the primary sources

we considered are literary texts; we did not take into account different primary sources like monumental inscriptions. This study is a crucial aspect of this research field since it allows reconstructing the cultural background of an author and its evolution in time. Usually, this knowledge is reported in commentaries, written by scholars in natural language, which the readers use to study and decode several aspects of the literary texts. These commentaries are called secondary sources, i.e. documents that relate

or discuss information originally presented in a primary source. A secondary source contrasts with a primary source, which is an original source of the information being discussed. Secondary sources involve generalization, analysis, and interpretation of the original information. Unfortunately, commentaries are usually in natural language form and the knowledge contained in them is not machine-readable (Hendler, 2012), thereby limiting the possibility for scholars to obtain automatic inferences in order to derive new knowledge. For instance, mapping the distribution of the primary sources cited in a work would allow scholars to obtain a precise understanding of which sources influenced the author's work at a particular moment in time. In addition, some of the main research questions still unsolved that are at the basis of our research are: (1) Is it possible to retrieve from a text the textual fragments that refer to different primary sources? (2) What kinds of references are present in a text (e.g. explicit references, strict references)? (3) Is it possible to retrieve from a text the primary sources organized by thematic area (e.g. Aristotelianism, Astronomy, Physics)?

In order to overcome these limitations in the study of literary texts, we developed an ontology for representing the knowledge about primary sources and a Web application that allows visualizing this knowledge in a user-friendly format, such as charts and tables.

Our study is part of the 'Towards a Digital Dante Encyclopedia' project (2013–16), an Italian National Research project supporting scholars in formally expressing and accessing the knowledge about primary sources cited in the works of Dante Alighieri—the major Italian poet of the late middle ages—and more generally in literary texts.

In this article, we describe the methodology and the technical development process we followed to create DanteSources (<http://dantesources.org>), a Web application that allows free access to the information about Dante Alighieri's primary sources. We illustrate the entire process from the creation of the ontology for representing the knowledge on the primary sources to the development of the Web application.

One of the main goals of our work was to create an ontology providing a formal representation of the knowledge on primary sources of literary texts. To this aim, we adopted the Semantic Web technologies, in particular the Resource Description Framework Schema (RDF/S) vocabulary, which provides the terms to represent the knowledge in a machine-readable form. In the Digital Humanities literature, there are many ontologies focusing on different aspects of textual information. Each of these ontologies represents a set of possible interpretations of the source text(s). Our developed ontology (Bartalesi and Meghini, 2017) is the first attempt to formally represent this knowledge, since up to now vocabularies to express this information have not been developed and, to the best of our knowledge, at the moment there are no other research projects focussing on giving a logical representation of this knowledge. Furthermore, our ontology extends well beyond the specific author's primary sources we are considering in our project.

On top of our ontology, we developed the Web application DanteSources that allows visualizing the data in the form of charts and tables and exporting them in Comma Separated Value (CSV) format. In our project, these visualizations provide useful knowledge to the scholars who are working on the creation of a complete encyclopaedia of Dante's works. Furthermore, to the best of our knowledge, this is the first time that the information about Dante's primary sources is available in machine-readable format. Indeed, in the commentaries the knowledge about primary sources is expressed in natural language, thereby preventing automatic inferences of new information that may be useful for the studies of the scholars, e.g. the inferences can concern the total amount of references to a certain work. In addition to scholars and Dante's experts, our application could be useful for high school professors who want to analyse Dante's works with their students, discovering their primary sources in a quick and easy way. A benefit could derive also for university students in the Humanities and Digital Humanities who write their thesis on the primary sources used by Dante. For example, starting from the results of the DanteSources queries they could

compare the primary sources used by Dante before and after his exile.¹

The methodology and the tools we developed within the project are freely available on request for research aims, by e-mail to the authors. The tools are open-source and released under the GPLv3 licence.² Indeed, they are easily reusable in order to represent the knowledge about primary sources of other authors of the Italian and international literature, in addition to Dante Alighieri. Our methodology could also be applied to scientific literature in order to formally represent knowledge about its primary sources.

The article is structured as follows: in order to understand the functionalities of the DanteSources Web application, Section 2 reports the definitions of the main notions used in the Humanities field. Section 3 presents several applications developed about Dante Alighieri's works and a list of ontologies to represent textual knowledge. Section 4 describes the methodology we followed to develop our Web application. In Section 5 we present the details of the semantic model and its implementation to represent the knowledge about primary sources. In Section 6, we describe the software to populate the developed ontology. Then, Section 7 reports the description of the functionalities of our application and how the knowledge is presented to the final user. In Section 8, we report data about a user test we conducted in order to evaluate the usability of DanteSources. Section 9 reports our final remarks.

2 Background

The knowledge about the primary sources of literary works is usually reported in commentaries. A commentary is a text, usually reported below the text of the literary work, which contains a series of comments, or explanations about the work. Authoritative commentaries are those written by notable scholars and the scientific validity of them is recognized by the scientific reference community. These commentaries are composed of notes. A note is a textual fragment that refers to a particular expression in the original work. A note is composed of two parts: a piece of text of the original work and a

Manifestamente adunque può vedere chi bene considera, che pochi rimangono quelli che all'abito da tutti desiderato possano pervenire, e innumerabili quasi sono li 'mpediti che di questo cibo sempre vivono affamati. 7. Oh beati quelli pochi che seggiono a quella mensa dove lo pane delli angeli si manuca! e miseri quelli che colle pecore hanno comune cibo!³

I. 6.-7 & MANIFESTAMENTE: 'chiaramente'. & CHI BENE CONSIDERA: 'chi riflette attentamente' & L'ABITO DA TUTTI DESIDERATO: si tratta ovviamente della scienza. & SEGGIONO: 'siedono'. & SI MANUCA: 'si mangia'. Il pane degli angeli, nella tradizione veterotestamentaria è la manna (cfr. *Ps. 77, 25* «Panem angelorum manducavit homo»). Nel vangelo di Giovanni (6, 59) il nuovo pane del cielo, che si oppone alla manna dei padri, è Cristo che si offre come vero cibo. L'esegesi patristica, a partire da Agostino, aveva visto nel 'panis angelorum' il Verbo stesso (cfr. le *Enarrationes in Psalmos, ps. 77, 17, p. 1081*, seguite

Fig. 1 Example of commentary to Dante's *Convivio*. The highlighted note states that the fragment 'si manuca' refers to Psalm 77, the Gospel of John, and the *Enarrationes in Psalmos*

detailed comment of this piece of text developed by a scholar, where also information about primary sources are reported. An example is shown in Fig. 1. Commentaries and their notes neither have a standard structure, nor follow a standard codification; thus each scholar can codify her/his commentary using her/his preferred codification model. However, these codifications are usually very similar between them; thus it is possible to develop a software that, using different rules, automatically extracts the two components of the commentary, i.e. the reported text of the literary work and the related comment. An example of a commentary with a different codification from Fig. 1 is shown in Fig. 2.

Up to now, in our research we have taken into account six authoritative commentaries related to the following Dante Alighieri's works: *Convivio* (Alighieri, 2014a), a philosophical essay written between 1304 and 1307; *Vita Nova* (Alighieri, 2011b) and *Vita Nuova* (Alighieri, 1980) are two different versions of commentaries written by two different scholars on the same Dante's work, published in 1295, an expression of the medieval genre of courtly love; *De vulgari eloquentia* (Alighieri, 2011a), a Latin essay written between 1302 and 1305 discussing the relationship between Latin and vernacular; *Monarchia* (Alighieri, 2014b), a Latin treatise on secular and

II 7. UT PHYLOSOPHUS DOCET: «come il Filosofo insegna» (Anonimo, pp. 155-6); «come dice Aristotile» (Ficino); cfr. *Ethica ad Nicomachum*, 1094 b 12; 1098 a 27. Cfr. sopra, II II 1. EX MANIFESTIS SIGNIS ATQUE SAPIENTUM AUTORITATIBUS: sull'*auctoritas* come forma di premessa in un sillogismo che conduce a conclusioni probabili v. qui *Imbach* (p. 293), *Kay* (pp. 100-1, nota 21) e *Cassell* (pp. 304-5, nota 121), con rimando a Pietro Ispano, *Summulae logicales*, V, 36, ed. de Rijk, pp. 75-6, Cfr. più oltre, II V 6. ILLIUS POPULI GLORIOSI: cfr. sopra, II I 4 e più avanti, II V 5 e II IX 19.

Fig. 2 Example of commentary to Dante's *Monarchia*. The format of the commentary is different from that of *Convivio*

religious power; and *Rime* (Alighieri, 2011c), lyric poems that Dante wrote throughout his life, based on the poet's varied existential and stylistic experiences.

3 Related Works

3.1 Applications of Dante's works

Several Web applications that allow investigating different aspects of Dante's works are currently available. Unfortunately, none of them focuses on the primary sources or uses a Semantic Web approach to represent this information. In the following, we report an overview of the most important projects about Dante Alighieri's work.

The Dartmouth Dante Project (DDP)³ aims at publishing the full searchable text of the *Divine Comedy* and several of its commentaries in digital format (Hollander, 1989). On top of the DDP, the Web application Dante Lab⁴ was developed. This application allows the concurrent visualization of the original text of the *Divine Comedy*, some translations, i.e. English, French, and German, and more than seventy-five commentaries, each of them fully searchable. The Princeton Dante Project⁵ includes the full searchable text of the *Divine Comedy*, Dante's minor works, several commentaries, and multimedia resources (Hollander, 2013).

Differently from our approach, in the previous projects, the results of the queries are not exportable from the Web application and it is not possible to automatically extract and aggregate the primary

sources for each text. It is only possible to retrieve this information by reading the content of the commentaries on the computer screen.

The World of Dante⁶ (Parker, 2001) is a multimedia research tool that contains the whole text of the *Divine Comedy* with a basic semantic annotation that identifies people, places, deities and structures, and connections to digital objects such as images and music. However, the semantic knowledge is not represented through a formal ontology and is not available in a machine-readable format (Hendler, 2012).

Digital Dante⁷ offers the full text of all Dante's works with commentaries, illustrations, and recorded readings but not in a machine-readable form. Digital Dante provides a functionality, namely, Intertextual Dante, which links the *Divine Comedy*'s text to the corresponding Ovid's works. This functionality is limited to a single author, Ovid, and its static implementation is based on a manual annotation of the texts (Van Peteghem, 2015).

Dante Online⁸ is a website that allows visualizing and searching the full text of all of Dante's works. It also includes a biography with an interactive timeline, and a database listing hundreds of manuscripts of the *Divine Comedy* with bibliographic information. This site neither contains the text of commentaries to Dante's works nor information on the primary sources.

DanteSearch⁹ allows users to perform morphological and syntactic queries on the full text of Dante's works. This application focusses on the morpho-syntactic aspects of Dante's texts but no semantic annotation is performed on them.

Finally, DaMA¹⁰ is a digital archive containing the full text of Dante's works and several medieval primary sources in XML (eXtensible Markup Language)-TEI (Text Encoding Initiative) format. DaMA allows visualizing the text of the medieval primary sources used by Dante but, differently from DanteSources, it does not provide links between each fragment of a Dante's work and the corresponding primary sources.

3.2 Ontologies

In the Semantic Data Modelling field, several ontologies have been developed in order to represent

textual knowledge. We have investigated the scientific literature and the existing standards in the Digital Libraries field (Altman, 2006), both official and *de facto*. Up to now, a formal ontology for representing knowledge about the primary sources of literary texts is not present in the scientific literature. As explained in Bartalesi and Meghini (2017), we did not find a single ontology able to represent all the classes and properties we identified when analysing the commentaries on Dante's works. In order to maximize the interoperability of our ontology, we reused classes and properties from the following ontologies,¹¹ adding our own classes and properties only if necessary.

- **Functional Requirements for Bibliographic Records Object Oriented (FRBRoo)**. It is essentially the Functional Requirements for Bibliographic Records (FRBR) ontology (Riva, Doerr and Zumer, 2008) expressed in an object-oriented form that harmonizes the FRBR and the CIDOC Conceptual Reference Model (CRM) (Doerr, 2002) to better share library and museum information.
- **Dublin Core**. The Dublin Core Metadata Element Set¹² is a vocabulary of fifteen properties for usage in resource description. These elements of the Dublin Core are part of a larger set of metadata vocabularies and technical specifications developed by the Dublin Core Metadata Initiative.
- **Simple Knowledge Organization System (SKOS)**. It is a model for sharing and linking knowledge organization systems such as thesauri, classification schemes, subject heading systems, and taxonomies within the framework of the Semantic Web (Miles, Matthews, Wilson, and Brickley, 2005).
- **Friend of a Friend (FOAF)**.¹³ It is a machine-readable ontology linking people and information using the Web. In particular, FOAF describes persons, their activities, and their relations to other people and objects.
- **Document Components Ontology (DoCO)**. This ontology (Constantin, Peroni, Pettifer, Shotton, and Vitali, 2016) allows describing document components in Resource Description Framework (RDF). Such components are both structural (e.g.

block, chapter, heading, paragraph, section) and rhetorical (e.g. abstract, introduction, results, discussion, conclusions, acknowledgements, bibliography).

- **FRBR-Aligned Bibliographic Ontology (FaBiO)**. It is an ontology (Peroni and Shotton, 2012) to record and publish bibliographic records created by scholars on the Semantic Web. FaBiO entities are primarily textual publications such as books, magazines, newspapers, and journals and items of their content such as poems, conference papers, and editorials.
- **Citation Typing Ontology (CiTO)**. It is an ontology (Shotton, 2010) for the characterization of bibliographic citations, both factually and rhetorically and for their publication on the Web.
- **The Open Annotation Core Data Model**. The Open Annotation Core Data Model (Sanderson, Ciccarese, and Van de Sompel, 2013) specifies an approach to associate annotations with resources, using a methodology that conforms to the Architecture of the World Wide Web and the Linked Data initiative.

In the last decades, ontologies and Semantic Web technologies have been applied in the Humanities field in order to improve the interoperability of the data, e.g. through analysis of iconographic representation in manuscripts, data aggregation for digital history, semantic annotation of literary texts. For instance, in González-Blanco, Seláf, Del Rio Riande, Martínez Cantón, and Martos Pérez (2014) a semantic model to connect repertoires of poetic writings is described. Another example is Biblissima (Gehrke, Frunzeanu, Charbonnier, and Muffat, 2015), an online digital library, which provides easy and coordinated access to a huge and complex mass of documentation on manuscripts and early printed books from the 8th to 18th centuries. In order to handle the heterogeneity of the database formats and the variety of the data (manuscript cataloguing databases, textual editions, iconographic databases), in the Biblissima project the CIDOC CRM and FRBRoo are chosen as reference ontologies to facilitate the internal mapping to a single common model and to allow the partners of the project to expose their data in RDF compliant to a globally established standard.

In the WarSampo project (Hyvönen et al., 2016), the Semantic Web and Linked Data technologies are



Fig. 3 The main phases of our approach: from the textual knowledge to the web application

used for modelling, harmonizing, and aggregating data about the Second World War. This approach makes it possible, for both historians and laymen, to study history in a contextualized way where linked data sets enrich each other.

Wagner and Caesar (2016) describe a digital collection of more than a hundred sources from Iberian theologians and jurists of the 16th and 17th century. To formally represent data about the primary sources and the authors, the FOAF, BIO,¹⁴ Relationship,¹⁵ and SPAR¹⁶ ontologies were used. Niccolucci and D’Andrea (2006) show an ontology for the 3D visualization of cultural heritage; Lana et al. (2014) present an ontology for annotating geographical places in texts. Currently, to the best of our knowledge, a research project for the formal representation of the primary sources of literary texts using the Semantic Web technologies does not exist yet. Taking into account the previous projects about the works of Dante Alighieri and the developed ontologies for representing aspects of literary texts, we developed DanteSources, where we applied the technologies of the Semantic Web to literary texts in order to represent the knowledge on Dante’s primary sources.

4 Methodology

The main phases of the development process of DanteSources were: (1) analysing the representation requirements about citations to primary sources, in collaboration with a Dante’s expert; (2) creating an ontology responding to the requirements, providing a formal representation of the terms required to express knowledge about the primary sources, using the Semantic Web language; (3) developing a semi-automatic tool for populating the ontology and storing the resulting data in a knowledge base; (4) developing a Web application running on top of

the knowledge base that makes inferences based on the contents of the knowledge base and visualizes these inferences using charts and tables. Figure 3 summarizes the main phases of our approach.

First of all, semi-structured interviews were performed with four Dante’s experts who collaborated with us during the whole project, in order to clarify the requirements of the project, in particular, the meaning of the data and the functionalities needed for the citations to primary sources. We asked to the experts to identify the relevant knowledge we had to extract from the commentaries and to suggest to us how this knowledge had to be further elaborated.

According to the scholars, our ontology has to represent the following pieces of knowledge:

- the textual fragment of Dante’s works which the note refers to (e.g. ‘As the Philosopher says in the beginning of the first Philosophy’);
- the textual fragment of the note citing a primary source (e.g. ‘These are the words that open the Aristotle’s *Metaphysics* I 1, 980a 21’);
- the structure of Dante’s work containing the fragment (e.g. book, chapter, and paragraph, poem, *cantica*, etc.);
- the name of the author of the primary source (e.g. Aristotle);
- the title of the primary source (e.g. *Metaphysics*);
- the thematic area of the primary source (e.g. Aristotelianism, Bible, Medical Science, Scholastic Philosophy);
- the entire text of the commentary, thus making available to the scholars the context of the reference;
- the type of reference, which can be:

explicit, if the reference is explicitly made by Dante, as in ‘As the Philosopher says at the beginning of the First Philosophy’, where the

Philosopher is Aristotle and the First Philosophy is *Metaphysics*; *strict*, if the reference is indicated by a scholar, and refers to a specific work as in ‘SI MANUCA: it is the bread of the angels, the *manna* as called in the Old Testament (Ps. 77, 25 *Panem angelorum manducavit homo*)’; *generic*, if the reference is indicated by a scholar, and refers to a concept (e.g. ‘Medieval comments to Aristotle’s works’).

Once the semantics was understood, we created an ontology to represent it, with the additional requirement of maximizing reusability. To this end, we investigated several existing vocabularies to identify useful classes and properties to represent this knowledge. In particular, we analysed the ontologies addressing issues related to the textual domain, reported in Section 3. We chose the classes and properties from those ontologies that we considered the most appropriate. When we could not find appropriate classes and properties in the existing ontologies to represent the necessary kind of knowledge, we created our own. Following the Linked Data recommendations, we used HTTP Uniform Resource Identifiers (URIs) for denoting resources and we provided a description for every resource. Then, we validated our ontology by showing to the experts the semantic network resulting from encoding of the citations of *Convivio* (Alighieri, 2014a). Once the ontology was validated, we invited the domain experts participating in the research project to extract the relevant information about primary sources from authoritative commentaries about other Dante’s works, and to encode such information in our ontology. In order to facilitate the experts in this task, we developed a semi-automatic tool, allowing them to easily export the extracted knowledge in XML form. The resulting XML files were automatically exported in RDF format. The resulting RDF graph was saved into a Virtuoso triple store.

In parallel with the ontology population phase, we developed a Web application to explore the semantic network of Dante’s works and their primary sources. The requirements to develop the Web application we identified by interviewing the Dante’s experts are the following:

- extracting either the primary sources or the cited authors or the thematic areas for each Dante’s work;
- extracting all the Dante’s works citing a specific primary source, author and thematic area; and
- extracting the types of citation (explicit, strict, generic) from each Dante’s work and primary source

The Web application provides search functionalities corresponding to each requirement and visualizes the results through charts and tables. In the final stage of the project, a group of external users were asked to fill in and submit a survey for the evaluation of the Web application.

5 The Ontology

We developed an ontology for the semantic representation of the structure of the works of Dante and of the knowledge related to their primary sources. An ontology is a set of classes and properties that describe a knowledge domain, along with axioms that capture the formal meaning of these classes and properties. Our ontology is expressed in RDF/S, a vocabulary for data modelling recommended by the World Wide Web Consortium (W3C; Manola, Miller and McBride, 2004). The schema is available on the DanteSources website.¹⁷ The classes and properties defined in RDF/S are then used in an RDF graph, which is a collection of RDF statements representing some slice of reality. An RDF statement is a triple in the form of subject–predicate–object (World Wide Web Consortium, 2014).

In order to obtain an ontology for the semantic representation of the pieces of knowledge defined as requirements (see Section 4), we analysed several existing ontologies as reported in Section 3, and we chose the classes and the properties that we considered the most appropriate for our purposes, respecting the constraints established in the source ontologies, i.e. we preserved the definitions of the classes and the domains and ranges of the properties. Furthermore, we added our own classes and properties for representing pieces of knowledge that the existing ontologies had not defined.

Almost 90% of the classes and properties of our ontology come from existing ontologies.

In the following, the classes and properties we reused as well as the new ones are reported. A more detailed description of the ontology is reported in [Bartalesi and Meghini \(2017\)](#).

As a notational convention, we use prefixed qnames (e.g. foaf:Person) to denote terms reused from other vocabularies, whereas we use local names (e.g. hasCommentary) to denote terms of our own ontology.

From the FRBRoo ontology we reused two classes to represent the related aspects of our textual knowledge: efrbroo:Work, corresponding to a specific work, cited by the commentaries (e.g. *De consolatione philosophiae*) without reference to a specific edition; efrbroo:ExpressionFragment, corresponding to (1) a fragment of a Dante's work a commentary refers to and also (2) to a fragment of a commentary that refers to a specific primary source.

In our model a work has an author (e.g. Albertus Magnus) and a thematic area (e.g. Rhetoric, Scholasticism, Medical science, etc.). We represented these pieces of knowledge using the classes foaf:Person and skos:Concept, respectively. To link the work, author, and thematic area with their corresponding names in natural language, we used the property dcterms:alternative.

To link the efrbroo:Work class with its author and thematic area, we used two Dublin Core properties: dc:creator and dc:subject, respectively.

To describe the structure of the resource from which the fragment is extracted, we reused the following classes from the FaBiO and DoCo ontologies:

- (1) doco:Paragraph, a self-contained unit of discourse that deals with a particular point or idea;
- (2) fabio:Poem, an artistic work written with an intensity or beauty of language more characteristic of poetry than of prose;
- (3) doco:Line, a line in poetry, i.e. a unit of language into which a poem is divided;
- (4) doco:Chapter, a principle division of the body matter of a large document;

- (5) fabio:Book, a non-serial document that is complete in one volume or a designated finite number of volumes.

The Open Annotation Model is the standard adopted by the W3C, for this reason we decided to be compliant with this ontology and to guarantee the interoperability with it. The Open Annotation Model defines an annotation as composed of 'a single Body, which is the comment or other descriptive resource, and a single Target that the Body is somehow about'.

In order to link the annotation with its components, we used the following classes and properties:

- (1) oa:Annotation class represents a commentary;
- (2) oa:TextualBody, which identifies the textual content of the Annotation;
- (3) oa:hasBody, which is the property that links the Annotation and its textual body;
- (4) oa:hasTarget, which is the property that links the Annotation and its related resource (in our case, it represents the primary source);
- (5) efrbroo:ExpressionFragment, i.e. a fragment of an annotation that refers to a specific primary source.

We also defined a new property hasCommentary to directly link the fragment of Dante's text to the body of the note that describes it and identifies a related primary source.

Since in the content of the same commentary there can be several references to different primary sources, we added the new property hasCitingFragment to relate the commentary with its own citing fragments. We defined hasCitingFragment as a subproperty of efrbroo:hasFragment.

The fragment of a commentary refers to a primary source through three subproperties of the cito:cites property we defined: (1) citesAsExplicitCitation, (2) citesAsStrictCitation, and (3) citesAsGeneralCitation.

Figure 4 shows an example of the representation of a note in our ontology. The individual danteExpressionFragment represents the fragment of Dante's text 'Sì come dice lo Filosofo nel principio de la Prima Filosofia, tutti li uomini naturalmente desiderano di sapere' (As the Philosopher says in the beginning of the First Philosophy, all men by nature

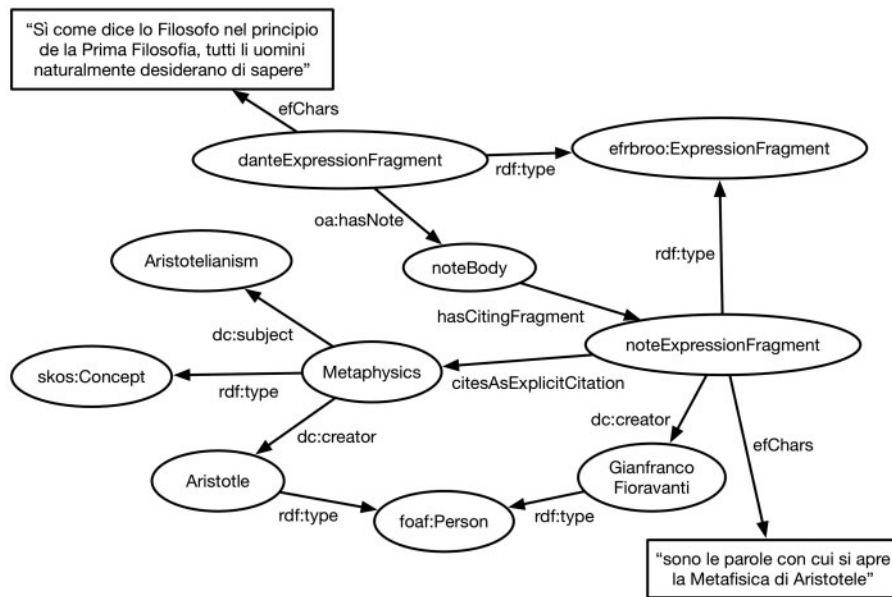


Fig. 4 An example of the representation of a note in our ontology

desire to know). Dante's fragment has a note, the individual `noteBody`, which is connected to it through the property `oa:hasNote`. The note contains another fragment of text, `noteExpressionFragment`, which states 'sono le parole con cui si apre la Metafisica di Aristotele' (these are the words that open Aristotle's *Metaphysics*). This fragment is connected to the primary source through the property `citesAsExplicitCitation`. The cited work is *Metaphysics*, which is Aristotle's First Philosophy. *Metaphysics* has `dc:creator` Aristotle and `dc:subject` Aristotelianism.

Using the semi-automatic tool described in Section 6, three Dante's experts populated the ontology with the knowledge extracted from the authoritative commentaries listed in Section 2, i.e. *Convivio*, *Vita Nova*, *Vita Nuova*, *De vulgari eloquentia*, *Monarchia*, and *Rime*. The obtained RDF graph was eventually stored in a Virtuoso triple store (Erling and Mikhailov, 2009).

The information stored in our knowledge base may be incomplete and can provide an incoherent representation of Dante's works. For instance, two commentaries written by two different scholars who identify two different primary sources for the same fragment of Dante's text may raise incoherency.

Indeed, on the one hand our ontology is able to represent such conflicting annotations because it does not impose a unicity constraint on the primary sources related to one piece of text. On the other, incompleteness may arise since in our knowledge base we stored only some commentaries to Dante's works. However, the ontology allows the addition of new commentaries at any time in order to increase knowledge completeness.

A further problem in order to formally represent knowledge is the fact that the commentary of a literary text can be open to different interpretations that can produce inconsistencies in the ontology. To avoid these types of inconsistencies, we are considering extending the ontology adding to the commentary the information about who has interpreted it if s/he is a different person from the author of the commentary itself.

The main advantages of using an ontology instead of a traditional database model are extensibility (researchers can add classes and properties in order to refine the ontology) and interoperability (researcher can reuse the data by integrating our ontology with theirs, and this is much easier with Linked Data than with traditional data).

Furthermore, it is possible to automatically check the consistency of the knowledge base with respect to the ontological model using a semantic reasoner (Mishra and Kumar, 2011). Moreover, the semantic reasoners allow to automatically infer new implicit knowledge. Furthermore, the ontology can be linked to other existing ontologies to extend the represented domain. Finally, any user can download and use the ontology schema freely, using the paradigm of Linked Data.

6 The Population Tool

In order to facilitate the population process of the ontology, we implemented a semi-automatic tool. The population of the ontology was carried out by three of the four Dante's experts who participated in our project (under the supervision of the fourth scholar, who coordinated with the Dante's expert group). The six commentaries (the full list is reported in Section 5) we started from were in PDF and MS Word format. The total size of the commentaries is about 1,600,000 words. Through a Java software we developed, we pre-processed the texts of the commentaries in order to automatically extract the following information: (1) the fragment of Dante's text which the note refers to, (2) the entire text of the note, (3) the title of Dante's work along with the number of the book, chapter, and paragraph which the fragment belongs to. Since in the Humanities a standard format to write a commentary does not exist and each scholar writes her/his commentary using different formats, in our software for each commentary we used different regular expressions in order to identify specific textual fragments and pieces of knowledge. Once extracted, these data were stored in an XML file. This automatic phase can be estimated in 6 person-days (7 h per day) for two Information Science researchers, including the development of different regular expressions for each commentary.

After this first phase of pre-processing, we developed a tool with a Web interface in order to facilitate the insertion of the remaining data by the experts. The tool is developed in HTML5 and JavaScript using the jQuery¹⁸ and jQuery UI¹⁹

libraries. It takes as input the XML file, where the data extracted from the commentaries are stored, and automatically shows these data in the corresponding fields of the interface. It also facilitates the user's work by storing the inserted data in an IndexedDB²⁰ database and by suggesting it to the user through autocomplete fields.

After analysing the notes to Dante's works reported in the commentaries, the experts inserted through the interface the following remaining pieces of knowledge: (1) the fragment of the note referring to a primary source, (2) the title of the primary source, (3) the name of the author, (4) the thematic area which the primary source belongs to, and (4) the type of reference. For primary sources, authors, and thematic areas, the experts also inserted the corresponding URIs. For primary sources and authors, the URIs were imported from the DBpedia²¹ knowledge base. For thematic areas, they were imported from the Nuovo Soggettario²² thesaurus. For instance, the URI for Aristotle's Poetics is [http://dbpedia.org/page/Poetics_\(Aristotle\)](http://dbpedia.org/page/Poetics_(Aristotle)), and the URI for Aristotelianism is <http://purl.org/bnct/tid/19144>. When the URIs were not available on DBpedia or in the Nuovo Soggettario, custom URIs were automatically assigned by the population tool.

Figure 5 shows the interface of the tool.

As the data were gradually added, they were saved as lists in a JavaScript Object Notation (JSON) file, which was used to implement the auto-completion menus of the interface. In particular, the auto-completion menus were implemented for: the primary sources' titles, the names of authors and of thematic areas, and the corresponding URIs. In the JSON file a mapping between primary sources, authors, thematic areas, and URIs was also defined. Therefore, when an expert chooses a title of a primary source, the interface automatically completes the related fields (i.e. author, thematic area, and URIs). This functionality allowed the experts to reduce (1) the time for populating the ontology and (2) the possibility to make mistakes by inserting the data manually. The insertion in the knowledge base of the data extracted from a portion of about 15,000 words of the first commentary analysed, i.e. *Convivio*, was done manually by a single expert, since the tool was under development. We

PREVIOUS	SAVE NOTE	NEXT
----------	-----------	------

Work:

Book: Chapter: Paragraph:

Dante's Text:

Note Body:

Fragment:

Primary Source: +

Cited Author: +

Thematic Area: +

Source Type: ↓

Reference Type: ↓

ADD CITATION

Source Title:

Source URI:

Author Name:

Author URI:

Area Name:

Area URI:

ADD NEW DATA

SAVE JSON

Fig. 5 The interface of the population tool

compared the number of errors identified in the manually inserted data with the number of errors in the data inserted using the population tool for another portion of *Convivio* of the same length. We noticed that the number of errors was about 20% greater in the manually inserted data than in the other. Furthermore, the manual insertion required about 30% more time than the insertion using the tool.

At the beginning of our research, before the development of the Java software that pre-processes the texts of the commentaries and of the Web interface used by the experts, we estimated that inserting all data manually (about 15,000 words for *Convivio*), the work of an expert is about 8 person-days (7 h per day). Using the Java SW and the Web interface the time was reduced to 2.5 person-days (7 h per day).

```
<Testo>SI MANUCA</Testo>
<body>'si mangia'. Il pane degli angeli, nella tradizione veterotestamentaria è la manna
<Source>
<Libro></Libro>
<Capitolo>1</Capitolo>
<Paragrafo>7</Paragrafo>
</Source>
<Citazione>
<Frammento>6, 59</Frammento>
<Tipo>CONCORDANZA STRINGENTE</Tipo>
<InfoOperaCitata>
<OperaURI>http://dbpedia.org/resource/Gospel_of_John</OperaURI>
<TitoloOpera>Vangelo di Giovanni</TitoloOpera>
<Edizione></Edizione>
<Autore>
<Name>Giovanni</Name>
<URI>http://dbpedia.org/resource/John_the_Evangelist</URI>
</Autore>
<Area>http://perunaenciclopediadantescadigitale.eu/resource/bibbia</Area>
<TipoOpera>WORK</TipoOpera>
</InfoOperaCitata>
</Citazione>
```

Fig. 6 A piece of XML file produced by the population tool

We estimated that, using the tool, the manual work of each expert for inserting all data was about 100 person-days (7 h per day). In addition, we can compute a further revision of the quality of the inserted data that we can estimate in about 15 person-days (7 h per day).

The data added by the experts are saved in the same XML file that the interface takes as input, enriching it. The XML Schema follows the structure of our ontology. Figure 6 shows a piece of an XML file produced by the tool.

The XML generated by the tool is loaded into a Java triplifier that converts it to intermediate Java objects. The knowledge stored in the Java objects is finally converted to RDF using the free and open-source Java framework Apache Jena,²³ which provides an Application Programming Interface (API) to write RDF graphs. The resulting graph is automatically stored into a Virtuoso triple store. The RDF file stored in Virtuoso has a size of 29.2 MB and contains 187,328 triples at present.

7 The DanteSources Web Application

The DanteSources Web application²⁴ allows extracting and displaying the information stored in the knowledge base to support scholars in discovering and exploring Dante's primary sources.

In the last year DanteSources had 4,956 unique visitors, for a total of 30,501 visited pages. The majority of the users are from Italy (66%), 7% of the

users are from the USA, 3% from Germany, 3% from France, 2% from the UK. The remaining 19% of the users are from different countries in the world. In all, 85% of the users access DanteSources from a desktop computer and 15% from a mobile device.

7.1 Software development

DanteSources is developed in Java and JavaServer Pages. JavaBeans and the Apache Jena library are used to perform SPARQL (Prud'Hommeaux and Seaborne, 2008) queries in order to extract information from the underlying RDF knowledge base stored in Virtuoso. It currently uses Virtuoso version 7.2 and Jena version 3.1.

The Web interface is written in HTML5 and CSS3, using JavaScript and Ajax functions. In particular, we used the jQuery JavaScript library and Ajax functions to implement the search menus, which are described in detail in subsection 7.2. jQuery was also used to implement the tables that show the content of the textual fragments (an example of the table visualizing textual fragments is reported in Fig. 5).

DanteSources shows the knowledge about the primary sources cited by Dante in the form of tables and column bar charts. To implement the charts, we used the Highcharts²⁵ JavaScript library. Highcharts allows exporting the charts in the following formats: PDF, PNG, JPEG, SVG. Furthermore, we implemented an additional JavaScript function allowing users to automatically export and download the data in CSV format.²⁶ This feature is important since it allows scholars to apply further data analyses on raw data, in addition to the ones already provided by the application's queries. DanteSources is provided with a responsive layout that allows an easy reading and navigation with a minimum of resizing, panning, and scrolling across a wide range of devices, from desktop computers to mobile phones (Bartalesi, Meghini, Metilli, and Andriani, 2016).

7.2 The user interface

Currently, DanteSources provides eight different predefined SPARQL queries to extract data. They can be distinguished into three different groups. This distinction is based on the results produced

Fig. 7 The search form for the first group of queries

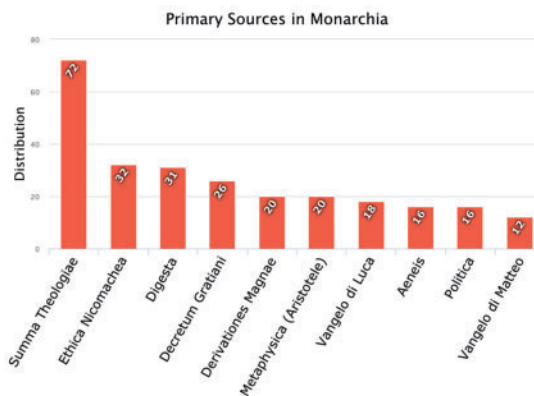


Fig. 8 The ten primary sources with the highest distribution in *Monarchia*

by the queries. Three queries are in the first group, which return knowledge about the distribution of the works, the authors, and the thematic areas mentioned by Dante. Three queries are in the second group, which return the distribution of a particular primary source, a cited author, and a thematic area. Two queries are in the last group, which return the distribution of the three types of reference to primary sources.

The three queries belonging to the first group produce column bar charts regarding the distribution of the works, the authors, and the thematic areas mentioned by Dante. Through the search form shown in Fig. 7, users can choose either one work of Dante or all his works and, in addition, a specific subpart of the work (e.g. a book).

Figure 8 shows the chart of the distribution of the ten most-cited primary sources in *Monarchia*. This visualization allows the scholar to immediately get an idea of the main sources that have influenced Dante in writing a specific work.

Dante's work	Book	Chapter or Poem	Paragraph or Verse	Fragment of Dante's work	Reference type	Fragment of the primary source	Thematic area of the primary source	Author of the primary source
Monarchia	1	11	12	SUA ... IURISDICTIONE TERMINATUR OCCEANO SOLUM	Strict reference	I 286-7: Nascetur pulchra Troianus origine Caesar, / imperium Oceano, famam qui terminet astris	Epica latina	Virgilio
Dante fragment (Monarchia , book 1, chapter 11, par. 12) Sed Monarcha non habet quod possit optare: sua nanque iurisdictione terminatur Oceano solum , quod non contingit principibus aliis, quorum principatus ad alios terminantur, ut puta regis Castelle ad illum qui regis Aragonum. Ex quo sequitur quod Monarcha sincerissimum inter mortales iustitie possit esse subiectum.					Source fragment (Aeneis , book 1, verses 286-287) cum domus Assaraci Phthiam clarasque Mycenae servitio premet ac victis dominabitur Argis. Nascetur pulchra Troianus origine Caesar, imperium Oceano, famam qui terminet astris , Iulius, a magno demissum nomen Iulo. Hunc tu olim caelo spoliis Orientis onustum.			
Monarchia	1	15	10	OMNES CONSEQUENTIE SUPERIORES	Strict reference	Non disponibile	Epica latina	Virgilio
Monarchia	2	3	6	DIVINUS POETA NOSTER VIRGILIUS	Explicit reference	Non disponibile	Epica latina	Virgilio
Monarchia	2	3	9	IDEM IN SEXTO	Explicit reference	Aen. VI 162-235	Epica latina	Virgilio

Fig. 9 The first rows of the table regarding Virgil's *Aeneid* cited in *Monarchia*

Clicking on the name of each primary source, author, and thematic area visualized on the charts, it is possible to receive more information. In particular, clicking on the title of a primary source, the tool shows a table reporting the following details: (1) the book, (2) the chapter, (3) the paragraph or poem and (4) the fragment of the Dante's work in which the primary source is cited, (5) the type of reference, (6) the reference to a fragment of the primary source cited in the commentary, (7) the thematic area, and (8) the author of the primary source.

Clicking on a table row, it is possible to visualize the fragment of Dante's text and the fragment of the corresponding primary source. An example of this table is shown in Fig. 9. We are importing in DanteSources the text of the primary sources included in DaMA in XML/TEI format. Up to now, the application includes the text of five primary sources but we are working to add the complete list. Figure 10 shows the chart of the distribution of the ten most-cited authors in *Convivio*.

Similarly, as shown in Fig. 11, clicking on the name of a cited author, the tool shows a table

reporting all the works of that author cited in Dante's text chosen by the user. Furthermore, for each primary source the tool reports the book, the chapter, and the paragraph (or poem) of Dante's text in which the author is cited. Through this table, the scholar can directly access the paragraph in Dante's work in which the poet refers to a specific primary source.

Figure 12 shows the chart of the ten thematic areas with the highest distribution cited in *De vulgari eloquentia*.

Clicking on one name of the thematic areas, the user obtains the following information: the primary sources included in that area, their authors, and the book, chapter, and paragraph (or poem) of Dante's work where the thematic area is cited. Figure 13 shows the first rows of the table regarding the thematic area Scholasticism in *De vulgari eloquentia*. Using this search, a scholar can immediately access all the primary sources belonging to a specific thematic area.

The three queries of the second group allow visualizing the distribution of a particular primary

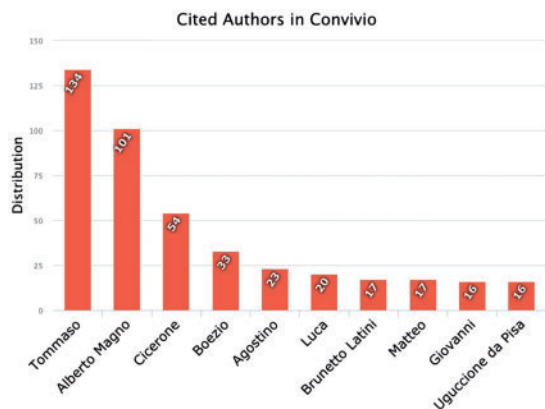


Fig. 10 The ten cited authors with the highest distribution in *Convivio*

Dante's Work	Book	Chapter	Paragraph	Primary Source
Convivio	1	1	7	Enarrationes in Psalmos
Convivio	1	2	14	Confessiones
Convivio	1	3	4	Contra Academicos
Convivio	1	4	9	Confessiones
Convivio	1	5	13	De civitate Dei
Convivio	2	1	1	De beata vita

Fig. 11 The first rows of the table regarding Saint Augustine cited in *Convivio*

source, a cited author, or a thematic area, respectively, in the form of charts. In order to improve the usability of the search, the application provides two different search forms for these queries: (1) an auto-complete menu where the user can type the title of the cited work, the author, or the thematic area; (2) an alphabetically ordered list where the user can select the title of the primary source, the name of the author, or the thematic area.

The distribution data are available for both an entire Dante's work, e.g. *Monarchia*, and its subparts, e.g. books, chapters. By clicking on one bar in the chart, the application allows to visualize details about the subparts of one work. Figure 14 shows the distribution of Aristotle's *Nicomachean Ethics* in Dante's works and in some of their subparts.

The two queries of the last group allow visualizing the distribution of the three types of reference to primary sources, explicit, strict, and generic, focusing either on Dante's works or on a single primary source.

Figure 15 shows the distribution of the three types of reference in the works of Dante. By clicking on the bars of the chart, it is possible to visualize the data regarding the distribution of the types of reference also for the subparts of Dante's works, i.e. books and chapters.

Figure 16 reports the distribution of the three types of reference for the works of Seneca cited by Dante in his texts. Having the type of reference for each author is a useful indication for the scholar to understand the influence of that author on Dante's work (e.g. several explicit references to a particular author indicate that that author had a significant influence for the poet).

Finally, DanteSources makes available a SPARQL endpoint so that any user can make her/his own query to the knowledge base.

Analysing the logs of the Apache Web server, we discovered that the most performed query is the one by primary source (37%), followed by the one showing all cited authors in a specific Dante's work (20%). The third is the one visualizing the distribution of a specific cited author in all of Dante's works (16%).

8 Usability Evaluation

We conducted a survey on the interaction with DanteSources via PCs and mobile devices to collect suggestions and comments on possible usability issues observed by the users. For a detailed description of the questions and the answers of our usability test, and for an aggregate analysis of the answers see Bartalesi et al., 2016.²⁷

The survey included twenty-eight questions about the main aspects regarding the usability of the Web application. We collected suggestions and comments from twenty-six users. In all, 25% of the users were students, 20% were researchers, 12% were university professors, 4% were school teachers, and 37% of the users are generic users. Likert five-

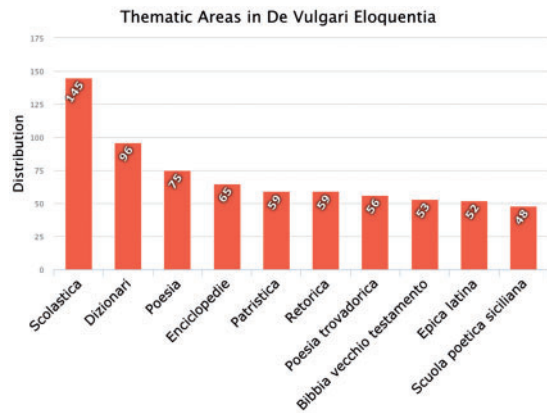


Fig. 12 The ten thematic areas with the highest distribution cited in *De vulgari eloquentia*

scale values from 1 (very negative) to 5 (very positive) were used to express the opinion.

Positive responses on the general interaction have appeared from the survey: 85% of the users expressed a distinctly positive opinion (scores 4 and 5). About 80% reported a satisfactory opinion on the navigation (scores 4 and 5) and on the readability of charts and tables (scores 4 and 5) as well. Hundred percent of the users who accessed DanteSources from a mobile device (five users) declared that the interaction was easy and clear (scores 4 and 5). On the other hand, the main problems identified concern: (1) the presence of a large blank space between the search forms and the results, (2) the hidden position of the explanation of the charts' functionalities, (3) the lack of text visualization for the cited fragments and for the entire texts of the primary sources and Dante's works. The users also suggested to enrich the knowledge base with other Dante's works, especially with the *Divine Comedy*. We are working to modify and enrich DanteSources taking into account the difficulties highlighted by the users and the given suggestions.

The scholars who took part in the evaluation confirmed that, to the best of their knowledge, this is the first time that the information about Dante's primary sources is available in digital format. This availability improves and makes the searches by Dante's experts more efficient. Indeed, in the commentaries in paper format, the

Dante's Work	Book	Chapter	Paragraph	Author	Primary Source
De vulgari eloquentia	1	1	2	Tommaso	Summa Theologiae
De vulgari eloquentia	1	1	2	Tommaso	Expositio libri Posteriorum Analyticorum
De vulgari eloquentia	1	2	2	Tommaso	Sententia libri Politicorum
De vulgari eloquentia	1	2	2	Tommaso	Sententia libri Politicorum
De vulgari eloquentia	1	2	2	Tommaso	Summa Theologiae

Fig. 13 The first rows of the table regarding the thematic area Scholasticism in *De vulgari eloquentia*

knowledge on the primary sources is expressed in natural language. This situation limits scholars in their advances insofar as it prevents automatic inferences of new information that may be useful for their studies, e.g. regarding the primary sources, the inferences can concern the total amount of references to a certain work. Eventually, the digitization of the knowledge about primary sources allows scholars to have a comprehensive overview of the data that can be easily exported to make further analyses.

9 Conclusions and Future Works

In this article we have presented the methodology and the development process we adopted to create DanteSources, a Web application that allows free access to the knowledge on Dante Alighieri's primary sources. In order to represent this knowledge, we developed an ontology, expressed in RDF/S, on top of which DanteSources was created. The main advantages of using an ontology instead of a traditional database are the following: (1) it allows researchers to add classes and properties thereby refining the ontology, (2) the ontology can be linked to other ontologies to extend the represented domain, (3) any user can download and use the knowledge base freely, using the paradigm of Linked Data. Furthermore, this is the first time that the knowledge about Dante's primary sources is coded in digital format. This availability allows

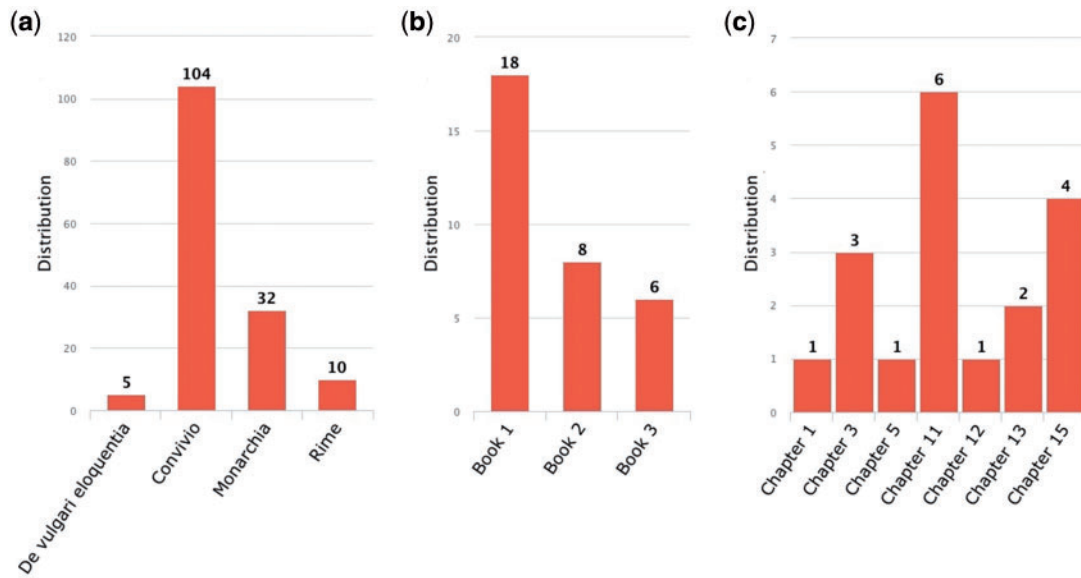


Fig. 14 The distribution of *Nicomachean Ethics* in Dante's works (a), in the three books of *Monarchia* (b) and in the chapters of the first book of *Monarchia* (c)

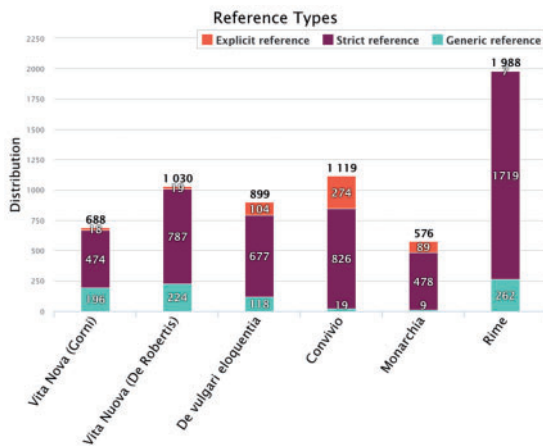


Fig. 15 The distribution of the three types of reference in Dante's works

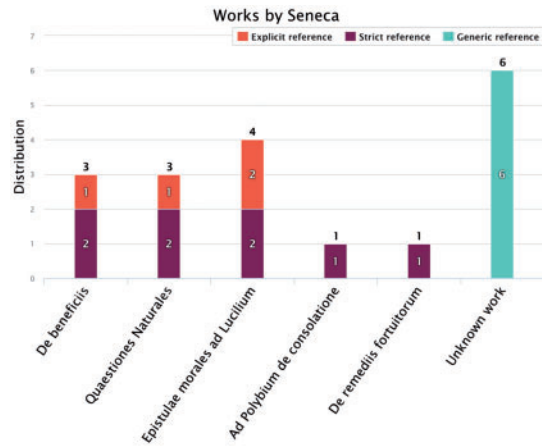


Fig. 16 The distribution of the three types of reference for the Seneca's works

improving and making more efficient the scholars' searches. Eventually, the digitization of the knowledge about primary sources and its visualization in the form of charts and tables allow the scholars to have a complete overview of the data. This simplifies the study on the evolution of Dante's cultural

background in time, also related to the diverse stages of his biography.

Summarizing, we have developed a methodology and a suite of tools to represent the knowledge about primary sources cited in literary texts. We have created a semantic model and an ontology

for representing such knowledge, a semi-automatic tool to populate the ontology, and a Web application that allows exploring the knowledge base. Our ontology, methodology, and tools are not created *ad hoc* for being exclusively applied to Dante's works, but they are easily reusable in order to represent the knowledge about the primary sources of other authors of the Italian and international literature. Furthermore, they could also be applied to other fields where it is important to represent primary sources, for instance scientific literature.

As future works, we envision: (1) further developing to the ontology in order to make it more adaptable to different kinds of textual works; (2) refining the population tool in order to make it more efficient, e.g. by making it easier for the user to import URIs from other repositories such as Wikidata²⁸ or VIAF²⁹; (3) building new visualizations to present the extracted data to the user from different points of view; (4) updating the software architecture to adopt newer technologies, such as JavaServer Faces. We also believe that our method could lead to the creation of entirely new applications, such as a new generation of digital critical editions built on top of structured knowledge.

References

- Alighieri, D.** (1980). *Vita nuova*. In De Robertis, D. (ed.). Milano, Napoli: Ricciardi.
- Alighieri, D.** (2011a). De vulgari eloquentia. In Giunta, C., Gorni, G., and Tavoni, M. (eds), *Opere. Vol. I: Rime, Vita Nova, De vulgari eloquentia*. Milano: Mondadori.
- Alighieri, D.** (2011b). Vita Nova. In Giunta, C., Gorni, G., and Tavoni, M. (eds), *Opere. Vol. I: Rime, Vita Nova, De vulgari eloquentia*. Milano: Mondadori.
- Alighieri, D.** (2011c). Rime. In Giunta, C., Gorni, G., and Tavoni, M. (eds), *Opere. Vol. I: Rime, Vita Nova, De vulgari eloquentia*. Milano: Mondadori.
- Alighieri, D.** (2014a). Convivio. In Fioravanti, G., Giunta, C., Quaglion, D., Villa, C., and Albanese, G. (eds), *Opere. Vol. II: Convivio, Monarchia, Epistole, Egloghe*. Milano: Mondadori.
- Alighieri, D.** (2014b). Monarchia. In Fioravanti, G., Giunta, C., Quaglion, D., Villa, C., and Albanese, G. (eds), *Opere. Vol. II: Convivio, Monarchia, Epistole, Egloghe*. Milano: Mondadori.
- Altman, M.** (2006). *An overview of digital libraries: issues and trends*. Cambridge: Academic Research Microsoft, pp. 1–10.
- Bartalesi, V. and Meghini, C.** (2017) Using an ontology for representing the knowledge on literary texts: The Dante Alighieri case study. *Semantic Web Journal*, 8(3): 385–94.
- Bartalesi, V., Meghini, C., Metilli, D., and Andriani, P.** (2016). Usability evaluation of the digital library DanteSources. In *Proceedings of International Conference on Human-Computer Interaction*, July 2016. LNCS, Springer International Publishing, pp. 191–203.
- Constantin, A., Peroni, S., Pettifer, S., Shotton, D., and Vitali, F.** (2016). The document components ontology (DoCO). *Semantic Web*, 7(2): 167–81.
- Dalton, M. S. and Charnigo, L.** (2004). Historians and their information sources. *College and Research Libraries*, 65(5): 400–25.
- Doerr, M.** (2002). The CIDOC CRM: an ontological approach to semantic interoperability of metadata. *AI Magazine, Special Issue*, 24(3): 75–92.
- Erling, O. and Mikhailov, I.** (2009). RDF support in the Virtuoso DBMS. In: *Networked Knowledge-Networked Media*. Berlin, Heidelberg: Springer, pp. 7–24.
- Gehrke, S., Frunzeanu, E., Charbonnier, P., and Muffat, M.** (2015). Biblissima's Prototype on Medieval Manuscript Illuminations and their Context. In *Proceedings of the First International Workshop on Semantic Web for Scientific Heritage at the 12th ESWC 2015 Conference, June 2015, Portorož, Slovenia*. Aachen, Germany: CEUR Workshop Proceedings, pp. 43–8.
- Gibert, Z.** (2015). *Jenseitsvisionen im islamischen Orient und christlichen Okzident: Jenseitsreisen im Islam und Dantes Göttliche Komödie im Vergleich*. Degree thesis in Languages, Georg-August-Universität Göttingen.
- González-Blanco, E., Seláf, L., Del Rio Riande, M.G., Martínez Cantón, C.I., and Martos Pérez, M.D.** (2014). Building a metrical ontology as a model to link digital poetic repertoires. In *Proceedings of the Digital Humanities Conference*. <http://dharchive.org/paper/DH2014/Paper-674.xml> (accessed 12 July 2017).
- Hellqvist, B.** (2010). Referencing in the humanities and its implications for citation analysis. *Journal of the American Society for Information Science and Technology*, 61(2): 310–18.
- Hendler, J.** (2012). *A primer on machine readability for online documents and data*. <https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data> (accessed 12 July 2017).

- Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, J., and Mäkelä, E.** (2016). Publishing second world war history as linked data events on the semantic web. In *Proceedings of the Digital Humanities Conference 2016*, July 2016. Cham: Springer, pp. 571–3.
- Hollander, R.** (1989). The Dartmouth Dante Project. *Quaderni d'italianistica*, **10**(1–2): 287–98.
- Hollander, R.** (2013). The Princeton Dante Project. *Humanist Studies and the Digital Age*, **3**(1): 53–9.
- Lana, M., Ciotti, F., Magro, D., Peroni, S., Tomasi, F., and Vitali, F.** (2014). Annotating texts with ontologies, from geography to persons and events. In *Poster at Digital Humanities Conference 2014*. <http://dharchive.org/paper/DH2014/Poster-136.xml> (accessed 30 January 2018).
- Malkmus, D.** (2010). “Old stuff” for new teaching methods: outreach to history faculty teaching with primary sources. *Libraries and the Academy*, **10**(4): 413–35.
- Manola, F., Miller, E., and McBride, B.** (2004). *RDF primer*. W3C Recommendation, vol. 10(1-107): 6. <https://www.w3.org/TR/rdf-primer/> (accessed 30 January 2018).
- Miles, A., Matthews, B., Wilson, M., and Brickley, D.** (2005). SKOS core: simple knowledge organisation for the web. In *Proceedings of International Conference on Dublin Core and Metadata Applications*, September 2005. pp. 3–10.
- Mishra, R. B. and Kumar, S.** (2011). Semantic web reasoners and languages. *Artificial Intelligence Review*, **35**(4): 339–68.
- Mussati, V.** (2014). *Le fonti della Vita Nuova. Trattamento informatico e formazione della “biblioteca di Dante”*. Degree thesis in Literature, University of Pisa.
- Nicolucci, F. and D’Andrea, A.** (2006). An ontology for 3D cultural objects. In *Proceedings of the 7th International conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, October 2006. Aire-la-Ville, Switzerland: Eurographics Association, pp. 203–10.
- Parker, D.** (2001). The world of Dante: a hypermedia archive for the study of the Inferno. *Literary and Linguistic Computing*, **16**(3), 287–97.
- Peroni, S. and Shotton, D.** (2012). FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Web Semantics: Science, Services and Agents on the World Wide Web*, **17**: 33–43.
- Prud’Hommeaux, E. and Seaborne, A.** (2008). *SPARQL query language for RDF*, vol. 15. W3C Recommendation. <https://www.w3.org/TR/rdf-sparql-query/> (accessed 30 January 2018).
- Riva, P., Doerr, M., and Zumer, M.** (2008). FRBRoo: enabling a common view of information from memory institutions. In *Proceedings of World Library and Information Congress: 74th IFLA General Conference and Council*, August 2008. http://archive.ifla.org/IV/ifla74/papers/156-Riva-Doerr_Zumer-en.pdf (accessed 30 January 2018).
- Sanderson, R., Ciccarese, P., and Van de Sompel, H.** (2013). Designing the W3C open annotation data model. In: *Proceedings of the 5th Annual ACM Web Science Conference*, May 2013. New York, NY, USA: ACM, pp. 366–75.
- Shotton, D.** (2010). CiTO, the citation typing ontology. *Journal of Biomedical Semantics*, **1**(Suppl 1): S6.
- Van Peteghem, J.** (2015). Digital readers of allusive texts: ovidian intertextuality in the ‘Commedia’ and the digital concordance on ‘Intertextual Dante’. *Humanist Studies and the Digital Age*, **4**(1): 39–59.
- Wagner, A. and Caesar, I.** (2016). The School of Salamanca on the Semantic Web. In *Proceedings of Digital Humanities Conference 2016*, July 2016. pp. 704–5. <http://dh2016.adho.org/static/data-copy/385.html> (accessed 30 January 2018).
- World Wide Web Consortium.** (2014). *RDF 1.1 concepts and abstract syntax*. <https://www.w3.org/TR/rdf11-concepts/> (accessed 30 January 2018).

Notes

- 1 Up to now, two theses have been written using DanteSources: [Mussati, 2014](#) and [Gibert, 2015](#).
- 2 <https://www.gnu.org/licenses/gpl-3.0.en.html>
- 3 <http://dante.dartmouth.edu>
- 4 <http://dantelab.dartmouth.edu>
- 5 <http://etcweb.princeton.edu/dante/>
- 6 <http://www.worldofdante.org>
- 7 <http://digitaldante.columbia.edu>
- 8 <http://danteonline.it>
- 9 <http://www.perunaenciclopediadantescadigitale.eu:8080/dantesearch/>
- 10 <http://perunaenciclopediadantescadigitale.eu/istidama/>
- 11 For the full list of the analysed ontologies see [Bartalesi and Meghini \(2017\)](#)

- 12 <http://dublincore.org/documents/dces/>
- 13 <http://www.foaf-project.org>
- 14 <http://vocab.org/bio/>
- 15 <http://vocab.org/relationship/>
- 16 <http://www.sparontologies.net>
- 17 <http://perunaenciclopediadantescadigitale.eu/dante-sources/schema/dante.rdf>
- 18 <https://jquery.com>
- 19 <https://jqueryui.com>
- 20 <https://www.w3.org/TR/IndexedDB-2/>
- 21 <http://dbpedia.org>
- 22 <http://thes.bncf.firenze.sbn.it>
- 23 <https://jena.apache.org/index.html>
- 24 DanteSources was the winner of the Digital Humanities Awards 2015 in the category 'Best DH Tool or Suite of Tools', <http://dhawards.org/dhawards2015/>
- 25 <http://www.highcharts.com>
- 26 <http://www.w3.org/2013/05/lcsv-charter.html>
- 27 http://link.springer.com/chapter/10.1007/978-3-319-39513-5_18
- 28 <https://wikidata.org>
- 29 <https://viaf.org>