Using an Ontology for Representing the Knowledge on Literary Texts: the Dante Alighieri Case Study

Valentina Bartalesi [a,*] and Carlo Meghini [a]

[a] *Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI) – CNR, via G. Moruzzi 1, 56127, Pisa, Italy*
E-mail: {valentina.bartalesi,carlo.meghini}@isti.cnr.it

Abstract. This paper describes a digital library developed within the "Towards a Digital Dante Encyclopaedia" project, a three years Italian National Research Project that aims at building services supporting scholars in creating, evolving and consulting a digital encyclopaedia of Dante Alighieri's works. The digital library is based on a knowledge base storing knowledge on the primary sources that Dante refers to in his works, i.e. the works of other authors Dante refers to in his texts. At present, this information is scattered on many paper books, making it difficult to systematically overview the cultural background of Dante and to obtain a well-founded perception of how this background was gradually set up in time. The same applies also to other authors, therefore the applicability of our work extends well beyond the specific author we are considering in our project. The digital library that we are building is based on an ontology for representing the knowledge on one author's works and on the primary sources embedded in the commentaries to these works. Following this approach, a semantic network of Dante's works and of references to primary sources of these works was created. Furthermore, a web application allowing users to explore the semantic network in various ways and to visualize statistical information about the references as charts and tables was developed.

Keywords: Ontology, Semantic Web, Digital Libraries, Digital Humanities

1. Introduction

Literary works, especially those written centuries ago, hold information which, for several reasons, remains to modern readers largely implicit at best, and outright inaccessible at worst. It is the task of specialized scholars in the humanities to help readers understand the implicit knowledge contained in literary works by carefully analyzing the text of such works. This typically happens in the so-called commentaries, modern works of interpretive research that decode, often line by line, several kinds of knowledge encoded in literary texts, and report it explicitly (e.g. [6]). One important kind of knowledge concerns the primary sources literary texts refer to. In our case study, primary sources are the works of other authors (eg.Aristoteles) referred to in his texts by Dante Alighieri (1265-1321) - the major Italian poet of the late Middle Ages. In their commentaries, scholars typically express knowledge about the primary sources referred to by Dante in natural language. Using natural language limits scholars in their advances insofar as it prevents automatic inferences of new information that may be useful for their studies. For instance, in the case of references to primary sources, these inferences may concern the total amount of references to a certain

---

*Corresponding author. E-mail: valentina.bartalesi@isti.cnr.it

work or to an author. Such knowledge may be useful to reconstruct the evolution of the author's cultural background. The goal of our work is to overcome this problem by creating an ontology providing a formal representation of the knowledge on an author's primary sources. 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). Up to now, an ontology for representing knowledge about primary sources of literary texts has not been developed yet and there are not projects focussed on giving a logical representation of this knowledge. Furthermore, our ontology extends well beyond the specific author we are considering in our project.

The work we present in this paper is part of the "Towards a Digital Dante Encyclopaedia", an Italian National Research project supporting scholars in formally expressing the knowledge about primary sources present in Dante's works and more in general in literary texts. Moreover, the project aims at developing a system to automatically discover complex semantic relations among different parts of the text or between the text and its related resources. This work presents the ontology developed on the basis of a previous preliminary study [1] to represent knowledge on the references to primary sources included in literary texts and in their commentaries, with a specific focus on Dante's works. In particular, after a deep analysis of the related ontologies, as reported in [1], in comparison with the preliminary study, we report here the result of the refine process of our ontology and we describe the web application that we developed on the basis of the final version of this ontology. We report the process that we followed to extract the relevant knowledge from natural language commentaries to Dante's works and to express this knowledge with a formal logic semantic representation. We started from an Excel file provided by an authoritative Italian scholar, reporting commentaries and references to primary sources embedded in a Dante's work. On the basis of the analysis of the contents of the Excel file, we developed a conceptualization, i.e. a set of classes and properties underlying the knowledge reported in the spreadsheet. Next, we reviewed existing ontologies in the Digital Libraries domain in order to create a vocabulary to express the identified classes and properties. We adopted several terms (i.e. the names and the corresponding definitions of classes and properties) from these ontologies to maximize the interoperability of our representation. Finally, we added our own classes and properties to

represent the terms that we did not find in the analyzed ontologies. The resulting ontology is expressed in RDF/S[1]. Once the ontology had been defined, we populated it with the knowledge included in the Excel file but also with the knowledge included in other commentaries to Dante's works. We stored the resulting RDF graph into a Virtuoso triple store [5]. On top of this graph, we developed a web application allowing a number of data visualizations of Dante's works and primary sources. Such visualizations are expected to provide useful knowledge to the scholars that work on the creation of a complete encyclopaedia of Dante's works.

The paper is structured as follows: Section 2 reports the analysis and the evaluation of the initial representation requirements. In Section 3, the review of the existing ontologies is reported, while Section 4 presents the classes and properties of these ontologies that we re-used. It also introduces the classes and properties that we added to capture missing aspects. Section 5 presents the ontology population process. Section 6 describes the web application that we developed on the basis of our ontology. Finally, in Section 7 we report our conclusive remarks.

2. Requirements

The references to primary sources are contained in commentaries to Dante's works written by authoritative scholars. Each commentary applies to a fragment of a work by Dante and asserts, among other things, that the Dante's fragment makes a reference to a work of another author. The reference may be to a specific fragment or to the totality of the primary source. Our first task was to understand how formally representing the described knowledge to implement automatic services on top of it. In order to achieve this goal, we started with the study of a spreadsheet, provided by an authoritative Italian scholar (Mirko Tavoni, professor of History of the Italian Language and Italian Linguistics at the University of Pisa) reporting the relevant knowledge relative to Convivio [6], a philosophical essay written by Dante between 1304 and 1307. Each row of the spreadsheet reports (see Figure 1):

- the number of the book, of the chapter (e.g. 1.01 indicates the first book and the first chapter of

---

Convivio) and of the paragraph to which the commentary is applied;

- the Dante's text fragment which the commentary is applied to (e.g., *"Sì come dice lo filosofo nel principio della Prima Filosofia"*/ As the Philosopher says at the beginning of the First Philosophy);
- a fragment of the commentary to the Dante's text;
- primary source, structured as:

  * author (e.g. Aristotle);
  * title (e.g. Metaphysics);
  * thematic area (e.g. Aristotelianism).

	A	B	C	D	E	F	G
1	Book - Chapter	Paragraph	Text	Note	Author	Work	Thematic area
2	1.01	1	Si come dice lo Filosofo nel principio della Prima Filosofia	le parole con cui si apre la Metafisica di Aristotele (I 1, 980a 21). Il sintagma "prima Filosofia" è già presente in Aristotele per distinguere questa scienza dalla Fisica (cfr. Metaph. VI 1, 1026 a 27-30).	Aristotele	Metafisica	Aristotelismo

Fig. 1. A row of the spreadsheet provided by the scholar.

After discussing with the scholar, two other pieces of knowledge were added to each row of the spreadsheet:

1. the entire text of the commentary, thus making available to the scholars the context of the reference;
2. the kind of the reference, which may be:

   - *explicit,* if the reference is explicitly made by Dante, as in *"Sì come dice lo Filosofo nel principio della Prima Filosofia"*/As the Philosopher says at the beginning of the First Philosophy;
   - *strict,* if the reference is indicated by a scholar, and refers to a specific work as in *"SI MANUCA: il pane degli angeli, nella tradizione veterotestamentaria e' la manna (cfr. Ps. 77, 25 "Panem angelorum manducavit homo"*/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).

The organization of knowledge on primary sources in thematic areas can be useful for scholars who can investigate the evolution in time of Dante's cultural background using this more general information level. The list of thematic areas that the scholar gave us is based on the knowledge of the scholar himself and not on standard resources. Our aim was instead to reuse available and standard resources for interoperability, thus we investigated the possibility of using a more widely known vocabulary. The choice finally was to draw the thematic areas from *Nuovo Soggettario*[2]. The Nuovo Soggettario, edited by the National Central Library of Florence, is a subject indexing tool for various types of information resources. It has been developed in compliance with the International Federation of Library Associations and Institutions (IFLA) recommendations[3] and represents an important asset for the Italian scholar community. The fact that an RDF version of this resource is freely available online, expressed through the SKOS [10] vocabulary, also played a positive role. Indeed, the expression in RDF format made it easy to insert Soggettario Nazionale into our knowledge base.

Besides Convivio, we received knowledge about references to other Dante's works along with their commentaries in textual format from a group of three scholars of the University of Pisa directed by Professor Tavoni. These works are: Monarchia [12], De vulgari eloquentia [16], Vita Nova [7], Vita Nuova [8]. These commentaries are currently accessible in paper format. They are protected by copyright, thus it was not possible for us to show to the users the entire text of the commentaries on our web application. Indeed, we could visualize only little portions of the commentaries annotated by the scholars. After a deep analysis of the commentaries, developed with the support of scholars, we verified that Convivio's knowledge structure reported in the Excel file provided by Professor Tavoni was compliant with the knowledge structure of the commentaries on the other Dante's works provided by the three scholars (that is the structure of the excel file of Convivo also worked for the other Dante's works). Moreover, we decided to include the whole text of the above mentioned works from Dante in our digital library, modelled in a semantic way (see next Section) according to their structure. We adopted this strategy to provide the users with as many context references as possible.

3. Reused Ontologies

The analysis described in the previous section allowed us obtaining a conceptualization of our subject

---

matter, which is a list of the relevant classes and properties comprising the slice of reality under study. As next step, we devised an ontology providing names for such classes and properties as well as axioms for expressing their meaning in a formal way. In order to carry out this step, we investigated the scientific literature and the existing standards in the Digital Libraries field [17], whether official or *de facto*. In particular, we took into account the ontologies addressing issues related to the textual domain. We did not find any ontology able to represent all the classes and properties we identified analyzing the commentaries on Dante's works. On the other hand, we could find ontologies offering classes, properties and axioms for some classes and properties. A brief account of these ontologies and of the re-used classes and properties is given below.

FRBR and FRBRoo. The Functional Requirements for Bibliographic Records model (FRBR) is a conceptual entity-relation model developed by the International Federation of Library Associations and Institutions (IFLA). FRBR includes a description of the conceptual model (the entities, properties, and attributes or metadata), a bibliographic record for all types of materials, and user tasks associated with the bibliographic resources described in catalogues, bibliographies, and other bibliographic tools [9]. FRBR contains many classes useful to describe a textual domain, e.g. the class Work, to represent an intellectual creation, or the class Expression whose members are a realization of a single work, usually in a physical form. FRBRoo is essentially the FRBR ontology expressed in an object-oriented form, which is more compatible with that of the CIDOC-CRM. The CIDOC Conceptual Reference Model (CRM) is an ontology that provides definitions and a formal structure for representing the implicit and explicit knowledge included in the cultural heritage documentation [3]. Since CIDOC-CRM is an ISO standard, it has been considered as a common vocabulary to represent information published on cultural heritage by archives, museums or libraries, and to map it to an equivalent digital representation [2] [4] [14]. As explained in the next Section, we used several classes and properties from FRBR and FRBRoo to represent our knowledge.

Dublin Core. The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for usage in resource description. These fifteen elements of the Dublin Core are part of a larger set of metadata vocabularies and technical specifications developed by the Dublin Core Metadata Initiative. We analysed the Dublin Core because its metadata can be used to de-

scribe a full range of web resources (video, images, web pages, etc.) but also of physical resources, such as books and objects like artworks. As explained in the next Section, we reused several classes from the Dublin Core vocabulary.

SKOS. Simple Knowledge Organization System (SKOS) 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 [10]. In particular, such vocabulary contains the term Concept that we reused to describe the thematic area of a work or a general concept like Comments to Aristotle's works.

FOAF. Friend of a friend (FOAF)[4] is a machine-readable ontology linking people and information using the Web. In particular, FOAF describes persons, their activities and their properties to other people and objects. The FOAF ontology was useful to represent the knowledge on the author of a work.

The SPAR (Semantic Publishing and Referencing) Ontologies[5] are a suite of complementary ontologies to describe all the aspects of bibliographic publications as comprehensive machine-readable RDF metadata. Since we wanted to describe the text structure, the textual publication type and the bibliographic citations, we specifically analyzed and reused some classes and properties from the following three ontologies, developed within this project:

- DoCO, Document Components Ontology[6], allows describing document components in RDF. Such components are both structural (e.g. block, chapter, heading, paragraph, section) and rhetorical (e.g. Abstract, Introduction, Results, Discussion, Conclusions, Acknowledgements, Bibliography).
- FaBiO, the FRBR-aligned Bibliographic Ontology [11], is an ontology 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.
- CiTO, the Citation Typing Ontology [15], is an ontology for the characterization of bibliographic

---

citations, both factually and rhetorically and for their publication on the Web.

Finally, we also reused classes and properties from the *Open Annotation Core Data Model* that provides classes and properties to represent textual commentaries and to create explicit properties between such commentaries and their related resources. The Open Annotation Core Data Model [11] 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. Within the Open Annotation Model, *"an annotation is considered to be a set of connected resources, typically including a body and a target, and it reveals that the body is related to the target"*[7]. Such ontology is especially interesting for our aims because it defines and represents the structure of the commentaries in detail. Indeed, as reported in the next Section, we reused many classes and properties from the OA vocabulary.

4. The Ontology

As previously mentioned, one of the aims of our project is to develop an ontology for expressing several kinds of knowledge on the structure, the content and the context of Dante's works and, more in general, of literary texts. Thus, we reused and extended some existing vocabularies of relevance to textual descriptions. We selected several ontologies [1] in order to identify classes and properties we could reuse. After a deep analysis of these ontologies, we selected those classes and properties from the ontologies described in the previous section that were useful for representing the specific kind of knowledge that we addressed in our study. In summary, almost 90% of the classes and properties of our ontology comes from standard ontologies. We reused classes and properties 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. We adopted this approach in order to avoid unwanted or unspecified interactions in semantics. In this section, the classes and properties we re-used and the new ones are reported in detail. As a notational convention, in what follows we use prefixed qnames (e.g., efrbroo:hasFragment) to denote terms re-used from other vocabularies, whereas we use local names (e.g. hasCitingFragment) to denote

terms of our own ontology. We started from the analysis of the CIDOC-CRM and FRBR vocabularies which are particularly relevant in terms of the transmission of cultural information and ideas expressed in written works. In particular, we considered the following FRBRoo classes as appropriate to represent the related aspects of our textual knowledge:

1. efrbroo:Work corresponding to a specific work, cited by the commentaries, (e.g., Metaphysics) without reference to a specific edition;
2. efrbroo:Expression corresponding to a precise edition of the Dante's work;
3. efrbroo:ExpressionFragment corresponding to (i) a fragment of a Dante's work a commentary refers to and also (ii) to a fragment of a commentary that refers to a specific primary source. We used this class instead of Self-Contained Expression because scholars annotate fragments of Dante's texts that were not intended by Dante himself to convey the concept of the work, but only part of it. For example, in our ontology ExpressionFragment also represents a single word written by Dante that the scholar has annotated. ExpressionFragment represents also a little piece of text of a commentary that cites a primary source, for example the title of a primary source reported by a scholar.

Each instance of efrbroo:ExpressionFragment is in textual form. In order to model the textual content of an ExpressionFragment instance, we followed the recommendations of the W3C's Content in RDF[8], and we used two specific subproperties we defined:

1. a subproperty of dc:format, called efFormat, taking as values MIME types. It gives the media type of the instance. For example, it allows distinguishing between embedded content in plain text versus that encoded in HTML;
2. a subproperty of cnt:chars[9], called efChars, taking literals as values. It represents the sequence of the characters of the content.

Furthermore, we represented the textual nature of an ExpressionFragment instance using two classes:

1. dctypes:Text class, to denote the fact that the instance represents a resource primarily intended to be read;

---

2. cnt:ContentAsText class represents the textual content of the instance.

Figure 2 shows the textual content representation in an RDF graph. In this and the following graphs, we have used the standard RDF graphical notation: URIs are written as qnames and are enclosed in ovals or used as arrow labels, whereas literals are enclosed in rectangles. Qnames without prefixes (e.g., ":ef1") use the base namespace, which is left undefined.
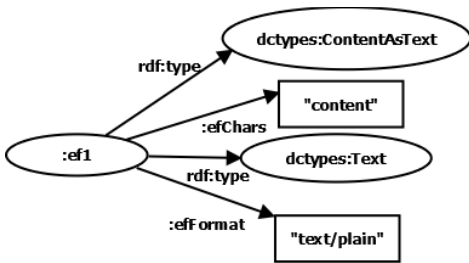

Fig. 2. Textual content of the expression fragment ef1.

In addition to the classes and properties reported so far, we used the concept of Selector from the Open Annotation Core Data Model, in order to identify in the most precise way a specific fragment of a text. A selector is a specifier that describes how to determine the segment of interest within a specified text. The text from which the segment was extracted is named Source in the Open Annotation Model. The nature of the Selector depends on the type of representation for which the segment is conveyed. In our case, we used the oa:TextPositionSelector that describes a range of text based on its start and end positions within a paragraph or a verse. The paragraph, or the verse, plays therefore the role of Source of the Selector. The connection between a fragment of a work and its Source is established using the property oa:hasSource. In Figure 3 we report the representation of the Selector and Source terms in our RDF graph.

To describe the structure of the resource from which the fragment is extracted we reused three classes of Fabio and Doco ontologies:

1. doco:Paragraph, which represents a self-contained unit of discourse that deals with a particular point or idea;
2. fabio:Poem, that is an artistic work written with an intensity or beauty of language more characteristic of poetry than of prose;
3. doco:Line, that is a line in poetry, i.e. a unit of language into which a poem is divided;
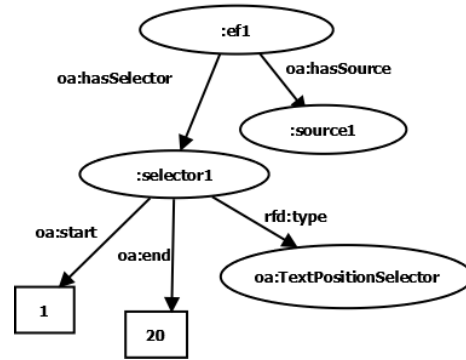

Fig. 3. Selector and Source.

4. doco:Chapter, which represents a principle division of the body matter of a large document;
5. fabio:Book, which defines a non-serial document that is complete in one volume or a designated finite number of volumes.

The model of a work structured in paragraphs, chapters and books is reported in Figure 4. To relate the paragraph with the related chapter and book, we defined three subproperties of the frbr:isPartOf: (i) isParagraphOf, (ii) isChapterOf, and (iii) isBookOf. In the case of Dante's poems (e.g. Rime) in our model the class :source1 in Figure 3 corresponds to a verse of a poem (represented by the class doco:Line and by the class fabio:Poem, respectively) in which the instance of the Expression Fragment class is included. On the other side, for Divina Commedia, we plan to define two specific classes: the Source is a verse, the verse is part of a *canto* that is a part of a *cantica*. We did not report examples from Rime and Divina Commedia because these works have not been inserted in our knowledge base yet.
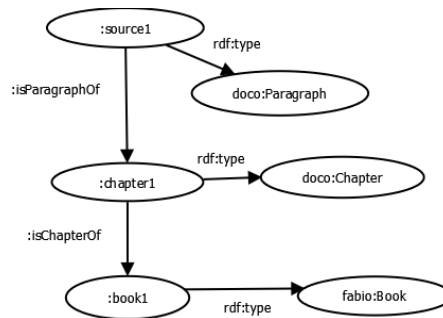

Fig. 4. The structure of Convivio.

Since the Open Annotation Model is the standard *de facto* adopted by the WC3, we decided to be compliant

with this ontology and to guarantee the interoperability with it. For these reasons, we used the definition of the commentary reported by the OA: *"an Annotation has a single Body, which is the comment or other descriptive resource, and a single Target that the Body is somehow about"*[10]. In order to link the annotation with its components, we used the oa:hasBody and oa:hasTarget properties but we also defined a new property :hasCommentary to directly link the target text with the commentary. These properties are reported in Figure 5.
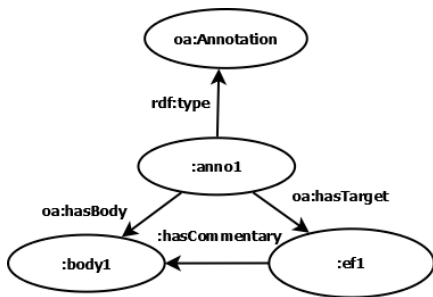

Fig. 5. Relation between the text and the commentary.

We modelled the content of a commentary as an instance of the class oa:Body. Figure 6 shows the representation of a commentary having a textual content, in the same way as the Expression Fragment.
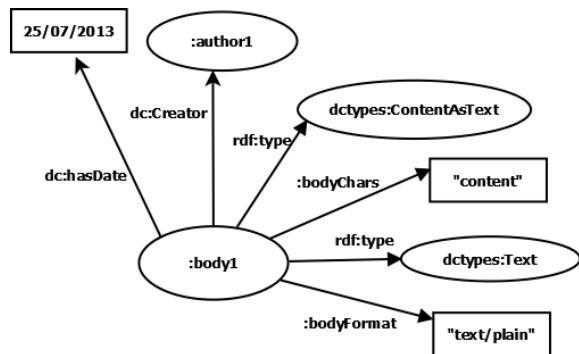

Fig. 6. Representation of a commentary.

A commentary can have more than one fragment, where a fragment in a commentary is the portion of text within the commentary that cites a specific primary source. The frbroo:ExpressionFragment is also used to represent a fragment of commentary. A Selector is used to identify the exact portion of the commentary that is part of the fragment.

In order to model the textual content of a fragment of a commentary and to distinguish it from a fragment of Dante's text, we defined two sub properties:

1. a subproperty of dc:format, named bodyFormat, taking as values MIME types. It gives the media type of the body.
2. a subproperty of cnt:chars, named bodyChars, which represents the character sequence of the body.

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 the hasCitingFragment as a subproperty of the frbroo:hasFragment.

The fragment of commentary refers to a primary source through three subproperties of the cito:cites property we defined: (i) citesAsExplicitCitation, (ii) citesAsStrictCitation, (iii) citesAsGeneralCitation.
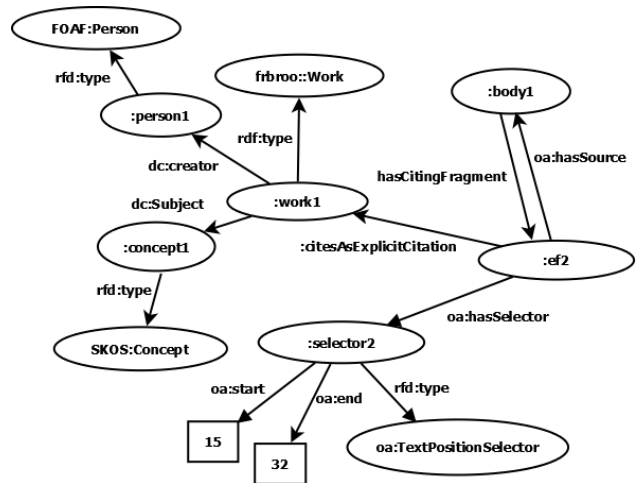

Fig. 7. The representation of the knowledge on a commentary.

As reported above, in order to describe a work referred to by the commentary, we used the class frbroo:Work. In our schema a work has an author and a thematic area (e.g. Rhetoric, Astronomy, Aristotelianism, etc.). We represented this additional knowledge using the classes foaf:Person and skos:Concept, respectively. To link the frbroo:Work class with the other two classes we used two Dublin Core properties: dc:creator, and dc:subject, respectively. In Figure 7 we present the representation of the knowledge on a commentary. Currently, we are working on translating the ontology from RDF language to OWL in order to add axioms that allow to infer new knowledge.

---

[10]http://www.w3.org/ns/oa

5. Ontology Population

A Java tool to transform our original data set of Dante's works in DOC format into an annotated corpus in RDF format, using a text processing approach, was developed. The tool allows automatically extracting the following pieces of information from the DOC and putting them into a CSV file:

- the number of the book, the chapter and the paragraph of the text fragment to which the commentary refers to;
- the text fragment to which the commentary applies (e.g., *"Sì come dice lo filosofo nel principio della Prima Filosofia"*/As the Philosopher says at the beginning of the First Philosophy);
- the entire text of the commentary;

After this first step, the following knowledge has been manually added to CSV by the three scholars working with Professor Tavoni:

- author of primary source (e.g. Aristotle);
- title of primary source (e.g. Metaphysics, the title is reported preferably in Latin);
- thematic area of the primary source (e.g. Aristotelianism, extract from the Nuovo Soggettario, as reported in Section 2);
- fragment of the text of the commentary citing a primary source (e.g.*"Queste sono le parole con cui si apre la Metafisica di Aristotele"*/These are the words that open the Aristotle's Metaphisics);
- kind of reference (explicit, strict or generic reference)

At the same time, we developed an XML schema (XSD), compliant with the ontology, in order to represent our knowledge in a pre-axiomatic form. The schema defines the permitted elements, the related data and the hierarchy of elements. Then, our tool automatically translates the CSV into an XML file, following the XML schema. Finally, we extended the tool in order to translate the XML into an RDF/XML file, which we stored into Virtuoso along with Soggettario Nazionale in RDF format.

6. Using the Ontology: the Web Application

In order to extract and display the knowledge stored in our digital library, we developed a web application to support scholars in writing a complete encyclopaedia of Dante's works. The application extracts knowledge by making SPARQL queries to the ontology. It is able to produce column bar charts in order to show the data about primary sources cited by Dante. We used the Highcharts[11] JavaScript library to implement these charts. This library allows exporting the charts in various well-known formats: PDF, PNG, JPEG, SVG. Furthermore, we implemented an additional JavaScript function allowing users to automatically export and download all the data extracted by the application in CSV format[12]. This feature is particularly important since it allows scholars to obtain and manage raw data, in order to apply further data analyses.

Up to now, six different predefined SPARQL queries are available to extract data for a representation using column bar charts. They can be distinguished into two different groups. The first group includes three queries. For these queries the user can choose one among Dante's works (or even all his works) and, in addition, a specific sub part of the work (e.g. a book). The queries return data regarding the distribution of the works, the authors and the thematic areas cited by Dante. Additional information about primary sources, authors and thematic areas are available by clicking on their names. The chart in Fig. 8 shows the distribution of some primary sources cited in the Book 1 of Convivio.
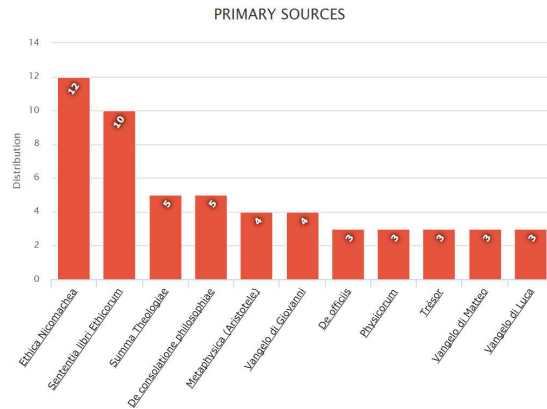

Fig. 8. The chart represents the distribution of some primary sources cited in the Book 1 of Convivio.

The three queries of the second group return several charts reporting the distribution of a particular primary source, cited author or thematic area chosen by the user. The data regarding the distribution are available not only for an entire Dante's work, like Convivio,

---

but also for its sub parts as books, chapters and poems. For example, the chart in Fig. 9 shows the distribution of the cited author Aristotle in the Dante's works stored in our knowledge base.
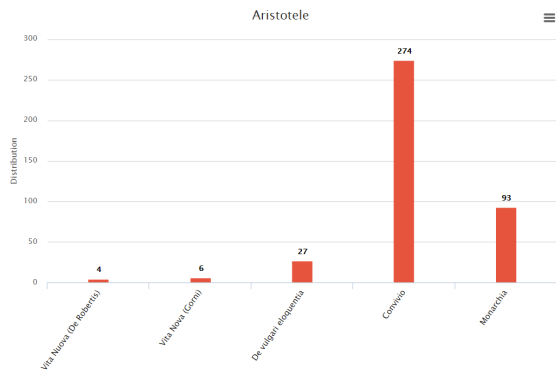

Fig. 9. The chart shows the distribution of the cited author Aristotle in the Dante's works stored in our knowledge base.

The application is freely available on the web[13].

7. Conclusions and Future Work

We have presented an ontology representing the knowledge on primary sources included in literary texts and in their commentaries, with a specific attention to Dante Alighieri's works. We have described the complete process to create a formal semantic representation of this knowledge included in commentaries to Dante's works. We have started from an Excel file containing commentaries and references to primary sources. We have analyzed this file and we have developed a conceptualization, i.e. a set of classes and properties underlying the knowledge reported in the spreadsheet. On the basis of the conceptualization we have created an ontology, expressed in RDF/S language. This ontology re-uses classes and properties from other existing vocabularies and also embeds newly defined terms. We have populated the ontology with the knowledge included in several commentaries to Dante's works. On top of the RDF graph, we have developed a web application that visualizes the knowledge in the form of charts and tables. Gathering the current information on the primary sources used by Dante in his works, and having this information available in digital format, improve and make more efficient

the research of primary sources by scholars. Having all this 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 (e.g., by type of source, or by author, by work, etc.), allows exploring the dynamics of the multi-faceted culture of Dante in relation to the diverse and often conflicting stages of his biography. Such visualizations are useful to the scholars who work on the creation of a complete encyclopaedia of Dante's works. From our description, it emerges that our ontology and our web application can be applied also to other authors and commentaries. As future work, we plan to map our ontological classes onto the TEI [18] elements in order to import TEI documents in our knowledge base. Furthermore, we are working to evaluate the usability of our web application by means of user tests.

8. Acknowledgements

References

[1] V. Bartalesi, C. Meghini, E. Locuratolo, L. Versienti. 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 Press, New York, 2013.

[2] C. Binding, K. May, D. Tudhope. Semantic interoperability in archaeological datasets: Data mapping and extraction via the CIDOC-CRM. *In Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries, ECDL '08*. Springer-Verlag, Berlin, Heidelberg, 2008, pp. 280-290.

[3] M. Doerr. The CIDOC conceptual reference module: An ontological approach to semantic interoperability of metadata. AI Mag. 24 (3) 2003, pp. 75-92.

[4] O. Eide, A. Felicetti, C. E. Ore, J. Holmen. Encoding cultural heritage information for the semantic web. Procedures for data integration through CIDOC-CRM mapping. *In Proceedings of EPOCH Conference on Open Digital Cultural Heritage Systems*, 2008, pp. 1-7.

---

[13]http://perunaenciclopediadantescadigitale.eu/dantesources/

[5] O. Erling, I. Mikhailov. RDF Support in the Virtuoso DBMS. Networked Knowledge-Networked Media. Springer-Verlag, Berlin, Heidelberg, 2009, pp. 7-24.

[6] G. Fioravanti. Opere. Vol. II: Convivio, Monarchia, Epistole, Egloghe. Eds. Gianfranco Fioravanti. Mondadori, Milano, 2014.

[7] G. Gorni. Opere. Vol. I: Rime, Vita Nova, De vulgari eloquentia. Eds. Guglielmo Gorni. Mondadori, Milano, 2011.

[8] D. De Robertis. Dante Alighieri - Vita Nuova. Eds. Domenico De Robertis. Ricciardi, Milano-Napoli, 1980.

[9] O. M. Madison. The IFLA functional requirements for bibliographic records. Library resources & technical services 44 (3) 2000, pp. 153-159.

[10] A. Miles, B. Matthews, M. Wilson, D. Brickley. SKOS core: simple knowledge organisation for the web. *In Proceedings of International Conference on Dublin Core and Metadata Applications*, 2005.

[11] S. Peroni, D. Shotton. FaBiO and CiTO: ontologies for describing bibliographic resources and citations. Web Semantics: Science, Services and Agents on the World Wide Web 17 2012, pp. 33-43.

[12] D. Quaglioni. Opere. Vol. II: Convivio, Monarchia, Epistole, Egloghe. Eds. Diego Quaglioni. Mondadori, Milano, 2014.

[13] R. Sanderson, P. Ciccarese, H. Van de Sompel. Designing the W3C open annotation data model. *In Proceedings of the 5th Annual ACM Web Science Conference*, ACM, 2013, pp. 366-375.

[14] P. Sinclair, M. Addis, F. Choi, M. Doerr, P. Lewis, K. Martinez. The use of CRM core in multimedia annotation. *In Proceedings of Workshop on Semantic Web Annotations for Multimedia*, 2006.

[15] D. Shotton. CiTO, the citation typing ontology. Journal of Biomedical Semantics 1 (S-1) 2010 S6.

[16] M. Tavoni. Opere. Vol. I: Rime, Vita Nova, De vulgari eloquentia. Eds. Mirko Tavoni. Mondadori, Milano, 2011.

[17] M. Altman. An overview of digital libraries: issues and trends. Academic Research Microsoft, Cambridge, 2006, pp. 1-10.

[18] TEI Consortium. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Eds. Lou Burnard, and Syd Bauman. TEI Consortium. http://www.teic.org/Guidelines/P5/