

The Biographical Dictionary of Friulians - “Nuovo Liruti” Online

A Biographical Dictionary Based on Semantic Web and Linked Open Data

Stefano Allegrezza¹(✉) and Nicola R. Di Matteo²(✉)

¹ Università degli Studi di Udine, Udine, Italy
stefano.allegrezza@uniud.it

² Dalhousie University, Halifax, NS, Canada
nicola.dimatteo@dal.ca

Abstract. Most biographical dictionaries today are becoming more and more online based. For example, Treccani has produced a digital edition of the Italian Biographical Dictionary, which is more readily available and more up-to-date than the paper edition. In fact, the traditional edition, which is organized according to an alphabetical order, has the entries relating to the letter “A” published in 1960 and those related to the letter “M” published in 2011; therefore it is currently obsolete. The Rosi Dictionary of the Renaissance Risorgimento is fully accessible online, as well as the Biography of Women and Men of Resistance of the National Association of Partisans of Italy, the Online Biography of Protestants in Italy of the Society of Studies Valdesi and many other examples that could be made. A similar trend amongst these is the American National Biography, the Diccionario Biográfico Español, the Slovenska Biografija show, only to make some significant examples. However, these online dictionaries often don’t have an intuitive interface and often are not so “attractive” to entice the user into navigation. Moreover, semantic web tools are almost never used, but they can find their own special and interesting application in this particular field. The aim of this project was to realize not only a digital edition of the printed version of the “Biographical Dictionary of Friulians” (“New Liruti online”), but also one of the richest and structured cultural and historical information deposit on Italian web sites, thanks to the application of semantic web methodologies for the digital edition of biographical dictionaries, with the opportunity to reach a much more ambitious and potentially unlimited audience than the paper edition - as well as being one of the most important cultural initiatives within the broadest project on “Cultural Identity of Friuli (ICF)”. It is also an example of effective collaboration between humanists and computer scientists, without whom the project would not be possible.

Keywords: Biographical dictionaries · Biographies · Semantic web
RDF · Linked data

1 Introduction

Most biographical dictionaries now tend to become more and more online. For example, Treccani has produced a digital edition of *the Italian Biographical Dictionary* [1], which is more readily available and more up-to-date than the paper edition (which is organized according to an alphabetical order, with the entries relating to the letter “A” published in 1960 and those related to the letter “M” published in 2011), is currently obsolete; the *Biography of Women and Men of Resistance* [2] of the National Association of Partisans of Italy, the *Online Biography of Protestants in Italy* [3] of the Society of Studies Valdesi and many others examples that could be made. Abroad the trend is almost identical, as *the American National Biography* [4], the *Diccionario Biográfico Español* [5], the *Slovenska Biografija* [6] show, only to make some significant examples. However, these online dictionaries have often a not intuitive interface and often not so “attractive” to entice the user into navigation. Moreover, semantic web tools are almost never used, but they can find their own special and interesting application in this particular field. This project aims to be a best practice for the application of innovative methodologies in the realization of digital editions of biographical dictionaries, and also an example of collaboration between humanists and computer scientists, without which the project would not have been possible.

2 The Idea of the Project

The idea of the project started with the happy intuition of the Pio Paschini Institute for the Church’s history in Friuli, which in 2016 proposed to create a digital edition of the “New Liruti. Biographical Dictionary of the Friulians”¹ [7] in collaboration with the institutions that had promoted the printed edition. The entries already published would have been revised and integrated by about four hundred bio-bibliographic profiles of the so called “Ongaro Supplement” (made by Maiko Favaro on the basis of the eighteenth century manuscripts by Domenico Ongaro), and by the voices of Onomasticon, which was planned during the presentation of the whole work. It was therefor not a question of simply transposing the digital version of the printed edition of the “New Liruti” or of making its electronic version available (the PDF file) but something much more: a real online biographical dictionary with a captivating graphic interface with numerous functionalities that can be used both by the scholar and the simple citizen who want to deepen the history and culture of Friuli Venezia Giulia. The project was funded by the Province of Udine, the Friuli Foundation and the Archdiocese of Udine; it was promoted by the Pio Paschini Institute for the Church History in Friuli, the Patriarch Deputies for Friuli, the Historical Institute of the Ancient Book, the Friulian Philological Society and the Department

¹ Although spellings Friulian and Friulan are both used, Friulian is more used and acknowledged [9, 10].

of Humanities and Cultural Heritage of the University of Udine². The scientific directors are Cesare Scalon and Claudio Griggio while the technical director is Stefano Allegrezza. The work was entrusted to Nicola Raffaele Di Matteo. The project was officially presented on April 3, 2017, Friuli’s home day, and is now available at the address <http://www.dizionariobiograficodefriulani.it> (see Fig. 1).



Fig. 1. The home page of the Biographical Dictionary of Friulians

3 Strengths of the Project

The project has started from a preliminary phase of analysis of biographical dictionaries published on the web, both in Italy (such as the Italian Dictionary of Biography, the Rosi Dictionary of the Risorgimento Renaissance, the Biography of Women and Men of Resistance by the National Association of Partisans of Italy, the Online Biography of Protestants in Italy by the Society of Studies Valdesi, etc.) and abroad (such as the American National Biography, the Deutsche Biographie, the Slovenska Biografija and the Diccionario Biográfico

² Among the partners of the project are the Accademia San Marco in Pordenone, the Academy of Letters and Arts in Udine, the Diocesan Archives and Patriarchal Library of Udine, the State Archives of Gorizia, the State Archives of Pordenone, the State Archives of Udine, the Guarneri Civic Library of San Daniele del Friuli, the Civic Library of Pordenone, the Civic Library “Vincenzo Joppi” in Udine, the Isontina State Biological Bureau of Gorizia, the “Institute of Social and Religious History” of Gorizia, the popular University of Udine.

Español, only to make some significant examples) [15]. From this analysis we have become aware of the almost universal trend of printed biographical dictionaries to become online biographical dictionaries. The reasons behind this tendency are many and will be briefly examined in the following because they are the same that are at the basis of this project.

3.1 Continuous and Real-Time Update

Compared to the printed edition, the online edition has the undoubted advantage of providing an up-to-date and continuously updated reference. Once a new biographical entry is inserted, it will be immediately visible online. Even correcting any mistakes or denials is very easy and immediate (this is obviously not possible with the printed edition). Anyway this requires the presence of an editorial committee that continuously follows the editorial activities.

3.2 Better Usability

Usability of content that is made available online could be better than the printed edition, since it is possible to not only a sequential reading mode (such as the printed edition) but also a hypertextual reading mode (taking advantage of the links that were included in the text to highlight the most interesting links). In addition, the 2620 biographical entries, re-checked and updated where necessary, are available in multiple navigation modes: not only following an alphabetical order - as is evident - but also following chronological, geographic or thematic paths.

3.3 Unlimited Scalability

The online edition of the dictionary is based on an infrastructure that puts no limit to the variety and amount of information that can be hosted; the dictionary can be expanded with content of any kind (think, for example, about audio recordings that can be associated with a musician's biographical entry or video recordings that can be associated with a director's entry, etc.). There are basically no limits to the expansion of the online dictionary.

3.4 Possibilities of Interaction with Users

The electronic medium allows extremely varied forms of interaction with users and in some cases also suitable for content editing, as it is possible, for example, in Wikipedia. Although mechanisms of this kind are widely used with good results [8], it could be necessary to have a drafting committee to verify the content inserted by users.

As a result, it has been decided to allow users to interact with the site only by leaving comments on biographical entries. It is also possible to interact with major social media (Facebook, Twitter, Google+, Instagram, LinkedIn). In the future it will be possible to review this choice by enabling users to interact with the site by

adding materials that can enrich a given biographical entry. For example, think of an artist’s biographical entry and the likelihood that a user who has digital copies of some of his works can add them, thus enriching the biographical entry with potentially interesting material. This will require the definition of “strategies” to check and “evaluate” the content by an editorial committee.

3.5 Possibility to Carry Out Very Sophisticated Searches

The true richness of the digital edition is certainly the ability to carry out searches among the most disparate and sophisticated. Compared to the printed version, which basically allows you to search only alphabetically by browsing the 7285 pages of printed volumes one by one.³ The digital version allows you to easily and freely perform both full-text searches on all content [11], that advanced searches (by specifying the appropriate search criteria so that you can quickly get the content you want to see). To achieve this, all biographical entries have been associated with a series of metadata (place and date of birth, place and date of death, places and dates important in person’s life, profession, sex, curator of biographical voice, etc.). This allows the user to retry information about persons or the facts through different search criteria that have been defined; for example query by name and surname, date or place of birth and death, profession (jurists, literate, typographers, musicians, etc.), sex, etc. Therefore, it is possible to know which illustrious Friulians are related to a certain city or territory; which anniversaries fall into a certain year and which anniversaries to celebrate; it is possible to further refine the search in order to know which writers, poets, storytellers, philologists, filmmakers, artists, sportsmen, etc. are linked to a certain city or territory; it is also possible to carry out more targeted research by combining the various search criteria between them. It allows for answers to extremely targeted questions, such as:

- who are the illustrious people linked to the city of Palmanova and whose celebration is expected in 2018?
- what happened on the 3rd April?
- who are the illustrious women who made the Latisana city famous in the world?
- who are the Friulian athletes of the 20th century?
- who are the illustrious female Friulian poets?
- who are the illustrious people who have simultaneously performed the activities of writer, poet, painter and director?

There is no limit regarding the research which can be carried out. All this is possible thanks to the preliminary phase of finding the correct metadata to associate with each dictionary entry; this aspect has been the subject of a in-depth study to achieve the greatest flexibility in subsequent research phases (Fig. 2).

³ The printed version of the “Biographical Dictionary of Friulians. New Liruti” consists of three parts: “1. The Middle Ages”, published in 2006; “2. The Venetian Age”, published in 2009; “3. The Contemporary Age”, published in 2011, for a total of 7285 pages.



Fig. 2. The search form

3.6 Ability to Reach Users All over the World

Who will use the Biographical Dictionary of Friulians online? Certainly the Friulians, but let us not forget that the illustrious people in the dictionary are not only known in the region but throughout Italy and abroad as well. Let us not forget the thousands of Friulians in the world who will certainly appreciate such a tool; we think of scholars in various disciplines; in general let us think about anyone interested in a certain illustrious person and want deepen his biography online. In order to reach the widest visibility, special attention has been paid to the predisposition of all SEO (Search Engine Optimization) techniques aimed at getting the site to be the first result in searches made on various search engines. The aim is to let a user type for example “Caterina Percoto” and obtain as the first result the corresponding biographical entry on the dictionary. Particular importance will be given to verifying the number of accesses and pages viewed by users, using web analytics tools.

4 Technological Solutions Adopted

From a technical point of view, the key choices have been made on the basis of four guiding principles:

- adoption of open source technologies;
- using the most advanced and modern standards;
- making the resources available free of charge;
- independence from devices.

In particular:

- adoption of open technologies: the dictionary was developed using a nearly universal Content Management System (Wordpress). Although it is necessary constantly update the program, updates are freely available and programmers can access the source, therefore the widest guarantees against obsolescence is provided [12]. In other words, the work done will not become obsolete within a few years or even a few months (as is often the case when proprietary technologies are used). In fact, this it allows not to be bound by the company that has made the site since being an open technology, anyone in the future will be able to make changes to the dictionary or to further develop the work.
- use of the most advanced and modern standards: the infrastructure is based on today’s universally accepted standard technologies such as semantic web, RDF, facial navigation, etc. In addition, an endpoint was implemented through the sparQL query language. This means that anyone will have the ability to interface with the dictionary site to retrieve information of their own interest and export this information to their site (maximum openness). As far as we know this is the first case, if not all over the world at least in Italy, to apply these technologies to an online biography dictionary.
- resources available free of charge: access to biographical entries is available for everyone freely and free of charge; anyone can enjoy the work done and enrich their knowledge or even simply satisfy their curiosity by reading biographical entries of the dictionary available without geographical and temporal boundaries.
- independence from the device: the online edition makes content available in a richer and more accessible form by using a responsive technology that allows you to enjoy content using any type of device available today - not just computers but also tablets, phablets, smartphones and any other device, so that the dictionary can reach a much larger audience and regardless of the technology platform used.

5 Significant Technological Aspects

In the context of the world wide web, a semantic annotation provides information about the meaning of a resource and is intended to formally express its content, enabling it to be processable by machines [13]. Automatic resource annotation is an unresolved problem and usually involves human beings with the support of computer tools [16]. In this project, we have decided to make annotations to all biographical entries in the form of RDF triples, starting from information entered by a humanist team within the text (intext) and in external labels (meta tag). RDF triples, stored in a triplestore accessible from server resources exposed by the site, can also be queried with SPARQL, having made available a SPARQL endpoint [14]. In addition to representing one of the conditions required to make biographical entries and their content available as a linked data, the structure allows to carry out very complex queries (for example: what are the musicians who worked in the period 1820–1840 in the city of Aquileia? or: who are the

illustrious people that the city of San Daniele del Friuli must celebrate in 2018?). For example the query

```
PREFIX p:
SELECT DISTINCT ?postId ?sesso0 ?luogoNascita0
WHERE
{
?s p:postId ?postId .
?s p:sesso ?sesso0 .
?s p:luogoNascita ?luogoNascita0 .
FILTER ( ?sesso0 = "F" && ?luogoNascita0 = "Udine" )
}
```

retrieves the illustrious women born in Udine.

One of the objectives of the project was to identify a methodology that would allow the adding of semantic annotations to biographical entries by a team of humanists without geographical or temporal limitations and without the need to be a computer specialist. To achieve this goal, dedicated tools have been built on an easy-to-use and extremely popular Wordpress platform [17]. As a first step, the 2700 biographical entries, available in the PDF format used for the printed edition, were migrated in a hypertext format with automatic recognition of image positions and bibliographic references within the structure. The result was achieved by using an open source tool (pdf2html) that generated XML files with the appropriate formatting instructions; by elaborating such XML files, it was possible to highlight common patterns that allows us to associate them with the semantic aspect and thus rebuild the biographical entries in their original structure (title, subtitle, body, bibliography); it was also possible to extract the first external metadata (for example, the author of the biographical entry). The files thus produced have been read and imported into the database of the Wordpress platform, properly configured and customized. Subsequently, the 2620 biographical entries have been reviewed and annotated semantically using a specially developed tool. The tool that was used to facilitate the work of the review and annotation workgroup is an application that allows you to select and enter RDF element values with your mouse, limiting to the maximum the input of terms from the keyboard. A Wordpress plugin was then created to allow you to select the subject and indicate its property by choosing it from a drop down menu which lists the properties available in that context. The use was very easy and the simple and friendly graphic interface of Wordpress did the rest, allowing the team to get almost complete annotation in less time than originally expected, with great satisfaction from the workgroup. The metadata needed for semantic notation has been inserted in an extremely intuitive manner; in fact, it was enough to select the object element and assign it the appropriate tag (which represents the property) by choosing it from a context-sensitive drop-down list. A section has also been prepared for inserting the semantic annotation outside of the text, leaving editors the ability to enter metadata and postpone a post-text-revision phase to the creation of a controlled vocabulary for predicate

objects. This solution has been chosen to avoid the lengthy amount of time needed to create an internal ontology and to minimize the learning time of an external one. The bibliography has also been annotated and used to create RDF structures which describe external resources. The processed text is then dynamically read by a parser that creates RDF elements, which can take into account both text and external annotations. The terms to be inserted are based on a controlled lexicon and as far as this is currently local for the application, there is a configuration section (now accessible by the code) that will allow you to choose the ontology to use to represent the data outside and create correspondences between vocabulary and the internal one. The availability of a triplestore that can be queried by SPARQL has allowed to offer advanced search and navigation tools. It was possible to create pre-set searches for the user and a search form that proposes the properties and objects for the query, associating the requests with a SPARQL query and returning the responses to the user.

6 Future Developments

The online edition of the dictionary has also been an opportunity both for updating the biographical entries in the light of recent research and studies and to complete the work with the creation of a Supplement that is already under development: so new biographical entries currently missing (for various reasons, including the fact that in the paper edition there are no biographical entries of the characters who died after the printing date) will be added. The “engine” on which the site is based is extremely flexible and powerful and will allow further development in the future. For example, a feature that will be implemented will be geo-referencing the biographical entries, this will allow the user to “point and click” on a Friuli Venezia Giulia city and view all the illustrious people that are related to that city or to map geographically (on maps) the results of any search. Additional functionality can be implemented based on the feedback that will be received from users.

7 Conclusion

In conclusion, the Biographical Dictionary of Friulians (“New Liruti online”) has the ambitious aim of being not only the “digital version” of the printed version of the “New Liruti” but one of the richest and most structured deposits of cultural and historical information on the Italian web, based on the most innovative technologies, with the ability to reach a wider and potentially unlimited audience than the paper edition (consisting not only of scholars and researchers but also of students and ordinary people) so to become one of the most important cultural initiatives within the broadest project on “Cultural Identity of Friuli”.

References

1. Treccani - La cultura Italiana - Biografie. <http://www.treccani.it/biografie/>
2. Donne e uomini della Resistenza. <http://www.anpi.it/donne-e-uomini/>
3. Dizionario biografico dei protestanti in Italia. <http://www.studivaldesi.org/dizionario>
4. American National Biography Online. <http://www.anb.org>
5. ArchiDocWeb-RAH. <http://www.rah.es:8888>
6. Umetnosti, S.: Slovenska biografija. <http://www.slovenska-biografija.si>
7. Scalon, C., Griggio, C., Rozzo, U., Bergamini, G.: Nuovo Liruti. Editrice Forum, Udine (2011)
8. Greenstein, S., Zhu, F.: Do Experts or Collective Intelligence Write with More Bias? Evidence from Encyclopædia Britannica and Wikipedia. Harvard Business School (2014)
9. Documentation for ISO 639 identifier. <http://www-01.sil.org/iso639-3/documentation.asp?id=fur>
10. Google Books Ngram Viewer. https://books.google.com/ngrams/graph?content=Friulan%2C+Friulian&year_start=1800&year_end=2000&corpus=15&smoothing=3&direct_url=t1%3B%2CFriulan%3B%2Cc0%3B.t1%3B%2CFriulan%3B%2Cc0
11. Dizionario Biografico dei Friulani. <http://www.dizionariobiograficodeifriulani.it/?s>
12. Corrado, E.M.: The Importance of Open Access, Open Source, and Open Standards for Libraries. *Issues in Science and Technology Librarianship* (42) (2005). ISSN 1092–1206
13. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Sci. Am.* **284**(5), 28–37 (2001)
14. dizionario Biografico dei Friulani - SPARQL End Point. <http://www.dizionariobiograficodeifriulani.it/sparql/>
15. Reinert, M., Schrott, M., Ebnet, B.: Team deutsche-biographie.de: from biographies to data curation – the making of www.deutsche-biographie.de (2005). <http://ceur-ws.org/Vol-1399/paper3.pdf>
16. Tosi, D., Morasca, S.: Supporting the semi-automatic semantic annotation of web services: a systematic literature review. *Inf. Softw. Technol.* **61**, 16–32 (2015)
17. Munford, M.: How wordpress ate the internet in 2016... and the world in 2017, *Forbes* (2016). <https://www.forbes.com/sites/montymunford/2016/12/22/how-wordpress-ate-the-internet-in-2016-and-the-world-in-2017/#6926823d199d>