

SCIBA - A Prototype of the Computerized Cartographic System of an Archaeological Bibliography

Eleonora Bernasconi¹, Paolo Boccuccia², Marco Fabbri³, Alessia Francescangeli⁴, Roberto Marcucci⁴, Massimo Mecella¹, Maura Medri⁵, Alberto Morvillo¹, Marcella Pisani³ and Emiliano Tondi³

¹*Sapienza Università di Roma, Department of Computer, Control, and Management Engineering Antonio Ruberti (DIAG), via Ariosto, 25, 00185 Rome, Italy*

²*Museo delle Civiltà, piazza Guglielmo Marconi, 14, 00144 Rome, Italy*

³*Roma Tor Vergata, Department of History, Cultural Heritage, Education and Society, Urban Landscape Archeology Laboratory (ARPAE), via Columbia 1, 00133 Rome, Italy*

⁴*L'Erma di Bretschneider, via Marianna Dionigi, 57, 00193 Rome, Italy*

⁵*Roma Tre, Department of Humanistic Studies, Cultural Heritage Laboratory (Digital Humanities Laboratory), via Ostiense, 234, 00146 Rome, Italy*

Abstract

Archaeological and historical-artistic studies, together with correlated multidisciplinary research, regardless of the language of writing or technical terminology, find a natural connection between them in the topographical positioning. Studying a monument or area from an archaeological and historical-artistic point of view cannot be separated from a long search for the material published in specialized libraries. SCIBA aims to drastically reduce search times and provide an instant overview of particular themes through knowledge extraction from reference texts and the visualization of the concerned cartography. With a simplified user experience, the user will be able to identify the area, location, archaeological site or single monument of interest, exactly as already used to do with contemporary navigation systems, and in addition will be able to view in real-time all the related bibliography and to navigate the key concepts in a knowledge graph.

Keywords

Knowledge extraction, Digital library, Archaeology,


1. Introduction


The SCIBA (Sistema Cartografico Informatizzato della Bibliografia Archeologica) project envisages the development of an innovative bibliographic search system of an archaeological and historical digital library on a cartographic basis. The platform, which can then be expanded on a national and international scale, will focus on the Lazio region (Italy), with the support

Joint Proceedings of RCIS 2022 Workshops and Research Projects Track, May 17-20, 2022, Barcelona, Spain

✉ bernasconi@diag.uniroma1.it (E. Bernasconi); paolo.boccuccia@beniculturali.it (P. Boccuccia); fabbri@uniroma2.it (M. Fabbri); alessia.francescangeli@lerma.it (A. Francescangeli); roberto.marcucci@lerma.it (R. Marcucci); mecella@diag.uniroma1.it (M. Mecella); maura.medri@uniroma3.it (M. Medri); morvillo@diag.uniroma1.it (A. Morvillo); marcella.pisani@uniroma2.it (M. Pisani); emiliano.tondi@uniroma2.it (E. Tondi)

ORCID 0000-0003-3142-3084 (E. Bernasconi); 0000-0002-9730-8882 (M. Mecella); 0000-0001-5154-6095 (A. Morvillo)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

of the “L’Erma di Bretschneider” publishing house, and will allow the visualization of the existing bibliography concerning some selected topographical elements. It will also be possible to expand the search to further related themes, automatically extracted from the system based on the contents and metadata of the integrated bibliographic material. The aim is to develop a search and comparative semantic analysis tool of great utility for research bodies, museums, municipalities and subjects interested in knowledge and management of the territory, and, at the same time, capable of generating an increase in volume sales with direct employment effects of the publishing chain.

2. Summary of the project

The SCIBA platform is based on the ARCA platform, a visual-search based system that allows the semantic exploration of a bookstore, and arises from the researchers’ need to explore knowledge bases in a cartographic context [1, 2, 3]. This is done by searching for links between the different topics related to a place or a key term, in such a way as to reveal unexpected connections during the exploration of contents and, thus, generating new ideas. These connections between concepts are visually represented by means of a graph that the user can arrange on her own, a solution already used in other projects, such as NOTAE, a tool to investigate the graphic symbols in order to capture all the possible historical implications [4], which uses a similar visual-search system to navigate a Knowledge Graph.

The content search will be based on semantics and cartographic visualization. Unlike the classic search, where the search engine or application, based on keywords, proposes the texts, documents and metadata in which those words are present as a list of results, the semantic search has the aim of expanding the results and improve their accuracy by eliminating unnecessary results such as in the case of homographs or by including conceptually relevant results. To carry out semantic searches, it is necessary that metadata and texts are “annotated”, that is, that unique identification codes (actually IRI, Internationalized Resource Identifier) are assigned to words contained in the texts, regardless of gender/number or, in the case of verbs, the inflected form and, in some cases, regardless of language. The actual search, then, will not be carried out on the single words but on the combination of semantic fields (the word, in all the genres and numbers in which it can be declined, synonyms, derivatives and the other words of the same semantic field).

Finally, by applying semantic search on a base cartography, it is possible to use a place name as a search key for a given theme. This is obtained by associating the cartography geometric and geographical information with the concept of “place” (i.e. a region or a point of interest), obtained from the reference texts with the use of an external AI. This AI will extract all the entities, represented by words from the texts, that will then linked by the platform to an external Knowledge Graph, specifically DBpedia, in such a way as to acquire the main information contained, included those concerning places. Geographical information of the places of interest are instead contained in the source maps files (i.e. GeoJSON file format¹).

Once internal data (entities extracted from PDFs and maps, provided by a manager-role user) and the external information (DBpedia, linked by the platform) have been obtained, these

¹<https://geojson.org/>

two will be conveyed by the platform, with the use of a triple store, into a specifically created Knowledge Graph: the “SCIBA knowledge graph” which can be explored in all its branches.

While the extraction of entities from the data, which is time-consuming, will have to be in batches, i.e., upstream of the whole process, every time the data sources (PDFs and maps) increase or undergo changes, the content search will be performed in real time and will have three different modes:

- **Location Search:** search by area indicated on a map, displaying all concepts related to it;
- **Entity Search:** search for a word or a person within the database and extracts all the relationships that branch off from the main entity being searched;
- **Road map:** plots the links between the relations of two searched entities (origin and destination), allowing to ‘navigate’ in the topic of interest for subsequent approximations and reach other contents and documents on further potentially interesting topics.

The user of the SCIBA platform, having tracked down all the contents of her interest, will be able to organize them according to her own needs and interests by querying the search engine, interacting with the expansion and multiplication of the information returned and exploring the virtual library and the graph.

3. Objectives and Expected Tangible Results

The aim is to create a prototype of the platform based on the sources provided by “L’Erma di Bretschneider” publishing house² and the cartographic files of the Lazio Region, which can be extended, at a later stage, to other editions of international publishers who have published research carried out in the Region.

The objectives of the project consist mainly in:

- location of the reference territory of the studies published by L’Erma di Bretschneider;
- integration of an automated semantic search system;
- identification and acquisition of possible stakeholders;
- elaboration of dissemination strategies capable of reaching the widest heterogeneity of markets to identify subjects particularly sensitive to the potential of the product. The SCIBA project also aims to increase the knowledge of minor archaeological sites, those which, while registering a low number of visitors, are of great cultural importance.

An essential aspect of the platform concerns its usability by users, with a user experience aimed at researchers in the field of archaeology and therefore emphasizing the cartographic aspect and reducing technicalities and mechanisms that could complicate its use.

²<https://www.lerma.it/>

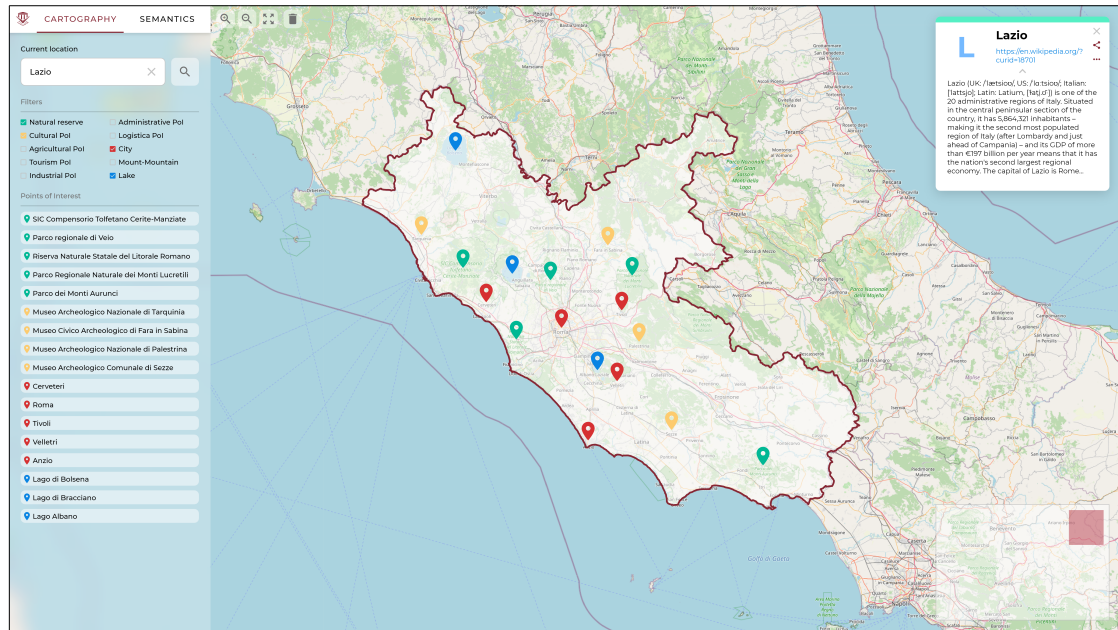


Figure 1: SCIBA platform working area mockup. The user can select the places (a region or a point of interest) on a background map, with a side panel for the semantic search while floating boxes shows the selected graph's entities.

4. Current Project Results

The SCIBA project started in October 2021 and is expected to end by April 2023 (18 months). So far, only User Experience research and platform definition activities have been performed. In particular:

- User experience and interaction have been defined. It must have a map-oriented interface with simplified interactions, being a platform intended for use by archaeological researchers.
- The system requirements has been set so they can satisfy the possibility of an on-premise installation with a virtualized environment.
- The platform structure has been defined. It must allow maintainability of the data sources by the management staff, so they can keep them updated without technical intervention.
- The geographical place integration in the SCIBA knowledge graph is still under analysis. The current approach is to use another knowledge graph to link geographical information to a concept from the generic knowledge graph (see Figure 2).

5. Conclusions

In this paper, we introduced the SCIBA project. The project's goal is to give a handy tool for researchers in archaeology that allows them to obtain information quickly, have reference

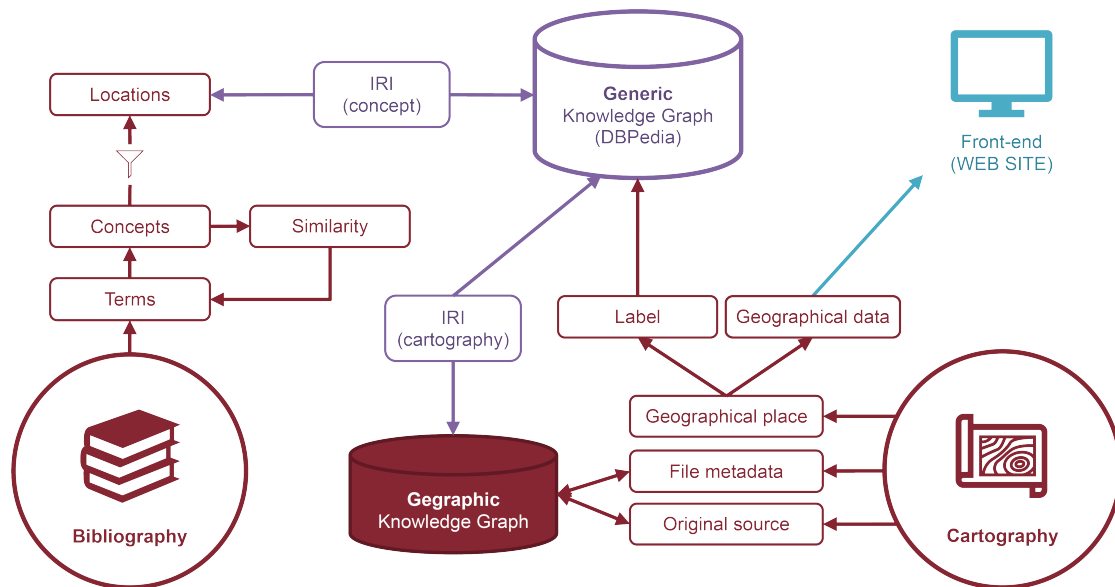


Figure 2: Current approach for the geographic contextualization of contents in SCIBA. Concepts extracted from the terms contained in the books are inserted in a knowledge graph, in which they will be identified through an IRI (Internationalized Resource Identifier). Those concerning a location will be associated to their respective cartographic data, also identified through an IRI.

cartography, and the possibility of extrapolating and linking concepts from a bibliography. The project is in an initial phase where only User Experience research and platform definition have already started.

At the first point, the fruition and management of cartographic files are still under development. The platform must allow to load, filter display and navigate the maps. Moreover, how the concepts extracted from the texts are associated with a geographical place has currently only been defined and not yet implemented. Each map file must be analyzed indexed, and its elements must be associated with a concept in the knowledge graph. Another challenge is to simplify the management of data sources. At present, loading and analyzing the contents requires complex operations that are difficult to carry out by a non-technician. Finally, the public website, which will allow the researchers to explore contents and items verified by SCIBA curators, is currently in the design phase to make it publicly available in 2023.

Acknowledgments

This work has been supported by Regione Lazio and MIUR [Determinazione n. G07413 del 16/06/2021]

References

- [1] M. Ceriani, E. Bernasconi, M. Mecella, A streamlined pipeline to enable the semantic exploration of a bookstore, in: *IRCDL 2020*, Springer International Publishing, Cham, 2020, pp. 75–81. URL: https://link.springer.com/chapter/10.1007/978-3-030-39905-4_8.
- [2] E. Bernasconi, M. Ceriani, M. Mecella, Exploring a text corpus via a knowledge graph, in: *IRCDL*, 2021, pp. 91–102. URL: <http://ceur-ws.org/Vol-2816>.
- [3] E. Bernasconi, M. Ceriani, M. Mecella, T. Catarci, M. C. Capanna, C. Di Fazio, R. Marcucci, E. Pender, F. M. Petriccione, Arca. semantic exploration of a bookstore, in: *Proceedings of the International Conference on Advanced Visual Interfaces*, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–3. URL: <https://doi.org/10.1145/3399715.3399939>.
- [4] E. Bernasconi, M. Boccuzzi, T. Catarci, M. Ceriani, A. Ghignoli, F. Leotta, M. Mecella, A. Monte, N. Sietis, S. Veneruso, et al., Exploring the historical context of graphic symbols: the notae knowledge graph and its visual interface, in: *IRCDL*, 2021, pp. 147–154. URL: <http://ceur-ws.org/Vol-2816>.