

ARCA. Semantic exploration of a bookstore

Eleonora Bernasconi
DIAG
Sapienza Università di Roma
Rome, Italy
bernasconi@diag.uniroma1.it

Tiziana Catarci
DIAG
Sapienza Università di Roma
Rome, Italy
catarci@diag.uniroma1.it

Roberto Marcucci
L'Erma di Bretschneider
Rome, Italy
roberto.marcucci@lerma.it

Miguel Ceriani
Dipartimento di Informatica
Università di Bari Aldo Moro
Bari, Italy
miguel.ceriani@uniba.it

Maria Cristina Capanna
DSA
Sapienza Università di Roma
Rome, Italy
mariacristina.capanna@uniroma1.it

Erik Pender
L'Erma di Bretschneider
Rome, Italy
erik.pender@lerma.it

Massimo Mecella
DIAG
Sapienza Università di Roma
Rome, Italy
mecella@diag.uniroma1.it

Clara Di Fazio
DSA
Sapienza Università di Roma
Rome, Italy
clara.difazio@uniroma1.it

Fabio Maria Petriccione
TSP - Tecnologie e Servizi
Professionali s.r.l.
Rome, Italy
fabio.petriccione@tspservizi.it

ABSTRACT

In this demo paper, we present ARCA, a visual-search based system that allows the semantic exploration of a bookstore. Navigating a domain-specific knowledge graph, students and researchers alike can start from any specific concept and reach any other related concept, discovering associated books and information. To achieve this paradigm of interaction we built a prototype system, flexible and adaptable to multiple contexts of use, that extracts semantic information from the contents of a books' corpus, building a dedicated knowledge graph that is linked to external knowledge bases.

The web-based user interface of ARCA integrates text-based search, visual knowledge graph navigation, and linear visualization of filtered books (ordered according to multiple criteria) in a comprehensive coordinated view aimed at exploiting the underlying data while avoiding information overload and unnecessary cluttering. A proof-of-concept of ARCA is available online at <http://arca.diag.uniroma1.it>

CCS CONCEPTS

• **Human-centered computing** → **Web-based interaction**; • **Information systems** → *Digital libraries and archives*; *Search interfaces*.

KEYWORDS

knowledge graph, books' catalog, visual search interface

ACM Reference Format:

Eleonora Bernasconi, Miguel Ceriani, Massimo Mecella, Tiziana Catarci, Maria Cristina Capanna, Clara Di Fazio, Roberto Marcucci, Erik Pender,

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

AVI '20, September 28-October 2, 2020, Salerno, Italy

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7535-1/20/09.

<https://doi.org/10.1145/3399715.3399939>

and Fabio Maria Petriccione. 2020. ARCA. Semantic exploration of a bookstore. In *International Conference on Advanced Visual Interfaces (AVI '20)*, September 28-October 2, 2020, Salerno, Italy. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3399715.3399939>

1 INTRODUCTION

Showing the results of the search on the content of a book can be a very hard task if the data are not organized and shown in a meaningful way. Digital libraries and publishing companies need to facilitate the navigation of their catalogs of books to attract most people in reading their catalogs. Hence there is the need to attach semantics to unstructured information like the text of the books, and this is possible with Named Entity Recognition (NER), Named Entity Linking (NEL) and Knowledge Graph (KG) techniques. In this way, the static pieces of the book's text, become interactive, explorable and interconnected concepts.

Some of the authors previously designed a pipeline for semantic enrichment and visualization of text corpus [3]. After the text is extracted from a book and sentences are split, the NER technique discovers the concepts presents in the sentences, like names of people, organizations, places, etc.; the NEL technique disambiguates the meaning of the identified concepts and establishes the relationships between each other; at the end, KG technique enables the explorations of the entities and their connections, through a search interface.

The pipeline has been implemented in a tool called ARCA. In this demo paper, we focus on the visual user interface of ARCA. ARCA was tested on a selected catalog of 110 books of L'Erma di Bretschneider, a publishing company in the field of archaeology and ancient Rome history. For the visualization of the knowledge graph, ARCA uses the free visualization tool Ontodia [6]. For the development of the other components, like the books' catalog component, we used the React framework¹. The development took place in parallel with the evaluation carried out by a small focus group of

¹<https://reactjs.org/>

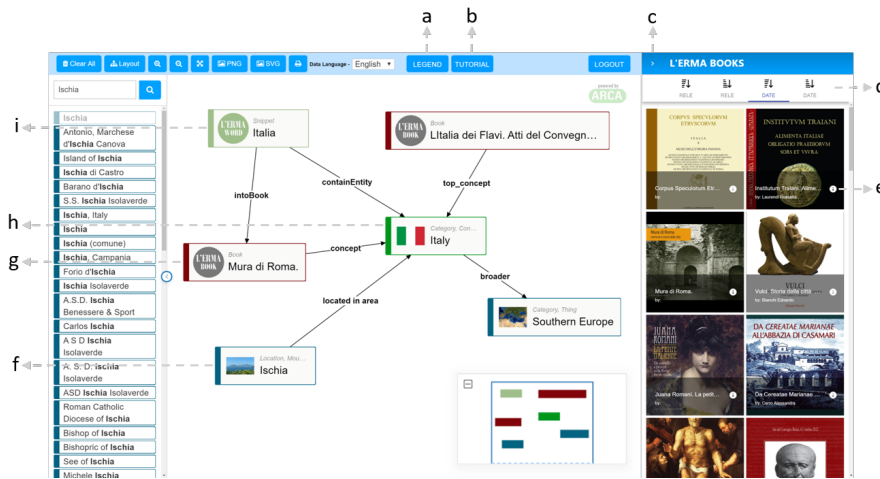


Figure 1: The ARCA visual search interface

six researchers in archaeology and history. This helped to refine details and features of the interface that improved its usefulness and usability, as the snippet component that shows fragments of text which contain the selected concept.

2 RELATED WORK

The strength and innovation of ARCA is its flexibility and adaptability to any type of textual content. Other interfaces for visualization and exploration of knowledge graphs have been investigated, but most of the analyzed tools are either monolithic systems like Yewno Discover [1] or are not-ready-to-use systems, like Apache Stanbol² or the GLOBDEF system [7], which are a set of components able to offer various services for semantic enrichment and the management of metadata, but they need to be integrated and composed before being usable.

3 ARCA INTERFACE

In the following, we briefly outline the main ARCA interface, by referring to Figure 1. The expected types of users are mainly two: researchers (mainly in humanities) who want to discover and connect multifarious information to find the books that can help them in their research activities; lay users who are interested in discovering new things.

Before starting with the search and the exploration of the contents, users can view the legend (a), the tutorials (b) and helps that will give them the instructions to perform the activities that ARCA allows. In Figure 1, they can be observed 4 different kinds of colors of the entities: green, light green, red and blue. The green entities (h) represent the concepts extracted from the books' corpus; the light green entities (i) represent the words that the NER identifies in a sentence as entities; the red entities (g) represent the books of the catalog (e.g., those ones of L'Erma) and finally, the blue entities (f) are the concepts present in the DBpedia knowledge base [4]. Starting from the search of a concept the users can observe the results and choose the topic to be explored by drag-drop it on the

central board. Only green entities (h) are connected with the catalog of the books accessible from the book button (c) on the top of the screen. The books catalog (e) shows the results of a query that ask to view only the books for which the selected concept is a "concept" or "top concept" namely, where "concept" is among the extracted entities that appear more than 20 times in the book and "top entities" are the 10 most frequent entities extracted from the book. The books can be ordered by relevance (concept/top concept) or by year of publication. If users want to view more information about a book, they can select the info button (d) in the bottom right of the book table and users can view the sentences in the book which contains the selected entities, they are referred to as "snippets". Furthermore, books from the right panel can be dragged to the central board to view them as nodes in the knowledge graph, revealing as connections all the related concepts.

4 CONCLUSIONS AND FUTURE WORK

This demo illustrates the ARCA interface, a concrete tool for enhancing access to a catalog of books through knowledge graph-based exploration. ARCA allows users to discover new things in the books' catalog through semantic search and the free exploration of content and connections. In future developments, the authors envision the implementation of new components to further enable multi-faceted content exploration. Such as the development of a storytelling component that connects geographic places and metadata to maps [2] and software that extracts images and caption from books [5] and integrate them into the ARCA knowledge graph.

ACKNOWLEDGMENTS

This research has been partly supported by project ARCA (POR FESR Lazio 2014–2020 - Avviso pubblico "Creatività 2020", domanda prot. n. A0128-2017-17189).

²<https://stanbol.apache.org>

REFERENCES

- [1] Manisha Bolina. 2019. Yewno Discover. *Nordic Journal of Information Literacy in Higher Education* 11, 1 (2019).
- [2] Cecilia Cadenas. 2014. Geovisualization: Integration and Visualization of Multiple Datasets Using Mapbox. (06 2014).
- [3] Miguel Ceriani, Eleonora Bernasconi, and Massimo Mecella. 2020. A Streamlined Pipeline to Enable the Semantic Exploration of a Bookstore. In *Digital Libraries: The Era of Big Data and Data Science*, Michelangelo Ceci, Stefano Ferilli, and Antonella Poggi (Eds.). Springer International Publishing, Cham, 75–81.
- [4] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, et al. 2014. DBpedia – A large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal* 5 (2014), 1–29.
- [5] Pengyuan Li, Xiangying Jiang, and Hagit Shatkay. 2018. Extracting Figures and Captions from Scientific Publications. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (Torino, Italy) (CIKM '18)*. Association for Computing Machinery, New York, NY, USA, 1595–1598. <https://doi.org/10.1145/3269206.3269265>
- [6] Dmitry Mouromtsev, Dmitry Pavlov, Yury Emelyanov, Alexey Morozov, Daniil Razdyakonov, and Mikhail Galkin. 2015. The simple, Web-based tool for visualization and sharing of semantic data and ontologies.
- [7] Maria Nisheva-Pavlova and Asen Alexandrov. 2018. GLOBDEF: A Framework for Dynamic Pipelines of Semantic Data Enrichment Tools. In *Proc. of MTSR 2018*. Springer, 159–168.