



Design, realization, and user evaluation of the ARCA system for exploring a digital library

Eleonora Bernasconi¹ · Miguel Ceriani² · Massimo Mecella¹ · Tiziana Catarci¹

Received: 22 November 2021 / Revised: 8 November 2022 / Accepted: 9 November 2022 / Published online: 16 December 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

This paper presents ARCA, a software system that enables semantic search and exploration over a book catalog. The main purpose of this work is twofold: to propose a general paradigm for a semantic enrichment workflow and to evaluate a visual approach to information retrieval based on extracted information and existing knowledge graphs. ARCA has been designed and implemented following a user-centered design approach. Two different releases of the system have incrementally and iteratively developed and evaluated. The first release has evaluated the quality and usefulness of the extracted data. The second release, whose design was a refinement based on the previous evaluation results, was assessed by several users. Moreover, a comparative test with other information retrieval systems was conducted in order to study the potential added-value of the system. ARCA is employed in a real editorial scenario to visually search and explore the books of a publishing house.

Keywords Semantic enrichment · Knowledge graph · Visual search interface

1 Introduction

Searching and exploring a vast text corpus have often arisen as a human need. Traditionally, the search process is based on manually curated metadata classifying documents by arguments, authors, metadata, and so on.

Although the metadata that are used to be stored in physical cabinets are now stored in databases, the process often remains similar.

Albeit being a decisive paradigm, the maintenance of metadata is costly and becomes progressively more expensive and less reliable with the increase in required detail. The transition to electronic documents (either created natively as such or digitized) enables direct text-based search of the content. Text-based search on the whole content of documents is a powerful tool, but it comes with its own limitations due to the inherent ambiguity of natural languages and the need for the user to anticipate the actual words used in the content, as the machine is not able to capture what the user and the corpus *mean*. This is called the semantic gap. Statistical methods can be successfully used for query expansion, mitigating the issue, but the user has no control of the process. Semantic enrichment methods, as *named-entity recognition and linking (NERL)* [28,37], aim instead directly at bridging the semantic gap between raw text and concepts, by associating words in the documents with entities in a knowledge base, often a knowledge graph (KG). NERL successfully enabled users to search and analyze text corpora [36] more effectively. Nevertheless, the navigation of semantic relationships (with their meaning, rather than just as generic connections) between extracted entities has seldom been adopted as a method for the exploration of a corpus, even if it is known that the cognitive processes in library searching are generally more complex than a single topic-based search [25]. Also, while knowledge extraction methods as NERL are now broadly used by

This is an extended and revised version of a preliminary conference paper that was presented in IRCDL 2021 [14].

✉ Eleonora Bernasconi
bernasconi@diag.uniroma1.it

Miguel Ceriani
miguel.ceriani@uniba.it

Massimo Mecella
mecella@diag.uniroma1.it

Tiziana Catarci
catarci@diag.uniroma1.it

¹ Dipartimento di Ingegneria Informatica, Automatica e Gestionale Antonio Ruberti, Sapienza Università di Roma, Via Ariosto 25, 00185 Rome, Italy

² Dipartimento di Informatica, Università di Bari Aldo Moro, Via Orabona 4, 70125 Bari, Italy

big players in the industry as well as in academic projects, their usage by small- to medium-sized organizations (which often have text corpora, either private or public, that they struggle to manage in a structured way consistently) is still minimal, in part due to the lack of an established standard workflow.

In this work, we analyze the applicability and usefulness of a corpus search and exploration paradigm based on the transparent use of knowledge graphs. To this purpose, a tool called ARCA has been developed. It includes a pipeline for semantic enrichment of textual content and a user interface that enables search and exploration of the corpus through visual navigation of a knowledge graph of topics.

The extracted semantics and user interface is supporting different general search behaviors, which were deemed useful in the specific use case we analyzed, and we conjecture they are of general interest. The main general supported search behaviors are the following:

- Find documents relevant to a specific topic;
- Expand or specialize searches by moving through related topics;
- Have visibility of available related resources, which could potentially be of interest;
- Visually organize the resources found by considering their relationships and properties;
- Find topics and documents at the crossing of multiple topics, possibly of different kinds (places, people, time periods, etc.).

1.1 Research questions

For the sake of the analytic approach, we frame our experimentation effort through a set of research questions. The questions elicited below are relevant to the application of KG-based approaches for the exploration of text corpora.

Q1 Would users exploring a corpus of text profit from the semantic navigation of the associated KG of topics?

Q2 What kind of user interface would effectively support such a navigation?

Q3 What kind of users, scenarios, and tasks would benefit from this interaction paradigm?

Q4 Do building and maintaining a semantic enrichment and KG creation pipeline necessarily involve high upfront costs and highly skilled developers?

1.2 Hypotheses

To reply to the questions above, we designed the presented study to test the following main hypotheses.

H1 (relevant to Q1 and Q2) Users will be able to effectively explore a text corpus through a KG-based user interface, which offers the following main functions: *a.* finding concepts through text search (among the ones pertinent to the specific domain), *b.* visually navigating the concepts and their relationships, and *c.* showing documents relevant to the selected concept.

H2 (relevant to Q3) The method, given a corpus of texts in a specific domain, will benefit both users with little knowledge of the domain (by supporting semantically relevant discovery) and domain experts (by enabling a topic-oriented visual organization of the documents).

H3 (relevant to Q4) It is feasible to build a ready-to-use complete system, including both semantic enrichment pipeline and web-based front end, which is able, with only some configuration, to be applied to any specific corpus to enable the KG-based exploration.

While the first two research questions and related hypotheses are relevant for investigating the benefits of the proposed approach for the end users, the last research question and hypothesis investigate the usefulness and portability of such a system to different contexts of use.

1.3 Approach

ARCA is the software system designed to enable the KG-based exploration of a given text corpus and test our hypotheses.

The system is organized according to the following main functions:

- Extraction of entities from a given text corpus;
- Integration between available metadata, extracted entities present in the text, and data from external knowledge bases;
- Consolidation of the local data in a KG stored in a triple store;
- Search and exploration of the corpus through the navigation of the KG in a composite user interface.

In order to ensure the whole solution is useful for potential users, it has been implemented and evaluated within a specific case study: exploring the book catalog of a medium-sized publishing house specializing in ancient history. The concrete case study offered the context for fruitful exchange among the stakeholders that are often involved in scenarios of information retrieval and library search:

- Who maintain the corpus (the publisher);
- Who need to search the corpus (researchers of the field and interested individuals);

- Who develop the software solution (in this case the authors of the present study).

The remaining sections are organized as it follows. Section 2 presents related work about visual information seeking. Section 3 reports the design process starting from identifying user requirements to the final interface's development and implementation. Section 4 introduces the relevant technical background, before describing our system in Sect. 5. Section 6 reports the evaluation process and analyses the findings. Finally, Sect. 7 draws conclusions and discusses future research directions.

2 Related work

In this section, relevant literature is surveyed for relevant works, starting from traditional systems for visual information seeking to tools for semantic enrichment of unstructured text and visualization/exploration of semantic data as KGs, both in the general case and in the specific case of a corpus of books.

2.1 Traditional systems

There has been a large amount of work in the literature about visual information seeking [4,21,38]. The first attempts to create a visual search interface have been done in the early 1990s [2], where some researchers had applied direct manipulation principles to search interfaces, creating what they called dynamic queries [1].

These are visual query systems, often based on the query-by-example paradigm [43]: search interfaces where users can manipulate sliders and other graphical controls to change search parameters. The results of those changes are immediately displayed to them in some visualization.

As an example relevant to our case, in the site of the publishing house “L’Erma di Bretschneider”¹ there is a traditional book search system that allows searching by keywords contained in book titles and by categories. For conciseness, we will call this search system *Lerma*.

Torrossa² is the digital search platform of “Casalini Libri” to which about 180 publishers, mainly Italian and Spanish ones, adhere with their contents. Torrossa allows an advanced search by metadata and by words contained in the books. A limit of these systems, for unstructured information like books, is that exploring and filtering by basic metadata (i.e., author, title, etc.) can be useful, but it is often insufficient.

¹ <http://www.lerma.it/>.

² <https://www.torrossa.com/>.

2.2 Semantic enrichment

There has been recently much research on how to attach semantics to unstructured data [17,36], through processes like NERL.

The GLOBDEF system [29,33] works with pluggable enhancement modules, which are dynamically activated to create on-the-fly pipelines for data enhancement. Apache Stanbol³ uses fixed, albeit configurable modules to process semantic enrichment and the management of metadata.

Both tools provide interesting paradigms to build a flexible pipeline for semantic enrichment.

In comparison with ARCA, they do not provide directly a front end to use the semantic information for information retrieval, which is crucial for the stated purposes and scientific questions of the present work. Furthermore, while GLOBDEF and Stanbol testify about the interest in this kind of solution, neither of them is actively maintained, the former being stuck in prototype status while the latter has been retired, and thus not practically usable to test the stated hypotheses.

2.3 Visualization of semantic data

The extracted semantics can then be extremely useful for exploring the data, but they are not fixed and homogeneous like a set of predefined metadata. Therefore, data models and visual user interfaces need to deal with these complex and heterogeneous data. The semantic web [9] and linked data [12] efforts deal with data modeling, integration, and interaction of this kind of data on the web. These efforts lately contributed to the emergence of KGs to organize complex datasets integrating multiple sources [16,39].

Many user interfaces for visualization and exploration of KGs exist, and new ones are developed every year, especially using semantic web and linked data technologies [10,15,24,34].

Metaphactory [18] is a platform for building KG applications that can be integrated into other software infrastructures. Metaphactory includes Ontodia, a user interface component for the visual exploration of KGs. The interaction paradigm is based on the idea of loading in the main panel of the tool the fragment of interest of the entire KG (which can consist of local data, a remote SPARQL endpoint, or the merge of multiple such sources). Entities can be found through textual search and then dragged to the main panel. Connections among entities are shown, and new entities can also be added by expanding the connections of shown entities.

ARCA adopts the interaction paradigm proposed by Ontodia, as part of an integrated user interface that includes a panel

³ <https://stanbol.apache.org/>.

with the list of documents related to the selected topic and other dedicated components.

Sampo-UI [20] is a framework that provides a set of reusable and extensible components, application state management, and a read-only API for SPARQL queries, which can be used to create a user interface for a semantic portal. Differently from Sampo UI, ARCA offer also a knowledge extractor service from unstructured data and a semantic enrichment service.

2.4 Exploration of a digital library

Many tools face the challenge of exploring the contents of a digital library, but two in particular go in the same direction of this work.

Yewno Discover [13] is an integrated system that offers classification and visual exploration of academic materials to help scholars in their research, but is not adaptable and flexible to different contexts of use, except with ad hoc adjustments. Furthermore, in respect to ARCA, it makes limited use of the KG structure for exploration, which is at the core of the research questions posed here.

Talk to Books⁴ is a tool by Google to explore ideas and discover books by getting quotes that respond to user's queries. It aims at helping users to find relevant books that may not be directly identified through keyword search, but does not provide a way for the user to autonomously explore the underlying knowledge base.

3 System requirements

The publishing house's specific use case offered the opportunity to adopt a user-centered design approach to identify and refine the system requirements. From informal interviews with publishing house representatives and a team of researchers in the same domain, an initial set of requirements has been identified:

- The user should be able to search entities textually;
- For an entity, the user should be able to see the relevant books;
- The user should be able to navigate among entities, following semantic relationships between them;
- For a document, the user should be able to access the basic information and be informed on how to obtain it (buy it from a bookstore, borrow it from a library, etc.);
- Any user should be able to perform operations without being taught how to, by following established interaction patterns and metaphors.

⁴ <https://books.google.com/talktobooks/>.

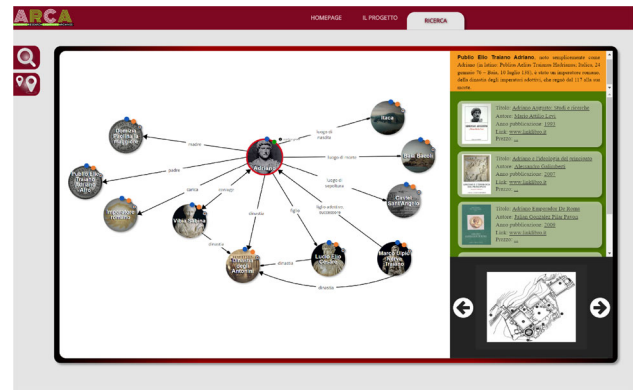


Fig. 1 Mockup of the user interface

A series of intermediate evaluations were carried on, including the following methods:

- Evaluation of extracted data quality by expert analysis;
- Tests and discussion with low fidelity prototypes (as an example, the reader can see the mockup in Fig. 1, which was one of the initial proposals, and can compare with the interface shown in Fig. 5);
- Tests and discussion with high fidelity prototypes (progressively closer to the final system).

During these iterations, the following additional requirements were identified:

- Entities which appear more frequently in a document (main topics) should be distinguished from less relevant entities;
- Users need to check the textual context in which an entity was found in a document.

4 Technical background

Preliminary, we want to describe the technologies underlying the proposed system briefly. Semantic technologies enable the transformation of unstructured information, like those present in the textual PDF documents, to structured data.

The semantic web [32,41], according to Berners-Lee, is “a web of things of the world, described by the data on the web” [11]. The concept is generic, but contains some crucial references:

- The network (graph);
- Things (objects related in a meaningful way);
- Data (no longer records, but connections among nodes of a network).

The concept of the semantic web is closely related to the concept of linked data as an effective method and technique

for simplifying and homogenizing solutions to identity interoperability problems, promoting the univocal identification of data in the dialogue between heterogeneous systems.

Linked data [19] are based on a set of techniques that, through shared vocabularies, allow non-human agents to understand content published on the web. The linked data initiative include both technology and a set of best practices for publishing data on the web in a readable, interpretable, and usable way by a machine.

A knowledge graph [35] is a set of graph-structured data where nodes represent entities of interest and edges represent relationships between them. Apart from the data model, what distinguishes a knowledge graph from a typical database is that it is not tied to a specific application. A knowledge graph is meant to hold information that is of interest for a company, a community, a domain of knowledge, or even include data from multiple domains. This graph is often enriched with various forms of schemata, rules, ontologies, and more, to help validate, structure, and define the semantics of the underlying graph. When considered at web scale, the idea of knowledge graphs overlaps with the concept of linked data.

Making data understandable to machines imply the sharing of a typical data structure. The RDF (Resource Description Framework) [26] is the language proposed by the W3C for achieving a standard data structure as a graph, in the context of linked data.

In RDF, data are organized around *resources*. Relations between resources are represented through *triples*, i.e., subject–predicate–object associations. Subject and object are a pair of related resources. The predicate is another resource which specify the meaning of the relation. Resources used in the predicate role are called *properties*. Furthermore, the object of a triple can also be a *literal*, i.e., a simple value conforming to some basic datatype as string, integer, date, etc. Resources have types too, which are specific resources called *classes*. A resource may have multiple types. An *RDF graph* is defined as a set of triples.

Resources (and datatypes too) in RDF are uniquely identified with *IRIs* (Internationalized Resource Identifiers), an extension of web URLs. To reduce the complexity of memorizing and writing down long IRIs, RDF provide a shortening mechanism with which the initial part of an IRI can be replaced with an abbreviation, called *prefix*, separated with a colon (“:”) from the final part of the IRI. Usually IRIs and prefixes are used so that a group of related concepts can be written using the same prefix. An initial part of IRI that is common to multiple related resources and is typically associated with a prefix is often called a *namespace*.

RDF can be serialized in a number specific data formats, in Listing 1 a fragment of an RDF graph is serialized in Turtle [5]. Multiple prefixes are used, associated with different namespaces.

To favor interoperability, *vocabularies* composed of RDF classes and properties are defined and shared. RDF Schema (RDFS) [3] is a data modeling vocabulary for RDF data that provide the means to describe vocabularies, including classes, properties, and their basic relationships (e.g., domain and range of a property, hierarchical relationship among classes). Beyond RDFS, the Web Ontology Language (OWL) [27] allows the specification of ontologies, which are vocabularies attached to stronger constraints enabling more expressive modeling.

SPARQL [42] is one of the key technologies of the semantic web, and it is used to retrieve and manipulate RDF data from the knowledge graphs available on the web. The SPARQL endpoints allow clients to issue SPARQL queries over a dataset, receiving direct results from the server.

5 The system

The software system has been implemented and tested in the specific use case, but it is designed for general use. The aim is to offer a ready-to-use package to explore any corpus of texts through a specialized KG visually. In this section, the software is described starting from its modular organization. Then, an overview of the data model and structure of the integrated KG is presented. Later the user interface is described. Finally, the implementation details are given.

5.1 Software modules

Figure 2 shows informally the system’s main software modules, different user categories, and the data flow among them. For clarity, modules and flows are organized in the main functional areas (numbered from 1 to 4). The system is composed of a pipeline to build the KG and a web-based front end to search the corpus using the KG.

The pipeline can be seen as composed of three steps, roughly corresponding to the functional areas 1, 2, and 3 of the diagram in Fig. 2: in the first step, newly added documents of the corpus enter the pipeline; in the second step, semantic enrichment services extract information from the documents; in the third step, the generated data are consolidated locally to be also integrated with additional data provided by external services.

RDF is used to represent all the data items in the pipeline, employing existing vocabularies and ontologies whenever possible and creating new terms if needed.

In the first step, the documents’ content (e.g., PDFs) is stored in the system, along with the relevant metadata. In the current version, the documents are loaded by copying them in a directory, but we plan to generalize it by having a repository that supports the linked data container API. The repository will then be maintained by the catalog maintain-

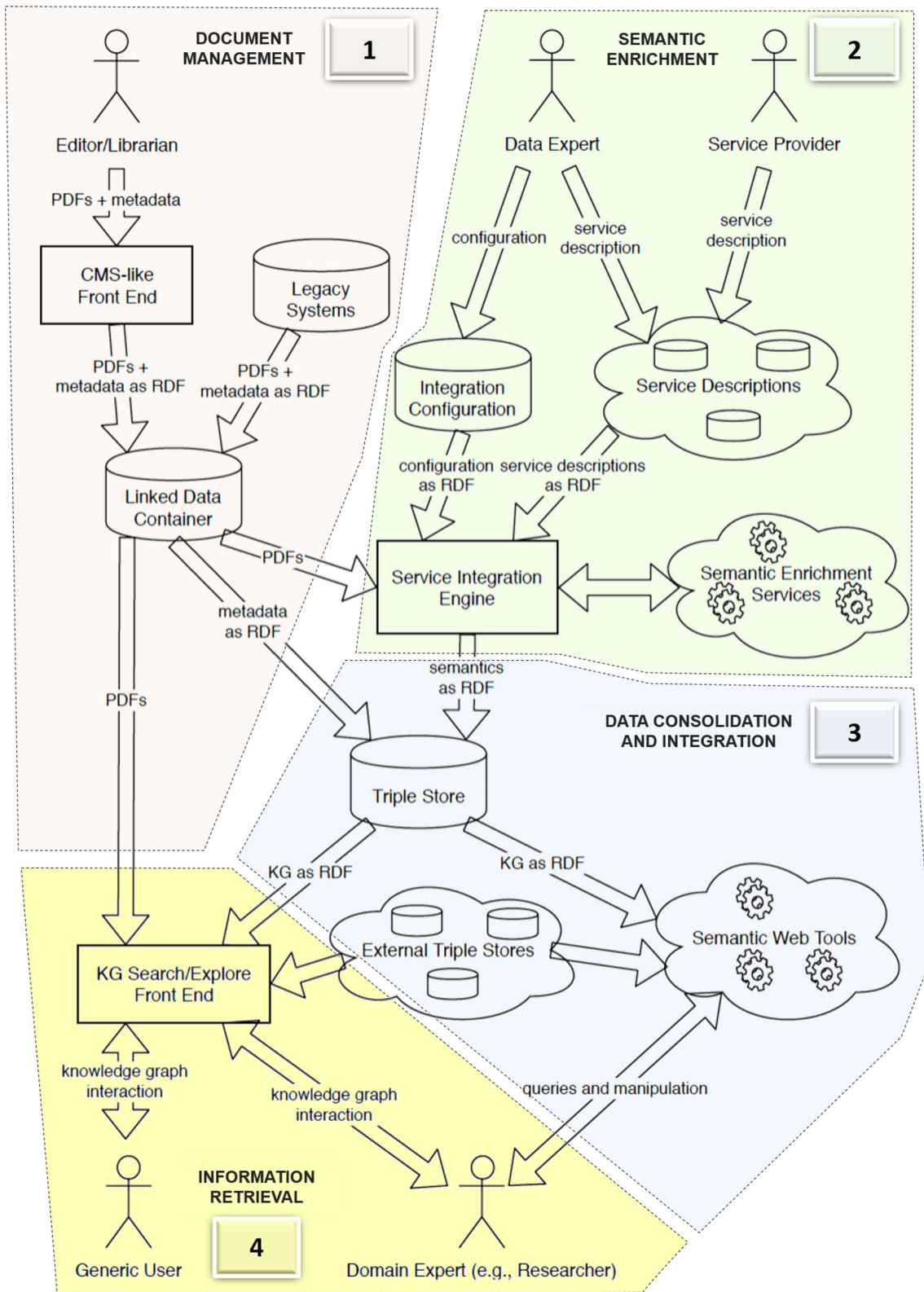


Fig. 2 Diagram describing the flow of data in the system

ers (e.g., editors or librarians) through a dedicated front-end application. It will also be possible to connect it directly with existing databases or systems for automatic content insertion or update.

In the second step, the documents analyzed by a set of semantic enrichment services give as output the knowledge extracted from the content expressed using existing models and KGs. Currently, a service providing entity extraction is called. The result is a set of the recognized entities (identified as DBpedia⁵ resources) alongside the document's point in which the entity was found. An adapter converts this information to RDF to be later integrated with existing metadata and the DBpedia KG.

In the third step, both the metadata coming from the linked data container and the knowledge extracted in the previous step are stored in a triple store as an integrated KG. Due to the distributed nature of linked data, relevant additional external data may be either added to the triple store in this step or kept separated and accessed on demand when needed.

Finally, the "Information Retrieval" functional area (number 4 in Fig. 2) refers to the actual usage of the KG to search and explore the corpus by generic users as well as domain experts. Users can use a web-based front end offering the visual user interface described in Sect. 5.3.

The front end is able to integrate on the fly data from the local triple store and other linked data sources. In the specific use case, the data from DBpedia are integrated on demand, so that the explored KG is a virtual graph obtained by merging the local KG with the DBpedia KG. Furthermore, the local data being in a triple store direct access through a SPARQL endpoint can be enabled, thus providing expert users with a means to perform advanced queries and further integration.

5.2 The data model

The data gathered in the process described in Sect. 5.1 are stored as a knowledge graph, which will be referred to as ARCA Knowledge Graph.

It is represented using RDF and describes:

- Information extracted automatically during the knowledge extraction process from books;
- Existing metadata associated with books.

The ARCA KG makes use of multiple vocabularies, providing a set of classes and properties to describe the given domain.

Tables 1 and 2 list, respectively, classes and properties employed to define ARCA KG. Listing 1 shows a fragment of RDF describing some of the information associated with a book.

Listing 1 RDF fragment

```

1 lermabook:DE000059 a arca:Book ;
2   rdfs:label "Ara Pacis Augustae." ;
3   dc:title "Ara Pacis Augustae." ;
4   dc:language "it" ;
5   dc:date 1986;
6   dc:about "In occasione del restauro della
7     fronte orientale." ;
8   schema:ISBN "9.78888E+12" ;
9
10  schema:author
11    author:la_rocca_eugenio,
12    author:zanardi_bruno,
13    author:ruesch_v;
14
15  dc:subject
16    metadata_subject:restauro_conservazione
17    ;
18
19  dct:terms:type
20    metadata_type:mostre_cataloghi ;
21
22  dct:terms:temporal
23    metadata_age:eta_classica ;
24
25  foaf:depiction lermabookimg:DE000059.jpg ;
26
27  arca:concept
28    <http://dbpedia.org/resource/Aeneas> ,
29    ...
30    <http://it.dbpedia.org/resource/Roma_(
31      citt%C3%A0_antica)> ;
32
33  arca:top_concept
34    <http://dbpedia.org/resource/Augustus>
35    ,
36    ...
37    <http://dbpedia.org/resource/Altar> .
38
39  author:la_rocca_eugenio a dbo:Person ;
40  rdfs:label "La Rocca Eugenio" .
41
42  author:zanardi_bruno a dbo:Person ;
43  rdfs:label "Zanardi Bruno" .
44
45  author:ruesch_v a dbo:Person ;
46  rdfs:label "Ruesch V." .
47
48  metadata_subject:restauro_conservazione a
49  madsrdf:Topic ;
50  rdfs:label "restauro conservazione" .
51
52  metadata_type:mostre_cataloghi a madsrdf:
53  GenreForm ;
54  rdfs:label "mostre cataloghi" .
55
56  metadata_age:eta_classica a madsrdf:Temporal ;
57  rdfs:label "eta classica" .
58
59  <http://dbpedia.org/resource/Aeneas> a arca:
60  Concept ;
61  skos:broader <http://dbpedia.org/resource/
62  Kings_of_Alba_Longa> ;
63  rdfs:label "Enea" .
64
65  <http://dbpedia.org/resource/Symbol> a arca:
66  Concept ;
67  skos:broader <http://dbpedia.org/resource/
68  Semiotics> ;
69  skos:broader <http://dbpedia.org/resource/
70  Semiotica> ;
71  skos:broader <http://dbpedia.org/resource/
72  Communication_design> ;
73  skos:broader <http://it.dbpedia.org/
74  resource/Sociologia_della_cultura> ;
75  rdfs:label "Simbolo" .
76
77  snippet:DE000059_SNI1 a arca:Snippet ;
78  rdfs:label "Augusto";
79  dc:description "D'altronde lo stesso
80  Augusto era conscio della sua
81  posizione";
82  arca:containEntity
83    <http://dbpedia.org/resource/Augustus>;
84  arca:intoBook
85    <http://www.lerma.it/index.php?pg=
86    SchedaTitolo&key=DE000059>.

```

⁵ <https://wiki.dbpedia.org/>.

Table 1 Classes

Name	Description
<i>From arca namespace</i>	
arca:Book	A “Book” is a PDF document analyzed and inserted into the ARCA system. The use case included books on ancient Roman history, but the ontological skeleton described here is applicable to any subject area
arca:Snippet	A “Snippet” is a text fragment which represents the textual contexts of the document in which a concept has been found
<i>From other namespaces</i>	
dbo:Person	A person (alive, dead, undead, or fictional)
madsrdf:Temporal	Describes a resource whose label represents a time-based notion
madsrdf:Topic	Describes a resource whose label represents a topic
madsrdf:GenreForm	Describes a resource whose label is a genre or form term. For example, biographies, catechisms, essays, hymns, or reviews, daybooks, diaries, directories, journals, memoranda, questionnaires

The data model incorporates a new vocabulary, described below, as well as the following existing vocabularies.

*SKOS*⁶ is a common data model for knowledge organization systems such as thesauri, classification schemes, subject heading systems, and taxonomies. The property `skos:broader` is adopted to define hierarchical relations among concepts. See lines 54 and 58–61 of Listing 1.

*FOAF*⁷ provides terms for describing people and organizations, documents associated with them, and social connections between people. The property `foaf:depiction` is used to associate the books with their cover image. See line 23 of Listing 1.

*SCHEMA*⁸ provides to mark up website content with metadata about itself. Properties from SCHEMA are used to connect the authors and the ISBN codes to the books. See lines 7 and 9 of Listing 1.

*MADS/RDF (Metadata Authority Description Schema in RDF)*⁹ is a data model for authority and vocabulary data used within the library and information science (LIS) community, which is inclusive of museums, archives, and other cultural institutions. Here classes from MADS/RDF have been adopted to identify different classifications of books present in meta-data. See lines 44, 47, and 50 of Listing 1.

*DC (Dublin Core)*¹⁰ is a metadata vocabulary used by many libraries. Properties from DC are used to connect the books’

title, date of publication, language, and abstract to the books. See lines 3–6 of Listing 1.

*DBO (DBpedia Ontology)*¹¹ is the ontology used within DBpedia. Here Person class from DBO has been adopted to define books’ authors. See lines 35, 38, and 41 of Listing 1.

Documents are defined by the ARCA class `arca:Book` (cfr. Table 1; line 1 of Listing 1). The metadata concern:

- The title, language, publication date, and abstract of each book (lines 2–6);
- The authors (see `schema:author` lines 10–12, 35, 38, and 41);
- The topic of the book (see `dc:subject` lines 15 and 44–45);
- The type of book (see `dcterms:type` lines 18 and 47–48);
- The era of which the book narrates (see `dcterms:temporal` lines 21 and 50);
- The cover of the book (see `foaf:depiction` line 23);

The information extracted automatically concerns all the concepts contained in each book (see `arca:concept` line 25) and the main ten concepts that describe a book (see `arca:top_concept` line 30) together with the position of the character of beginning and end in the text where the concepts were extracted. This last information is used to generate *snippet* resources (lines 64–70), having type `arca:Snippet` and associated with a text context (`dc:description` line 66) of the extracted concept (`arca:containEntity` line 67) from a specific book (`arca:intoBook` line 69).

⁶ see <https://www.w3.org/TR/swbp-skos-core-spec/>.

⁷ Friend Of A Friend—see <http://xmlns.com/foaf/0.1/>.

⁸ Schema.org—see <https://schema.org/>.

⁹ MADS/RDF—<https://id.loc.gov/ontologies/madsrdf/v1.html>.

¹⁰ Dublin Core—see <https://dublincore.org/>.

¹¹ DBpedia Ontology—see <https://www.dbpedia.org/resources/ontology/>.

Table 2 Properties

Name	Description	Domain	Range	Cardinality
<i>From arca namespaces</i>				
arca:concept	Represents the concept of a document	arca:Book	owl:Thing	Multiple
arca:top_concept	Represents the top concept of a document	arca:Book	owl:Thing	max(10)
arca:containEntity	Refers to the snippet for the extracted concept	arca:Snippet	owl:Thing	Functional
arca:intoBook	Refers to the document in which the snippet appears	arca:Snippet	arca:Book	Functional
<i>From other namespaces</i>				
rdfs:label	Represents the human-readable name of a given resource	rdfs:Resource	rdfs:Literal	Functional
dc:date	A point or period of time associated with the publication date of the document	rdfs:Resource	rdfs:Literal	Functional
dc:language	A language of the document	rdfs:Resource	dbo:Language	Functional
dc:title	The title of a document	rdfs:Resource	rdfs:Literal	functional
dc:about	further specifications on the document	rdfs:Resource	rdfs:Literal	Functional
dc:subject	A topic of the document	arca:Book	madsrdf:Topic	Functional
dc:description	The description used to specify the content of a Snippet	arca:Snippet	xsd:string	Functional
dcterms:type	The genre of the document	arca:Book	madsrdf:GenreForm	Functional
schema:ISBN	The ISBN of the book	rdfs:Resource	xsd:string	Functional
schema:author	The author of the document	arca:Book	dbo:Person	Multiple
dcterms:temporal	Temporal characteristics of the document	arca:Book	madsrdf:Temporal	Functional
foaf:depiction	Represents the cover image of a book	arca:Book	foaf:Image	Functional
skos:broader	Explicit expression of the relationships between and among concepts	arca:Book	owl:Thing	Multiple

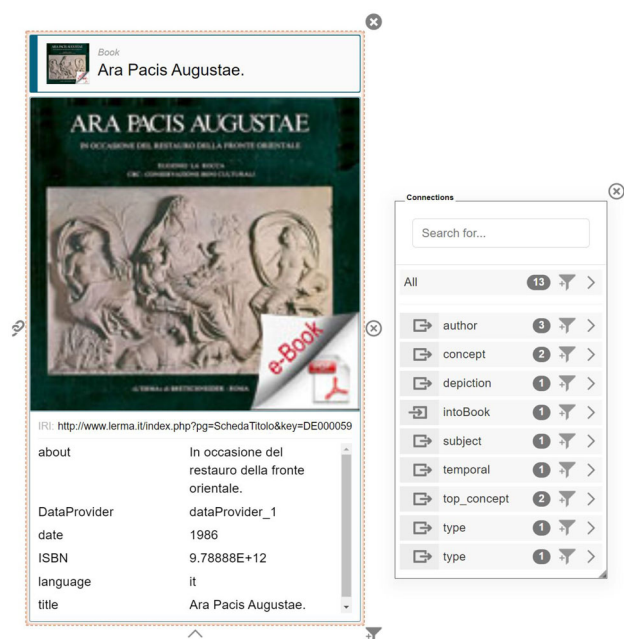


Fig. 3 Visual exploration of properties associated with the resource `lermabook:DE000059` in the RDF fragment in listing 1

Four dedicated namespaces have been defined to build unique IRIs from categories in the existing metadata. In the Listing 1, they are associated with the following prefixes:

- `author`: for authors;
- `metadata_subject`: for topic;
- `metadata_type`: for book genres;
- `metadata_age`: for historical time periods.

Figure 3 shows some of the properties described in the RDF fragment, as part of the user interface described later. Figure 4 shows another view of the user interface, in which part of the RDF graph described in the fragment is represented visually. The `arca:Snippet` class, described in Table 1, is illustrated in Fig. 6.

5.3 The user interface

The visual user interface¹² is composed of two main components (see Fig. 5). The first component contains the visualization and search of the entities contained in the KG. It is a customized version of the Ontodia workspace (briefly described in Sect. 2.3). The second component shows the list of documents associated with the selected entity, offering further interaction.

¹² <http://arca.diag.uniroma1.it:5000>.

5.3.1 Exploration of the knowledge graph

The knowledge exploration component (see part 1 of Fig. 5) has the following features.

Searching graph entities The left panel enables search for entities in the knowledge graph, corresponding in the use case mainly with entities from DBpedia. For example, typing “Rome” the user gets all the entities containing that string in their label. One or more of the returned entities (e.g., the one corresponding to Rome’s city) may be loaded to the graph navigation panel through drag and drop.

Knowledge graph navigation The central panel allows the user to navigate the KG. Starting from any shown entity, its connections, i.e., RDF triples in which the given entity is subject or object, can be expanded (hence adding the connected entities to the graph). Rather than expanding all the connections the user may select specific RDF properties (e.g., *birthplace*). Figure 3 shows the box with expanded information about a book along with the box to chose connections by RDF property. Furthermore, the connections among shown entities are shown by default, as they may be of interest. The navigation panel is coordinated with the document list panel (described below and shown in part 2 of Fig. 5) so that the latter shows the list of documents which include as topic the entity currently selected in the former.

Documents as entities Apart from being shown in the document list panel, documents can also be explored as entities themselves in the KG exploration.

They are linked to their topics by two types of semantic connections: *concept* for any entity found in the text, *top concept* for the ones recognized as main topics for that text. This choice enables further ways to interact with the system:

- Starting from a document, to explore its topics and then possibly other documents from them (e.g., in Fig. 5, from the book *The Tale of Cupid and Psyche* to the topic *Rome* and then to the book *Scutulata Pavimenta*);
- From shown entities, to visualize which documents are about two or more of them (e.g., in Fig. 5, the book *The Tale of Cupid and Psyche* is both about *Rome* and, specifically, about *Castel Sant’Angelo*).

Kinds of entities Different colors are used as an aid to distinguish three broad sets of entities:

- DBpedia entities not found in the corpus are in blue;
- DBpedia entities found at least once in the corpus are in green;
- Documents are in red.

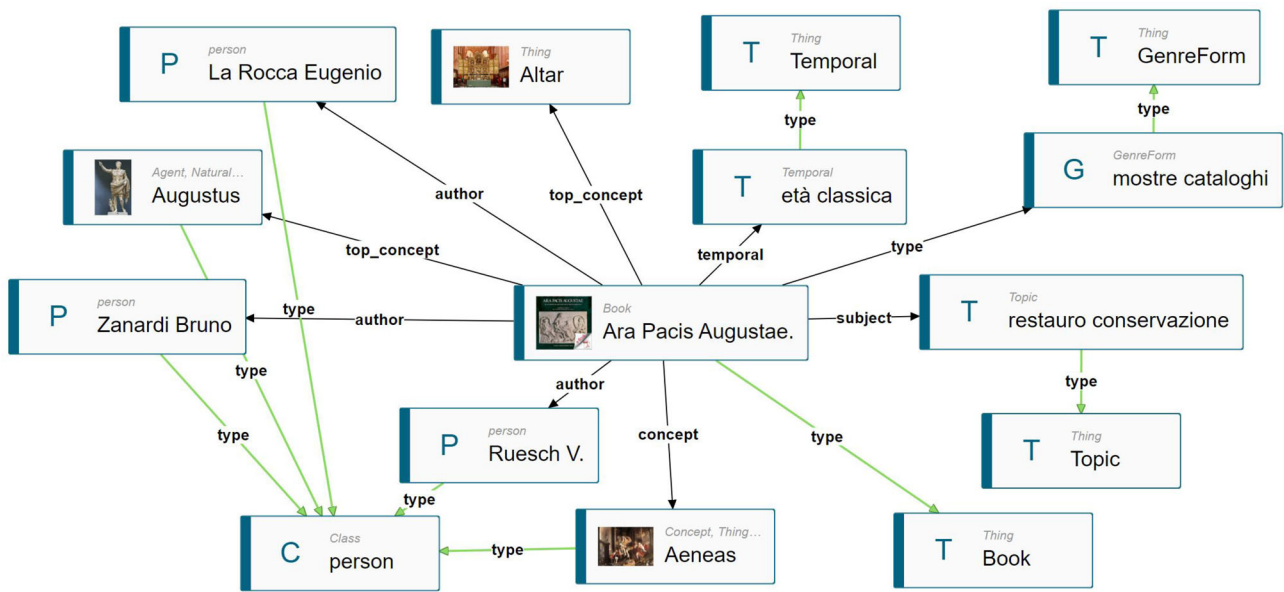


Fig. 4 Visual exploration of the RDF fragment in Listing 1

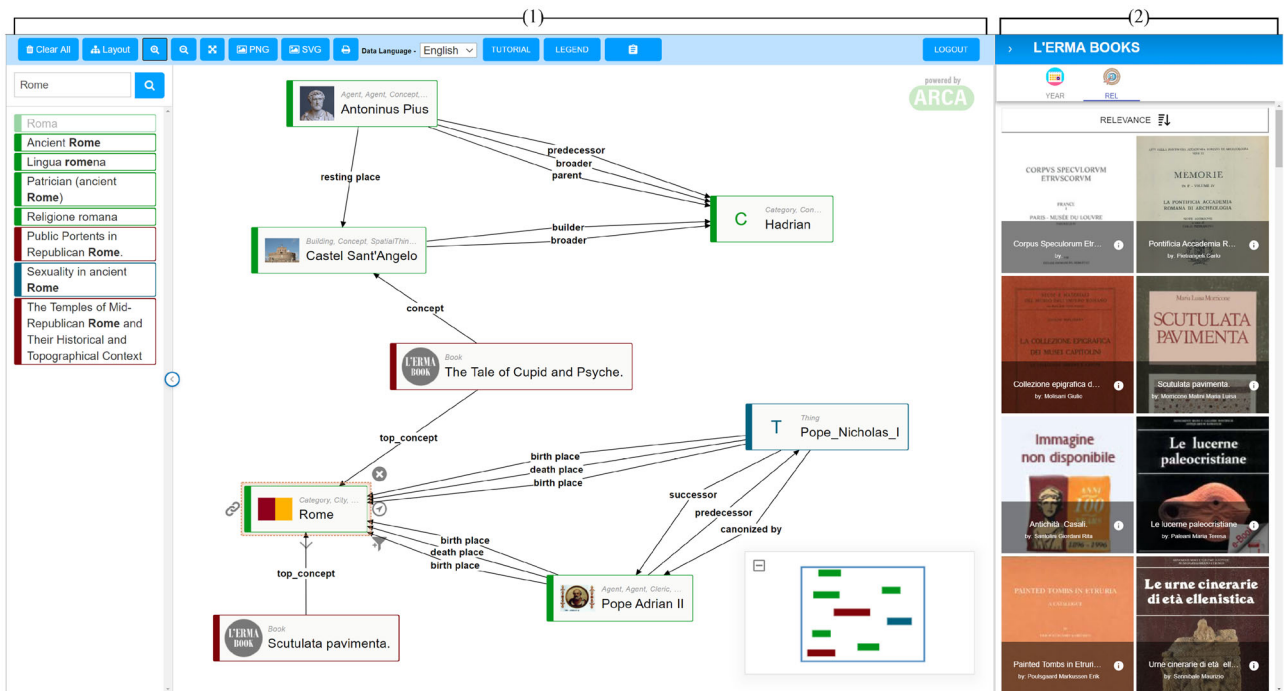


Fig. 5 User interface

5.3.2 Document list

The document list panel (part 2 of Fig. 5), which can be shown or hidden as needed, shows the list of documents associated with the entity currently selected in graph exploration panel, i.e., the documents whose extracted entities include that one. The documents may be shown ordered by year of publication or by relevance (if for that document it is a *main topic* or just

a *topic*). By clicking on the *info* button of a book, a modal window with further information on the document is opened (see Fig. 6). The information includes the list of *snippets*, i.e., all the textual contexts of the document in which the selected concept has been found.

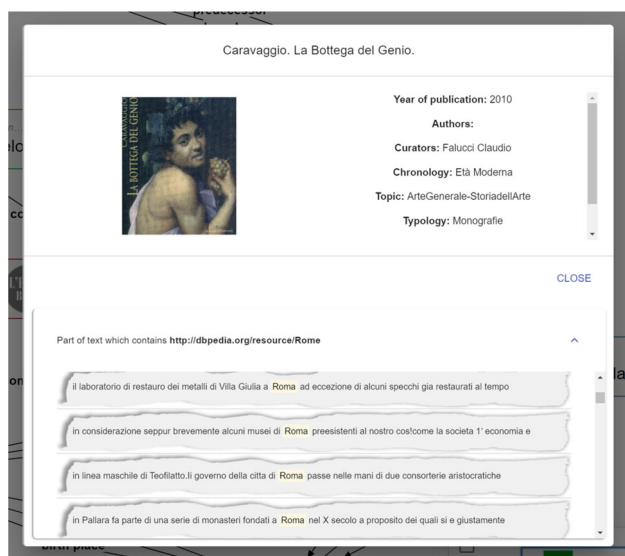


Fig. 6 Sentences

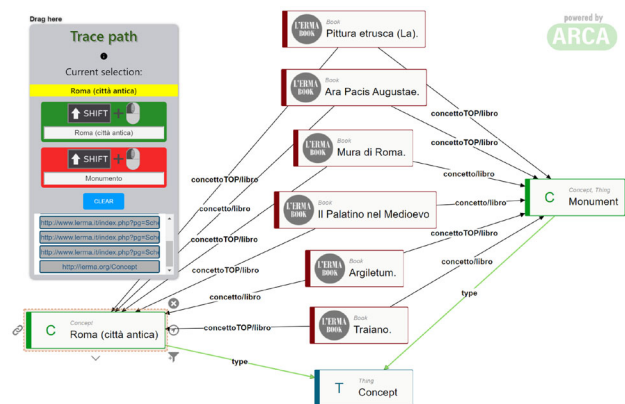


Fig. 7 Trace path

5.3.3 Trace path

The tool shows the connections between two selected entities. Thanks to the complex queries that can be processed on the knowledge graph, this tool identifies all the books connected to the first selected entity (in the case of Fig. 7 “Ancient Rome”) and all the books connected to the second selected entity (“Monument”) and makes an intersection on the two sets, showing only the books and links in common.

5.4 Implementation

For the triple store at the core of the system, we use Blazegraph,¹³ while the pipeline which builds the KG is developed in Python.

¹³ <https://www.blazegraph.com/>.

The web front end is developed using the React framework¹⁴ for modularity. It includes customized components from Ontodia as well as components built from scratch. The code is maintained in a public repository on GitHub.¹⁵

In our use case, some books do not exist natively in electronic format, for which the scanned pages go through the OCR of the software ABBYY FineReader Pro 15.¹⁶ For semantic enrichment, we are using the external *entity extraction* (NERL) web service offered by the Dandelion API¹⁷ which offers several text analysis services for many languages: entity extraction, text similarity, text classification, language detection, and sentiment analysis. Dandelion relates segments of the input text to resources in Wikipedia, along with a confidence value. Nevertheless, given the flexibility of the ARCA service integration mechanism, the system is not tied to this specific service.

5.5 Choices and motivations

In Sect. 5, the sub-elements of the system have been described in detail. In this final subsection, we want to deepen the purposes that led to the design and development of each system element.

Following, we list the choices adopted in the development of the system and its sub-elements that support the motivations described in Sect. 1 and the requirements in Sect. 3.

- The system’s modular design (cf. Sect. 5.1 *Software Modules*), which allows easy management and maintenance of a ready-to-use system. Thanks to the modularity, each module independently manages different phases of knowledge extraction, semantic enrichment, entity linking, connection to other knowledge bases, and complex queries called up through intuitive and usable interface components.
- The search bar (cf. Sect. 5.3.1 *Searching graph entities*), which allows to search for a topic and to:
 - Find concepts consistent with what is researched and extrapolated directly from the documents of the corpus of texts inserted in Arca;
 - Find documents whose title is semantically consistent with what is sought (cf. Sect. 5.3.1 *Documents as entities*);
 - Find other semantically coherent resources deriving from the knowledge graphs integrated into the search system (cf. Sect. 5.3.1 *Kinds of entities*);

¹⁴ <https://reactjs.org/>.

¹⁵ <https://github.com/EleonoraAI/ARCA-frontend>.

¹⁶ <https://www.abbyy.com/en-eu/finereader/>.

¹⁷ <https://dandelion.eu/>.

- The graph navigation mode (cf. Sect. 5.3.1 *Knowledge graph navigation*), which supports the user in accessing connected resources and discovering new information following the philosophy of serendipity.
- Direct access to the list of documents (belonging to the digital library included in Arca) in which the topic of interest is discussed.
- The possibility of finding topics in common with different resources (cf. Sect. 5.3.3 *Trace path*).

6 Evaluation

The system has been tested in the context of a specific use case: exploration of the book catalog of medium-sized publishing house, specialized in classical antiquity. The anticipated final users of the tool can be roughly classified in two categories:

- Domain experts who may adopt a new approach to search and discover resources in the context of their research;
- Curious people who want to explore new topics.

ARCA's evaluation process lasted two years and was characterized by three phases:

- An evaluation of the extracted data, from the point of view of quality and usefulness, with the help of domain experts;
- A small-scale qualitative user-based evaluation of the tool with a some researchers of the field;
- A larger and richer user-based evaluation of the tool, both on its own and in comparison with other existing solution, which involved both students and researchers of the field.

In the first two evaluation phases, discussed in previous work [8], we focused on analyzing the limits and margins for improvement in ARCA with the involvement of researchers experts in the relevant domain, who had also participated in the design process. It was possible to identify problems in the extraction of topics from the books and get feedback to improve the whole system.

In this paper, we focus on the third phase and discuss the results obtained. This phase involved 30 users and included both a comparative evaluation with other three tools offering similar functionality (Lerma, Torrossa, and Yewno) and a specific evaluation of ARCA on its own. Both parts of the evaluation are task-based and contain questions designed to evaluate multiple factors of the user experience and elicit perceived strengths and limitations of the tool. The following subsections describe in detail this experiment: the setup, the obtained results, and their discussion. The raw data gathered

from the test are also publicly available online on Zenodo [7].

6.1 Setup

As anticipated, the third phase of evaluation has been a user test involving 30 users, who were students and researchers in the field of classical antiquity. Participants had different levels of academic education: 9 secondary school qualification, 5 bachelor's degree, 7 master's degree, 7 PhD.¹⁸

Choosing students and researchers of the specific field considered in the use case was crucial to the experiment: it allowed to assume at least some level of interest for the considered topics and provided a way to partially predict the level of relevant background knowledge based on the reached academic qualification.¹⁹ Albeit all the involved users study the field at an academic level, their diverse level of academic education partially covers the requirement of H2 to test the tool with users of varying levels of knowledge of the domain.

In the third evaluation phase instead, the tool was novel for the users and they had varying level of domain expertise. In order to carry on a comparative evaluation, we planned a task-oriented setup in which the users were able to access four different tools on equal grounds, in random order, and remaining unaware of the fact that one of the tools (ARCA) is developed by us.

The comparative evaluation included two tools providing simple text-based search (Lerma and Torrossa) and two tools providing search enhanced by semantics (Yewno and ARCA).

The task-oriented comparative evaluation was compounded by a part of the evaluation focusing on specific aspects of ARCA. This part was scheduled in the end, to preserve the fairness of the comparative evaluation.

Just when we were about to start to plan the last phase of evaluation, the COVID-19 pandemic broke out and it was not possible to carry out the third phase of the evaluation in presence. For this reason, we have re-designed the process to make users autonomous and able to perform the required activities and answer questions while using the system from home.

Given the richness and complexity of the evaluation, the goal of comparing multiple tools, and the necessity for the users to be able to perform the whole process autonomously, we put a lot of effort in a carefully designed user interface

¹⁸ Two replied *other*.

¹⁹ We are aware the relationship between education and level of knowledge of a field is far from straightforward, but other measures have also limits (self-reporting is highly subjective and directly testing the knowledge with limited available time would also be very problematic). Furthermore, we use the academic level mainly as a way to ensure some diversity among the users, rather than an exact proxy of knowledge of the domain.

that was able to guide the users step by step through each required activity and each question.

Based of existing literature of the evaluation of search tools [30], we identified multiple measures belonging to the two following categories:

- Subjective self-reported measures given by users, like quantitative answers on Likert scales or qualitative answers to open questions;
- Objective measures, such as the log of the events from the user interface, the time to complete a task, and the words searched.

For the organization of the questions and the activities to be carried out by users during the test, we follow the scheme proposed by Kelly [22], who propose to organize the questionnaires to evaluate interactive information retrieval systems (IIR) in five parts: demographic (e.g., gender, age), pre-task (e.g., prior knowledge of the system or topic), post-task (e.g., task satisfaction), post-system (e.g., the overall experience of interacting with an information system), and exit (e.g., cross-system comparisons of ease of use or preference). Below are the categories of questions used in this work.

User info:

- Demographic (gender, age, level of education);
- Pre-task (prior knowledge of relevant topics).

System evaluation (phases repeated for each of the four compared systems):

- Task (the user is asked to navigate the system in order to retrieve a piece of information);
- Post-task (quantitative evaluation of efficiency, effectiveness, and satisfaction to measure the usability).

ARCA system in-depth evaluation:

- Task (the user is asked to navigate the system in order to retrieve a piece of information);
- Post-task (quantitative evaluation of efficiency, effectiveness, and satisfaction to measure the usability);
- Post-system (the overall experience of interacting with ARCA system).

When planning the evaluation, we paid attention to respecting the reliability and validity of the results.

To ensure the evaluation results' reliability, we established the following criteria:

- The whole process was executed through a self-administered web questionnaire, which ensured a level of distance between researcher and participant;
- Users started directly with the comparative evaluation of the search systems without ever discussing ARCA or its characteristics before;
- The four compared systems were presented in the same way and in a different order for each user group, asking them at the end of each navigation for feedback on their usefulness, satisfaction, and ease of use.

To ensure the validity of each evaluation request's results, we established a clear objective on what to evaluate and with which metrics to do it best. For example, to measure usability, we measured efficiency, effectiveness/usefulness, and satisfaction with Likert scales with a rating of one to five.

Key Factors. During the test, the proposed activities and subsequent questions were aimed at investigating the user interaction experience with the interface. In particular, the questions asked at the end of the activity tried to extrapolate an assessment of the key factors listed below.

- *Satisfaction* Namely, investigate how good is the system for the research objective, intended as the discovery and retrieval of information in a digital library.
- *Effectiveness* How effective is the system in showing users the information.
- *Support* How much the system supports the user during the searches and exploration of a digital library.
- *Usefulness* It directly impacts the usage of any system [40]. Therefore, usefulness can be considered a critical usability factor.
- *Learnability* Analyze the way users adopt and get familiarized with the systems.

Use case Inside ARCA 112 books have been inserted concerning the history of Roman archeology.

Users Fifty-two people were selected for the test, including students and researchers from the domain of books contained within ARCA.

Communication All communications between the ARCA team and the evaluating users took place by email.

The first email sent gave each user their login credentials and asked them to perform the three phases of the test:

- Comparison of the four platforms proposed for searching books;
- In-depth evaluation of a single search platform (ARCA);
- Reply to a set of open questions on the whole process.

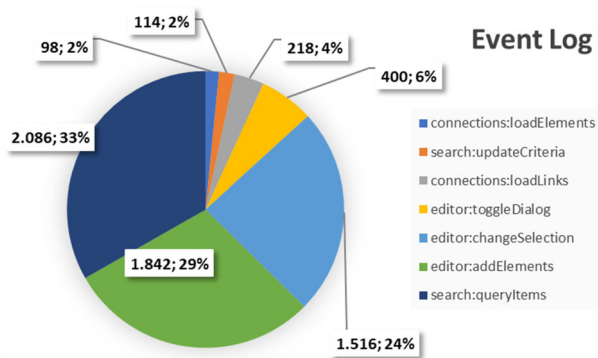


Fig. 8 Percentage of event by type of action

6.2 Results

The test, started in December 2020 and lasted for a month, is composed of four parts that were performed in this order:

1. The collection of personal data and self-assessment on knowledge background;
2. The comparison test of four book search platforms: ARCA, Yewno, Lerma, Torrossa;
3. The evaluation test of ARCA;
4. The test with open questions to express final evaluations.

This compilation order was chosen to respect the impartiality of judgment during the tests’ execution: to not put ARCA in a more favored position than the other search systems.

Regarding the number of participants who have completed each test part, we have:

- Twenty-five users who completed all four parts;
- One user who has completed the first three test parts;
- Four users who completed the first two parts.

On average, users completed the three parts of the test for more than an hour.

The event log, traced during the evaluation, revealed that most of the interactions with the interface. We consider all interaction actions with the user interface, excluding those that do not significantly affect the flow of the interaction (such

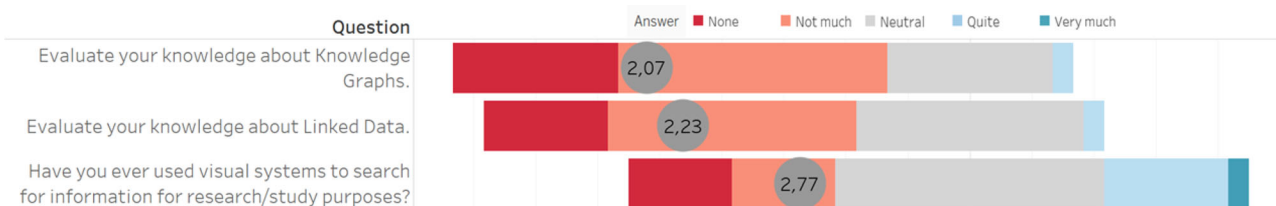


Fig. 9 User background

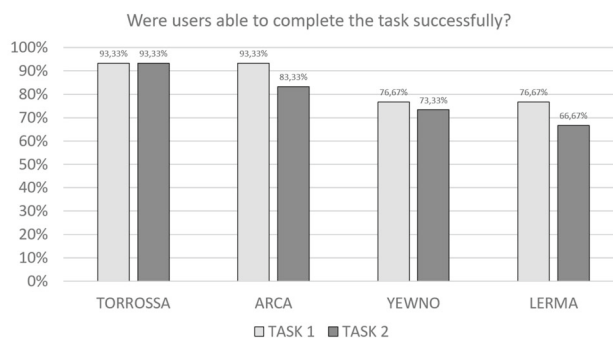


Fig. 10 Completion of the comparative evaluation tasks

as clicking on the buttons with the tutorials and consulting the search history). The 33.25% of user interactions concerned the search for terms (see Fig. 8); 29.36% involved adding concepts to the whiteboard; 24.16% concerned the selection of elements; 6.38% concerned the elimination of concepts from the dashboard; 5.03% (“connections:loadLinks” and “connections:loadElements”) concerned the exploration of the selected concepts; finally, 1.82% concerned modifying the searched keyword in the search bar.

The first part of the evaluation contains personal data and general information about the user’s background on information visualization, knowledge graph, and visual interfaces for querying and interacting with data. Figure 9 shows, with a Likert scale, that general knowledge on the required topics is very low.

Users who participated are mostly students and researchers in the humanities field. They are aged between twenty and forty-four, in particular 55% are between twenty and twenty-nine, 41% are between thirty and thirty-nine, and 4% are over forty. Regarding the gender 49% are men and 51% women. Regarding education, 37% have a diploma, 40% have a degree, and 23% a P.h.D.

6.2.1 Comparative evaluation

After assigning the user the first search platform to test (randomly between Arca, Yewno, Lerma, and Torrossa) and introducing him to its use, the task was to search for two books about two Roman hills. In Fig. 10, the task one, associated with each search platform, indicates the search of the

first book which talk about two Roman hills, while two indicate the search of the second book.

This task was created to encourage the user to follow multiple search paths without forcing him to a linear, and a sequential path to better evaluate the search platform's usefulness, as explained in the research carried out by Liu et al. [23].

Not all users were able to complete the task, that is, to find the two required books. We checked the positive results to verify that the books' titles, indicated by users, actually contained two Roman hills. The return is that all users, who managed to complete the task, correctly indicated the books' titles.

After completing the task, the user was asked five quantitative questions with a Likert scale. The numerical score chosen for the likert scale is from one to five which represents the corresponding qualitative evaluation from "None" to "Very Much." Figures 11 and 12 show the distribution of the qualitative scores given to each research platform used: Lerma, Torrossa, Arca, and Yewno.

For the statistical analysis of the results, we conducted a one-way ANOVA to determine any statistically significant differences between the means of given scores to evaluate the key factors of satisfaction, effectiveness, support, usefulness, and learnability for the four interfaces. Since the result of analysis of variance (F -value) is significant, it is necessary to use a Post-Hoc test [31], to identify which samples are different because the ANOVA test shows only that there is a difference between the means but does not indicate which means are different. The t-test is the selected Post-Hoc in this work. We select $p < 0.05$ as our significance threshold.

Learnability We evaluated the *learnability* factor with the following post-task question: "How easy was it to complete the task?" The one-way ANOVA revealed that there was a statistically significant difference in mean value of ease in completing the research activity between at least four groups ($F(4.60, 1.09) = [4.20], p = 0.01$).

T-Test for multiple comparisons found that the mean value was significantly different between:

- Arca and Yewno ($p = 0.01$, statistics = 2.86)
- Lerma and Torrossa ($p = 0.03$, statistics = -2.29)
- Torrossa and Yewno ($p = 8.32 \times 10^{-4}$, statistics = 3.73)

At the same time, there was no statistically significant difference between:

- Arca and Lerma ($p = 0.17$, statistics = 1.39)
- Arca and Torrossa ($p = 0.33$, statistics = -1)
- Lerma and Yewno ($p = 0.06$, statistics = 1.93)

The results show no statistical significance between the ease of use of Arca compared to that of Lerma and Torrossa. This result underlines that although Arca is based on a graph information visualization system, which the user is less accustomed to, it is considered as easy as Lerma and Torrossa, systems based on schematic and tabular navigation. Furthermore, Arca is significantly better in learnability than Yewno, although both are based on displaying the results via a graph. However, Yewno does not allow the interactive exploration of the graph nodes, and we believe that this is the reason for its lower ease of use compared to that of Arca.

Support We evaluated the *support* factor with the following post-task question: "How supported do you feel?" The one-way ANOVA revealed that there was a statistically significant difference in mean value of support in conducting the research activity between at least four groups ($F(10.14, 0.83) = [12.20], p = 10^{-6}$).

T-Test for multiple comparisons found that the mean value was significantly different between:

- Arca and Lerma ($p = 1.54$, statistics = 6.01)
- Arca and Torrossa ($p = 0.02$, statistics = 2.40)
- Arca and Yewno ($p = 8.32 \times 10^{-4}$, statistics = 3.72)
- Lerma and Torrossa ($p = 10^{-3}$, statistics = -3.63)
- Lerma and Yewno ($p = 0.02$, statistics = -2.45)

At the same time, there was no statistically significant difference between:

- Torrossa and Yewno ($p = 0.07$, statistics = 1.88)

The tests show that users felt more supported by the Arca platform, with a significant statistical difference compared to the other three platforms tested. We hypothesize that this derives from the users' need (detected and satisfied) for more significant support because Arca has distinctive features compared to other tools for searching for information in a digital library. For example, the trace path allows users to find information common to several concepts; the snippets show the part of the text that deals with the concept explored; the graph exploration allows users to trace search paths and view connections directly on the dashboard.

Effectiveness We evaluated the *effectiveness* factor with the following post-task question: "How satisfied are you with the results you have found?" The one-way ANOVA revealed that there was a statistically significant difference in mean value of the results shown during the research activity between at least four groups ($F(8.28, 1.09) = [7.55], p = 1.17 \times 10^{-4}$). T-Test for multiple comparisons found that the mean value was significantly different between:

- Arca and Lerma ($p = 2.56 \times 10^{-4}$, statistics = 4.16)

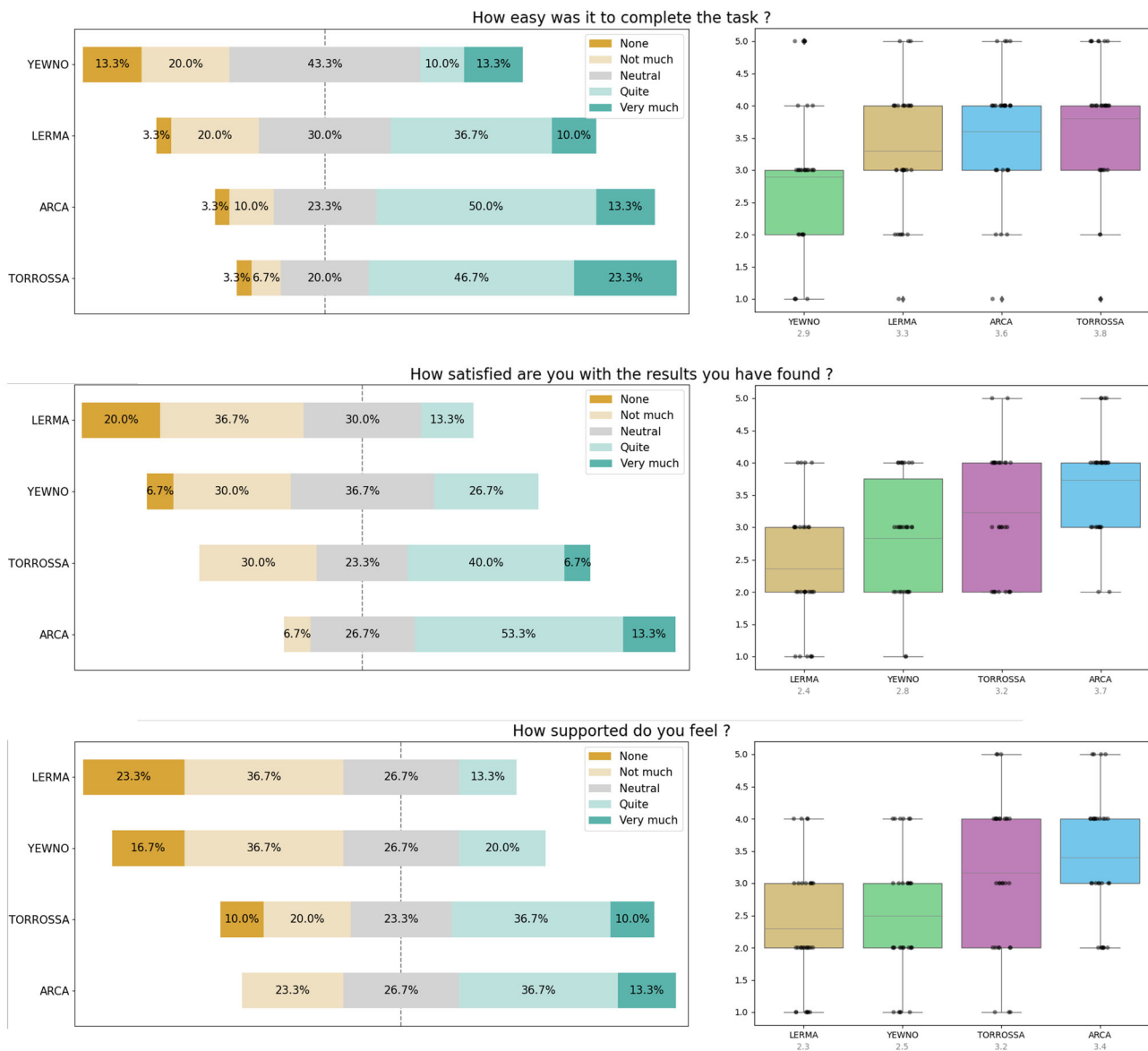


Fig. 11 Comparative evaluation: distributions and averages-part 1

- Arca and Yewno ($p = 10^{-3}$, statistics = 3.66)
- Lerma and Torrossa ($p = 4.15 \times 10^{-3}$, statistics = -3.11)
- Torrossa and Yewno ($p = 0.01$ statistics = 2.94)

At the same time, there was no statistically significant difference between:

- Arca and Torrossa ($p = 0.32$, statistics = 1.02)
- Lerma and Yewno ($p = 0.42$, statistics = -0.81)

User Satisfaction We evaluated the *satisfaction* factor with the following post-task question: “Is the information you viewed satisfactory for you?” The one-way ANOVA revealed

that there was a statistically significant difference in mean value of the satisfaction perceived by the resulting information during the research process between at least four groups ($F(7.83, 1.06) = [7.40]$, $p = 1.41 \times 10^{-4}$).

T-Test for multiple comparisons found that the mean value was significantly different between:

- Arca and Lerma ($p = 3.52$, statistics = 4.88)
- Arca and Yewno ($p = 2.61 \times 10^{-4}$, statistics = 4.16)
- Lerma and Torrossa ($p = 2.94 \times 10^{-3}$, statistics = -3.25)
- Torrossa and Yewno ($p = 0.01$ statistics = 2.97)

At the same time, there was no statistically significant difference between:

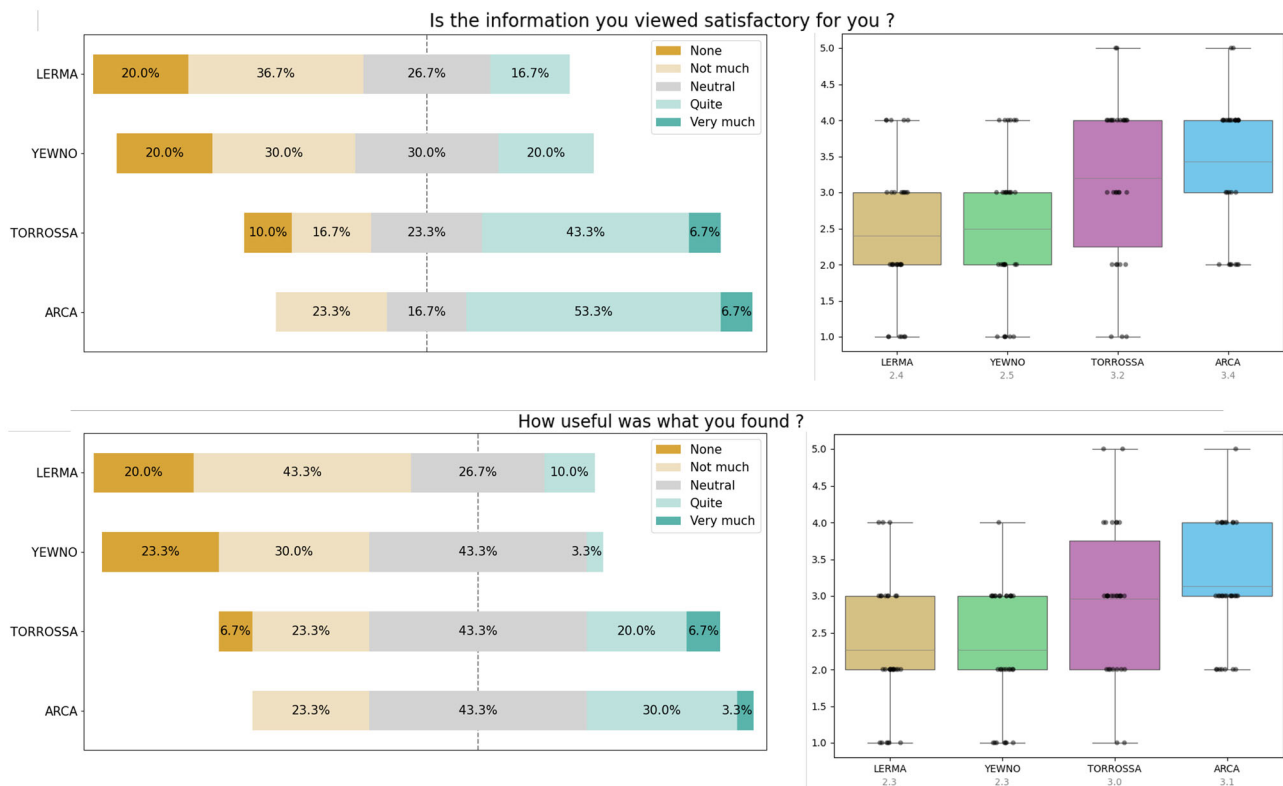


Fig. 12 Comparative evaluation: distributions and averages—part 2

- Arca and Torrossa ($p = 0.27$, statistics = 1.13)
- Lerma and Yewno ($p = 0.68$, statistics = -0.42)

As for the analysis of the satisfaction of the information shown (user satisfaction) and the results found (effectiveness), the user feels satisfied both with the use of Arca and Torrossa. On the contrary, the search results shown by Yewno and Lerma are considered less satisfactory by the user than the other two systems. We can detect from the tests that although Yewno is based on the same technologies as Arca, this does not enhance it compared to Torrossa, which is still considered better in the information displayed and the search results. We hypothesize that Arca and Yewno, based on knowledge graph, semantic search, and visualization of information on a graph, allow reaching more information and links than those Lerma and Torrossa, based on key-text search and visualization of tabular results. Despite our hypothesis, the results just commented support this for Arca, but not for Yewno. As already noted, this may be due to the explorability of the resources allowed by Arca.

Usefulness We evaluated the *usefulness* factor with the following post-task question: “How useful was what you found?” The one-way ANOVA revealed that there was a statistically significant difference in mean value of perceived usefulness by the resulting information during the research

process between at least four groups ($F(6.28, 0.81) = [7.73]$, $p = 9.4 \times 10^{-5}$).

T-Test for multiple comparisons found that the mean value was significantly different between:

- Arca and Lerma ($p = 2.47 \times 10^{-4}$, statistics = 4.18)
- Arca and Yewno ($p = 2.47 \times 10^{-4}$, statistics = 4.18)
- Lerma and Torrossa ($p = 2.9 \times 10^{-3}$, statistics = -3.25)
- Torrossa and Yewno ($p = 2.34 \times 10^{-3}$, statistics = 3.34)

At the same time, there was no statistically significant difference between:

- Arca and Torrossa ($p = 0.39$, statistics = 0.87)
- Lerma and Yewno ($p = 1.0$, statistics = 0.0)

The usefulness of the searches carried out with the Lerma and Yewno tools is significantly lower than that carried out with the Torrossa and Arca tools. Torrossa and Arca are not significantly different. We can deduce that in this case, the knowledge graph and semantic research have generated informative content as useful as Torrossa’s manual metadata.

On average, it took users 26 min to complete the comparative evaluation. In particular, they sailed on average:

- 9.4 min Arca;

- 6.6 min Torrossa;
- 5.8 min Yewno;
- 4.3 min Lerma.

Referring to Figs. 11 and 12, from the average of the scores assigned to each search platform, we derived a list of the preference. In order, users preferred the system (with a score from one [None] to five [Very much]):

- Arca with a score of 3.46;
- Torrossa with a score of 3.3;
- Yewno with a score of 2.6;
- Lerma with a score of 2.5.

The evaluation of Arca revealed that the system has good potential in:

- In producing satisfactory results;
- In supporting the user in searches;
- In being useful to the user for his research;
- In producing satisfactory results for the user.

As for ease of use, users preferred the Torrossa search platform more.

6.2.2 ARCA evaluation

For the evaluation of ARCA, we have set up research tasks with the aim of making users perform full navigation of ARCA, exploring every component and every functionality offered by the system to make the most of its potential in order to reach the required research goal. We first had users perform guided navigation of the system to make them able to know all the functions and possibilities of search and navigation, and then, we asked them for tasks and questions related to the research task just carried out.

Here are the three required tasks:

- Search and explore books about Rome in medieval times.
- Search and explore books about ancient Greek jewels.
- Research and explore books to deepen a topic, which is covered among the texts contained within ARCA.

The tasks have been chosen to leave the user the freedom to explore the system according to their creativity. To evaluate this aspect and serendipity, we asked users how favored they were at finding unexpected things (Fig. 13).

Free navigation of the system resulted in neutral utility and neutral general satisfaction level. By analyzing the next section, we will understand what the ideas for improving the system are.

6.2.3 Final questions

The previous tasks have been chosen to leave the user the freedom to explore the system according to their creativity. To evaluate this aspect and serendipity, we asked users how favored they were at finding unexpected things. Finally, three general questions were asked, shown below, along with an outline of user responses.

What are the most useful features of ARCA?

- The possibility of observing the connections between books by finding arguments in common;
- The practicality that facilitates bibliographic research;
- The transversal approach to the topics;
- The visual search;
- Semantic connections;
- Wide-ranging exploration (bibliographic and conceptual);
- Navigation of texts and concepts;
- The direct link to the catalog of books for finding the resources of interest;
- The amplification of the results;
- The type of result display that facilitates complex searches;
- The simplicity of use;
- The graphic environment, although it can be improved, has good potential;
- The interdisciplinary research of contents;
- The ability to find books on more than one subject at a time, and the fact that the search is not limited to titles;
- The possibility to organize diagrams with all the connections from basic research to peripheral publications;
- Being able to find texts with a common topic and above all with more topics in common.

What are ARCA's weaknesses?

- Difficult to use without having seen the tutorials;
- Research does not always lead to what is sought;
- Few results when searching for something specific;
- Some links are non-existent;
- The lack of filters on searches.

Are there any features that could improve the attractiveness and usefulness of ARCA?

- The introduction of components that allow more complex queries, such as the search for links in common to more than two resources;
- The inclusion of a more extensive catalog of books to expand the number of links and information;
- Investing in the graphics to make system navigation more comfortable and more intuitive;



Fig. 13 Serendipity evaluation

- The increase in the tutorials and guides to allow the user to exploit the full potential of the system.

6.3 Discussions and limitations

The system obtained a more than satisfactory performance in recommending relevant editorial products and received a high score in terms of usability; simplicity of use; user satisfaction with the results shown; consistency of the contents with the books domain of the publishing house; attractiveness of the system. Nonetheless, some users identified as an issue the relatively small amount of information contained in the internal KG (built with the concepts of 112 books and the related metadata). It is expected that when the catalog of books is numerically more significant, the chance of discovering new information and connections while browsing the KG will increase.

Furthermore, based on initial observations, it has been seen that using the system at first glance can be difficult without viewing the video tutorials. In fact, concerning the comparative assessment, users, to solve the tasks, took longer to navigate on ARCA than other tools. This fact may indicate a greater navigation complexity, but simultaneously, merged with positive feedback about usability, a greater exploratory interest. To reduce initial exploration difficulties, as a lighter alternative to video tutorials, a help component could be implemented to accompany the user in the first searches and thus make her independent in exploiting all the possibilities of exploration that the system offers.

Below we will discuss the analysis of the test results to support or re-evaluate the hypotheses elaborated in Sect. 1.2.

The current findings appear to validate the H1 hypothesis. The users have evaluated multiple aspects of their experience more positively than other tools. Users stated that they had obtained useful results for their searches and could explore and search within a texts corpus. Crucially, the tool was rated as easy to use as text-based search tools, which employ a paradigm that is certainly far more familiar to the users. In the open questions section, many users have stated that they appreciate the opportunity to explore the resources and the possibility of observing the connections between concepts.

Albeit the results relating to H1 are very promising, in a user study with limited available time is hard to evaluate advanced usage of semantic search for complex research scenarios and in-depth visual exploration of the knowledge

graph. For that purpose, an in-use evaluation with a longer time span may be designed in future. We argue that the availability of public tools for semantic-based information retrieval will allow to collect data and user feedback on how they are used and in turn enable the design of better paradigms and tools.

Regarding hypothesis H2, as anticipated when describing the setup of the user study in Sect. 6.1, the academic level has been considered as a partial indicator of the level of knowledge of the field. In that sense, if the hypothesis holds, we expect that usage of ARCA to be satisfactory across different levels of academic qualification. As already described, the users rated positively the experience with ARCA in respect to other tools. In none of the key factors, the level of education correlates significantly with the perceived quality of the experience.

The tracing of the logs shows that users with a higher education level (master's degree, doctorate) have dedicated on average 11.50% of total interactions to deepen concepts, while users with a lower level of education (diploma, three-year degree) devoted an average of 5.44%. This split of behavior hints at two different approaches to navigation:

- Find specific resources;
- Explore connections related to found resources.

It is possible that the approach to research of individual users determines this attitude in searching and exploring concepts. However, although users have appreciated the idea and the potential of the tool, the results shown so far by the platform do not fully satisfy the search wishes. In fact, in evaluating the satisfaction of the results shown, the average score expressed was 3.56 (on average satisfied) for users with a higher level of education, while 2.79 (not very satisfied) for users with a lower level of education. Regarding hypothesis H3, the generality of the implemented prototype goes in the direction of proving the generalizability of the pipeline. Furthermore, albeit not described here, part of the same pipeline is applied to another different use case, namely supporting research on ancient symbols²⁰ [6].

In conclusion, more work is required to fully evaluate the applicability of the tool in multiple contexts, but the results so far seem promising.

²⁰ <http://www.notae-project.eu/>.

7 Conclusions and future work

ARCA is an innovative system based on the visual semantic search and exploration of a text corpus. Through a knowledge graph-based navigation, the user can start from any relevant entity and reach other entities related to it, discovering in which books or articles each entity is present and evaluating which of these results are useful for their research. The user studies conducted so far confirm the amenability of the proposed system to domain experts which were able to perform non-trivial tasks of search and exploration, tasks that would be more cumbersome to execute with the search tools they are used to. Feedback gathered from users suggests that the proposed exploration mechanism tends to amplify the user experience by also offering opportunities for further study and discovery of sources, themes, and materials, which have the potential of enriching the research process with new ideas. In a comparative task-based evaluation with other tools for information retrieval on the same corpus, the users rated favorably multiple key factors of the experience with ARCA. Specifically, they rated it as easy to use as text-based search tools, notwithstanding the inherent complexity of the user interface due to richer functionality and novelty of the paradigm, and easier to use than another more static semantic-based visual tool presented to them.

Through the presented user study, a number of desiderata have been collected and they can be used to guide further development and experimentation in this context. Furthermore, in order to extend the evaluation to more users, a larger indirect observation study has been planned. In addition to the questionnaire, the analysis will be further completed with objective data gathered by tracking users' activity through interaction logs.

More work is needed to better evaluate one of the stated hypotheses and aims of the tool, i.e., if the tool can be easily applied to other use cases and what is needed to improve its generality. Furthermore, as a potential future direction of research and development, the scope of the semantic enrichment process could be broadened to other document elements, such as images and captions, enriching KG exploration.

Acknowledgements This work has been partly supported by projects ARCA (POR FESR Lazio 2014-2020 - Avviso pubblico "Creatività 2020," domanda prot. n. A0128-2017-17189), STORYBOOK (POR FESR Lazio 2014-2020 - Avviso Pubblico "Progetti di Innovazione Digitale," domanda prot. n. A0349-2020-34437), SCIBA (Domanda prot. n. 305-2020-35545 del 18/05/2020 Regione Lazio Intervento TE1 - Centro di Eccellenza Det. n. G00471 del 21/01/2020), and Social Museum and Smart Tourism (CTN01_00034_23154). Miguel Ceriani acknowledges funding from the European program PON Ricerca e Innovazione 2014-2020, within the project no. COD.AIM1852414, activity 2, line 2.1 and the project "Computational methods for the web economy," DM 1062/2021.

References

- Ahlberg, C., Shneiderman, B.: Visual information seeking: tight coupling of dynamic query filters with starfield displays. In: *The Craft of Information Visualization*. pp. 7–13, Elsevier (2003). <https://doi.org/10.1145/191666.191775>
- Ahlberg, C., Williamson, C., Shneiderman, B.: Dynamic queries for information exploration: an implementation and evaluation. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. Association for Computing Machinery, New York, NY, USA, pp. 619–626. 0897915135 (1992). <https://doi.org/10.1145/142750.143054>
- Allemang, D., Hendler, J., Gandon, F.: RDF Schema. Association for Computing Machinery, New York, NY, USA 9781450376174, (2020). <https://doi.org/10.1145/3382097.3382106>
- Baeza-Yates, R., Ribeiro-Neto, B.: *Modern Information Retrieval*, vol. 463. ACM press, New York (1999). <https://doi.org/10.5555/553876>
- Beckett, D., Berners-Lee, T.: Eric Prud'hommeaux, and Gavin Carothers. 2014 RDF 1.1 Turtle: Terse RDF Triple Language (2014). <http://www.w3.org/TR/2014/RECturtle-20140225>
- Bernasconi, E., Boccuzzi, M., Catarci, T., Ceriani, M., Ghignoli, A., Leotta, F., Mecella, M., Monte, A., Sietis, N., Veneruso, S., Ziran, Z.: Exploring the historical context of graphic symbols: the NOTAE knowledge graph and its visual interface. In: *Proceedings of the 17th Italian Research Conference on Digital Libraries, Padua, Italy (virtual event due to the Covid-19 pandemic), February 18–19, (2021a) (CEUR Workshop Proceedings)*, Dennis Dosso, Stefano Ferilli, Paolo Manghi, Antonella Poggi, Giuseppe Serra, and Gianmaria Silvello (Eds.), Vol. 2816. CEUR-WS.org, pp. 147–154. <http://ceur-ws.org/Vol-2816/short2.pdf>
- Bernasconi, E., Ceriani, M., Mecella, M.: Arca—Evaluation (2021b). <https://doi.org/10.5281/zenodo.4679175>
- Bernasconi, E., Ceriani, M., Mecella, M.: Exploring a text corpus via a knowledge graph. In: *Proceedings of the 17th Italian Research Conference on Digital Libraries, Padua, Italy (virtual event due to the Covid-19 pandemic), February 18–19 (2021c) (CEUR Workshop Proceedings)*, Dennis Dosso, Stefano Ferilli, Paolo Manghi, Antonella Poggi, Giuseppe Serra, and Gianmaria Silvello (Eds.), Vol. 2816. CEUR-WS.org, pp. 91–102. <http://ceur-ws.org/Vol-2816/paper8.pdf>
- Berners-Lee, T., James, A., Hendler, O.L.: 2001 the semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American* (2001). https://www.sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf
- Bikakis, N., Sellis, T.: Exploration and visualization in the web of big linked data: a survey of the state of the art. Preprint. [arXiv:1601.08059](https://arxiv.org/pdf/1601.08059) (2016). <https://arxiv.org/pdf/1601.08059.pdf>
- Bizer, C., Heath, T., Berners-Lee, T.: Linked data: the story so far. *Int. J. Seman. Web Inf. Syst.* **5**(3), 1–22 (2009). <https://doi.org/10.4018/jswis.2009081901>
- Bizer, C., Heath, T., Berners-Lee, T.: Linked data—the story so far. *Int. J. Seman. Web Inf. Syst.* **5**(3), 1–22 (2009). <https://doi.org/10.4018/jswis.2009081901>
- Bolina, M.: Yewno discover. *Nordic J. Inf. Lit. Higher Educ.* (2019). <https://doi.org/10.15845/noril.v11i1.2772>
- Ceriani, M., Bernasconi, E., Mecella, M.: A streamlined pipeline to enable the semantic exploration of a bookstore. In: *Digital Libraries: The Era of Big Data and Data Science*. Springer, Cham, pp. 75–81 (2020). https://doi.org/10.1007/978-3-030-39905-4_8

15. Dadzie, A.-S., Rowe, M.: 2011 Approaches to visualising linked data: a survey. *Seman. Web* **2**(2), 89–124 (2011). <https://doi.org/10.3233/SW-2011-0037>
16. Ehlringer, L., Wöß, W.: Towards a definition of knowledge graphs. In: SEMANTiCS (2016). <http://ceur-ws.org/Vol-1695/paper4.pdf>
17. Gagnon, M., Zouaq, A., Aranha, F., Ensan, F., Jean Louis, L.: An analysis of the semantic annotation task on the linked data cloud. *Int. J. Metadata Seman. Ontol.* **13**(4), 317–29 (2019). <https://doi.org/10.1504/ijmso.2019.102678>
18. Haase, P., Herzig, D.M., Kozlov, A., Nikolov, A., Trame, J.: metaphactory: a platform for knowledge graph management. *Seman. Web* **10**(06), 1–17 (2019). <https://doi.org/10.3233/SW-190360>
19. Hogan, A.: Linked data. *Web Data* **2020**, 515–625 (2020). https://doi.org/10.1007/978-3-030-51580-5_8
20. Ikkala, E., Hyvönen, E., Rantala, H., Koho, M.: Sampo-UI: a full stack JavaScript framework for developing semantic portal user interfaces. *Seman. Web* **2021**(1–16), 2210–4968 (2021). <https://doi.org/10.3233/sw-210428>
21. Keim, A.D.: Information visualization and visual data mining. *IEEE Trans. Visual Comput. Graph.* **8**(1), 1–8 (2002). <https://doi.org/10.1109/2945.981847>
22. Kelly, D.: Methods for evaluating interactive information retrieval systems with users. *Found. Trends Inf. Retr.* **3**(1–2), 1–224 (2009). <https://doi.org/10.1561/1500000012>
23. Liu, J., Belkin, N.: Personalizing information retrieval for multi-session tasks: examining the roles of task stage, task type, and topic knowledge on the interpretation of dwell time as an indicator of document usefulness. *J. Assoc. Inf. Sci. Technol.* (2014). <https://doi.org/10.1002/asi.23160>
24. Marie, N., Gandon, F.L.: Survey of linked data based exploration systems. In: IESD@ISWC. http://ceur-ws.org/Vol-1279/iesd14_8.pdf (2014)
25. Michel, D.A.: What is used during cognitive processing in information retrieval and library searching? Eleven sources of search information. *J. Am. Soc. Inf. Sci.* **45**(7), 498–514 (1994). <https://doi.org/10.5555/186200.186212>
26. Miller, E.: An introduction to the resource description framework. *Bull. Am. Soc. Inf. Sci. Technol.* **25**(1), 15–19 (2005). <https://doi.org/10.1002/bult.105>
27. Motik, B., Patel-Schneider, P.F., Parsia, B., Bock, C., Fokoue, A., Haase, P., Hoekstra, R., Horrocks, I., Rutenberg, A., Sattler, U., Smith, M.: OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax (Second Edition) (2009). <http://www.w3.org/TR/2012/REC-owl2-syntax-20121211>
28. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. *Linguisticae Investigationes* **30**(1), 3–26 (2007). <https://doi.org/10.1075/li.30.1.03nad>
29. Nisheva-Pavlova, M., Alexandrov, A.: GLOBDEF: a framework for dynamic pipelines of semantic data enrichment tools. In: Proceedings of MTSR 2018. Springer, pp. 159–168 (2018). https://doi.org/10.1007/978-3-030-14401-2_15
30. O'Brien, H., McCay-Peet, L.: Asking “Good” questions: questionnaire design and analysis in interactive information retrieval research. In: Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval (2017). <https://doi.org/10.1145/3020165.3020167>
31. Osborne, J.: Best Practices in Quantitative Methods. SAGE Publications, Inc. (2008) <https://doi.org/10.4135/9781412995627>
32. Patrício, H.S., Cordeiro, M.I., Ramos, P.N.: From the web of bibliographic data to the web of bibliographic meaning: structuring, interlinking and validating ontologies on the semantic web. *Int. J. Metadata Seman. Ontol.* **14**(2), 124–124 (2020). <https://doi.org/10.1504/ijmso.2020.108318>
33. Pavlova, M.N., Alexandrov, A.: Extending the GLOBDEF framework with support for semantic enhancement of various data formats. *Int. J. Metadata Seman. Ontol.* **14**(2), 158–158 (2020). <https://doi.org/10.1504/ijmso.2020.10030301>
34. Po, L., Bikakis, N., Desimoni, F., Papastefanatos, G.: Linked data visualization: techniques. *Tools Big Data.* (2020). <https://doi.org/10.1007/978-3-031-79490-2>
35. Reinanda, R., Meij, E., de Rijke, M.: Knowledge Graphs: An Information Retrieval Perspective. Now Publishers (2020). <https://doi.org/10.1561/9781680837292>
36. Ristoski, P., Paulheim, H.: Semantic web in data mining and knowledge discovery: a comprehensive survey. *J. Web Seman.* **36**(2016), 1–22 (2016). <https://doi.org/10.1016/j.websem.2016.01.001>
37. Shen, W., Wang, J., Han, J.: Entity linking with a knowledge base: issues, techniques, and solutions. *IEEE Trans. Knowl. Data Eng.* **27**(2), 443–460 (2014). <https://doi.org/10.1109/TKDE.2014.2327028>
38. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings 1996 IEEE Symposium on Visual Languages, pp. 336–343. <https://doi.org/10.1109/VL.1996.545307>
39. Singhal, A.: Introducing the Knowledge Graph: things, not strings (2022). <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/>
40. Tsakonas, G., Papatheodorou, C.: Analysing and evaluating usefulness and usability in electronic information services. *Journal of information..* (2006). <https://doi.org/10.1177/0165551506065934>
41. Willer, M., Dunsire, G.: Bibliographic information organization in the semantic web. Chandos Publ. (2014). <https://doi.org/10.1533/9781780633978>
42. Yu, L.: SPARQL: Querying the Semantic Web. A Developer's Guide to the Semantic Web (2014), pp. 265–353. https://doi.org/10.1007/978-3-662-43796-4_6
43. Zloof, M.M.: Query-by-example: a data base language. *IBM Syst. J.* **16**(1977), 324–343 (1977). <https://doi.org/10.1147/sj.164.0324>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.