



Language Independent Searching Tools for Cultural Heritage on the QueryLab Platform

Maria Teresa Artese  and Isabella Gagliardi  

IMATI-CNR, Via Bassini 15, 20133 Milan, Italy
{artese, gagliardi}@mi.imati.cnr.it

Abstract. The paper describes the tools for searching and visualizing local and web inventories related to intangible and tangible cultural heritage in the QueryLab platform. The pandemic outbreak has made more evident the need to offer users tools to query and enjoy interesting and educative websites by their homes. The tools presented are useful for users who are not experts in the domain of the inventories, offering predefined queries and semantic query expansion to interact with the archives. The visual suggestions, in the form of word clouds of the tags of the selected archives, help in querying the archives and retrieving the elements that come closest to the user's interests. As one of the QueryLab aims is to continue to add inventories, in the languages they are stored, visual suggestions help to overcome the language distance between the archives and the users to allow an easy and successful interaction. This paper presents QueryLab tools for searching, browsing, and displaying multimedia data, with some preliminary results.

Keywords: Intangible and tangible inventories · Innovative tools for searching and browsing · Semantic query expansion · Word cloud

1 Introduction

The pandemic outbreak of this year has made even more evident the need to allow scholars, tourists, or simply curious of the web to enjoy museums or collections of tangible and intangible cultural heritage by their homes, using tools created especially for them, whether they are inexperienced in the use of the web, subject matter or language.

Following the publication of the UNESCO's 2003 Convention for the Safeguarding of Intangible Heritage and the rules for the inclusion of endangered items in the UNESCO's list, numerous archives have been created [10]. Examples of online intangible cultural heritage archives, created after the Convention, are those in Scotland, France, Spain, concerning Europe, while South Korea, Japan, and China have defined strategies to safeguard their traditions much earlier than the UNESCO Convention. On the other hand, museums and cultural heritage collections, collectively called glams, have always had the awareness to offer their users, at different degrees of specialization, a view of their heritage.

The basic idea of QueryLab (<https://arm.mi.imati.cnr.it/QueryLab>) is to create a platform to integrate cultural heritage archives, whether local or remote, in a transparent

way for users who are not aware of where the data physically resides. This paper presents QueryLab tools for searching, browsing, and displaying multimedia data, which are some of the aims of the platform.

The paper is organized as follows: after a brief section on the related works, an overview of the QueryLab platform is proposed, and the multimodal search engine is described, highlighting its characteristics. Finally, it follows conclusions, preliminary assessments, and future developments.

In this paper, all issues related to archives of (intangible) cultural heritage, their integration, tools, and models for search, navigation, and enjoyment with serious games are seen from a technological point of view, leaving to scholars and experts in the field the purely cultural part.

2 Related Works

A number of collections of cultural heritage objects are on the web with the purpose of making the contents of museums available to the users. According to the works in [8, 9], several features are appreciated by virtual visitors in their utilization of digital collections. The most valuable feature for the engagement of users is the availability of Search/Browse tools for interacting with the web. Regarding applications designed to query and browse the museums archives, innovative tools can be exploited to manage different types of data [1, 5].

While there are several sites that are the entry point for museum or tangible cultural heritage contents, the most famous of which is Europeana [6], to our best knowledge, QueryLab is the first that also deals with intangible assets. In [2, 3] the architecture of the system, with some technical details related to RESTful web services and an overview of its fruition, is described.

3 QueryLab Platform

QueryLab has been designed to be able to handle databases and inventories both local and integrated via RESTful web services (from now called web inventories). This paper will describe in detail how to search and navigate the QueryLab prototype, highlighting the differences that are encountered in the indexing, search and use of data from local databases, compared to those from databases queried through web services.

Figure 1 shows the logical schema of the QueryLab platform (this paper dives into the greenish area). The interaction/query with the remote inventories, performed using the web services provided by each of them, makes the query phase transparent to the different database locations and the addition of new inventories easy and seamless at any time. The data are queried via web services, “at their home”, without any caching system or local copies that require constant updating to be always aligned to the remote inventories.

To speed up the query phase on the different local databases, an ‘ICH light metadata structure’ [2, 3] has been defined, starting from:

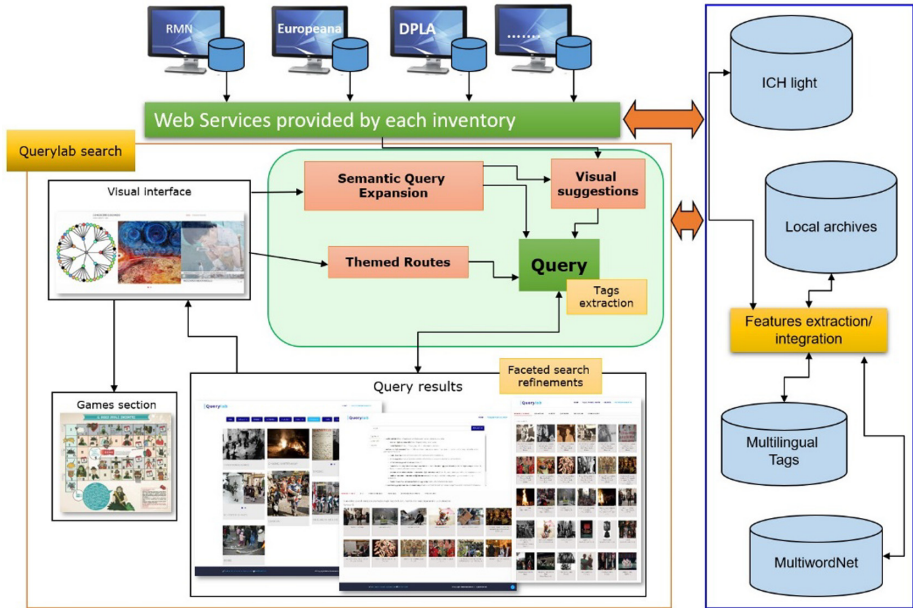


Fig. 1. Logical schema of QueryLab.

- the study and evaluation of the standard (de facto) metadata structures already in use, for example, EDM, Europeana Data Model [7],
- the structure provided by UNESCO to store information, which includes general information on cultural heritage, features, people that know and can transmit the knowledge, sustainability, data related to the inventory and references [11],
- the analysis of ICH inventories available on the web that share some common metadata, as title, UNESCO categories to which items belongs to, dates, places, ...

The QueryLab platform takes into account different ways to search, browse, visualize and interact with the data coming from different sources, so as to make the user able to interact comfortably and successfully with the web site even if he is not an expert in the field, or is not familiar with the content or language in which the terms are expressed.

Different ways to interact with the databases have been designed, according to the different types of users expected, depending on their information needs and their knowledge of the topics: Experts, Communities, Tourists and Web Users.

Table 1 describes the different inventories that participate in QueryLab. It can be noticed that web inventories, in general, contain both Intangible and Tangible items, related to traditions, interviews, photos, texts, manuscripts, etc. The archives store data in different/multi languages.

Table 1. The inventories in QueryLab so far, with some characteristics

Name (source)	Country	Data	Language	Description
Intangible search (local)	Italy	ICH	Italian, English, French, German	Arts and Entertainment, Oral Traditions, Rituals, Naturalistic knowledge, Technical knowledge
ACCU Databank (local)	Pacific Area	ICH	English	Performing Arts
Sahapedia (local)	India	ICH	English (used here) and other local idioms	Knowledge Traditions, Visual and Material Arts, Performing Arts, Literature and Languages, Practices and Rituals, Histories...
Lombardy digital archive (local)	Italy	ICH, CH	Italian	Oral history, Literature, Performing Arts, Popular traditions, Work tools, Religion, Musical instruments, Family...
Museum contemporary photography (local)	Italy	CH	Italian	Architecture, Cities, Emigration, Family, Objects, Work, Landscapes, Sport...
Europeana [6] (web)	European countries	ICH, CH	Multilingual	1914–18 World War, Archaeology, Art, Fashion, Industrial Heritage, Manuscripts, Maps, Migration, Music, Natural History, Photography...

(continued)

Table 1. (continued)

Name (source)	Country	Data	Language	Description
Victoria & Albert museum (web)	UK	ICH, CH	English	Architecture, Embroidery, Costume, Paintings, Photography, Frames, Wallpaper, Jewelry, Illustration, Fashion, Manuscripts, Wedding dress, Books, Opera, Performance...
Cooper hewitt (web)	USA	CH	English	Decorative Arts, Designs, Drawings, Posters, Models, Objects, Fashion, Textiles, Wallcoverings, Prints...
Réunion des Musée Nat – RMN (web)	France	ICH, CH	French with tags in English	Photography, Paintings, Sculptures, Literature, Dances, Modern Art, Decorative Arts, Music, Drawings, Architecture...
Auckland museum (web)	New Zealand	ICH, CH	English	Natural Sciences, Human History, Archaeology, War history, Arts and Design, Manuscripts, Maps, Books, Newspapers...

(continued)

Table 1. (continued)

Name (source)	Country	Data	Language	Description
Digital public library of America – DPLA (web)	USA	ICH, CH	English	American civil war, Aviation, Baseball, Civil rights movement, Immigration, Photography, Food, Women in science...

4 Querying the Archives

QueryLab offers multimodal means of navigation and search, e.g. guided tours, keyword analysis, and serious games. In this paper, only search and browse modes will be discussed in more detail: Themed Routes, Semantic Query Expansion, and Visual Suggestions, presented in increasing order of complexity and automation.

Whatever the mode is chosen for the data search, the same query is propagated to all databases, regardless of whether they are local, or web queried. The only difference is that by querying the databases locally, it is possible to have more control over the searches made than with web-services, whose structure and the queried fields are unknown.

4.1 Themed Routes

To allow users to have a “taste” of the contents of the different inventories, QueryLab offers “predefined queries” for database searches. Starting from the semantic tags composed by 1-gram or n-grams, defined or approved by ethnographers or experts, hierarchically organized in a multilevel-level structure – WordNet style –, the basic idea is to use these tags, available in the languages of the databases both local and web queried, to allow all users to easily interact with QueryLab. Users can browse among predefined paths, exploring and retrieving semantically similar documents.

The structure defined allows easy insertion of new structured tags at any time, is seamlessly adaptable to new inventories, highlighting themes and subjects of interest to users, or topics of relevance. These tags, originally created and defined for local ICH inventories, are used with success on all the QueryLab data.

4.2 Semantic Query Expansion

Scholars or expert users may be interested in querying the archives with specific terms or keywords. Besides the simple query using terms typed in by the users, which may or may not provide some results, QueryLab offers tools to expand semantically the queries, to enlarge or shrink results, by suggesting more general or more specific terms, according to WordNet and MultiWordNet. WordNet, a large lexical database in English, where nouns, verbs, adjectives and adverbs are grouped into groups of cognitive synonyms

(synsets), each of which expresses a distinct concept. MultiWordNet is a multilingual version of WordNet containing translation in different languages, as Italian, Spanish, Portuguese, etc.

By the integration of WordNet/MultiWordNet in QueryLab the semantics of terms is added [4], making it possible to:

- Seamlessly translate a term in any language of MultiWordNet;
- Structure flat list of tags into tree-shaped glossary;
- Enlarge or refine a query using the possible tree structures (associated with the different meanings of the selected term) of WordNet.

These semantic structure plug-ins can be used only on tags of the local databases. For web inventories, the visual suggestions are proposed.

4.3 Visual Suggestions

The web inventories are queried according to the RESTful protocol adopted by each inventory, generically on the descriptive data of the items. Until now, tags cannot be queried nor presented to the user in a list to be clicked. To overcome this limit and to offer all users suggestions of the possible queries, related to the one performed, the most relevant tags associated with the items are retrieved and displayed as word clouds.

These visual suggestions are part of a multi-step process to query the databases: the first step is performing a query using a simple term query, a semantic query expansion or selecting a thematic route.

The QueryLab system performs the query on all 10 databases. If the results do not satisfy the users, because a small number of items are retrieved, or none of them is significant, the visual suggestions may come to help. For local databases, the creation of visual suggestions is simple and immediate: the same tags that can be used as a possible refinement provide the material to be used. For remote databases, the data to be used are obtained via web services. Lacking a standard, each inventory required ad hoc analysis and procedures for tag extraction. By extracting the tags of the databases, the list of the tags, ordered by occurrence is then created and visualized as a word-cloud. By clicking on a tag, a new query is performed on the databases, and the process is repeated.

Figure 2 shows a word-cloud respectively for Digital Public Library of America (DPLA, USA) and Réunion des Musée Nationaux (RMN, France), after the query ‘wedding’ (‘mariage’ in French). It is important to note that RMN is a French-language database, so queries need to be translated before its use, because queries require terms in the language of the inventories.

In the case of RMN, visual suggestions are even more important, as the tags in the French language are extracted and the word cloud more useful for non-French users.

Visual suggestion, with its simplicity and its ability to extract tags in the language of the archive (and not necessarily in English), offers an extra tool to enhance the user’s ability to choose and retrieve those objects that are of interest to him, even if he does not know the language of the archive perfectly.



Fig. 2. Visual suggestions for Digital Public Library of America (DPLA, USA) (left) and Réunion des Musée Nationaux (RMN, France) (right) for ‘wedding’ (‘mariage’ in French) query.

5 Preliminary Results and Conclusions

The paper describes a work in progress for the development of a platform able to search and visualize two different types of inventories, the local ones and the ones queried through web services.

The tools presented are useful for users not-expert in the domain of the inventories, offering predefined queries and semantic query expansion to interact with the archives. The visual suggestions, in the form of word clouds of the tags of the selected archives, help in identifying the tags, sorted according to the number of occurrences, and therefore extracting that elements that come closest to the user’s interests. Users are provided with word clouds, a simple but expressive way to represent contents, as hints of the semantic contents of the databases and as suggestions to perform new queries.

As one of QueryLab aims is to continue to add inventories, in the languages they are stored, visual suggestions help to overcome the language distance between the archives and the users to allow an easy and successful interaction.

The work is still in progress, preliminary tests are giving positive results, but some issues have already been encountered:

- All the local databases are related only to ICH, while web inventories are mainly related to CH: no ICH web inventory has been found;
- Some web inventories are huge, with some millions of objects: a query refinement step is therefore needed to allow users to evaluate and enjoy the results. Visual suggestions could be used as a facet to refine queries and results;
- When the web inventories results are large, both tags extraction and word cloud visualization suffer: new solutions are therefore required;
- The databases are constantly growing in different languages, so new tests should be done to evaluate the results.

References

1. Artese, M.T., Ciocca, G., Gagliardi, I.: Evaluating perceptual visual attributes in social and cultural heritage web sites. *J. Cult. Herit.* **26**, 91–100 (2017). <https://doi.org/10.1016/j.culher.2017.02.009>
2. Artese, M.T., Gagliardi, I.: Sharing ICH Archives: Integration of Online Inventories and Definition of Common Metadata, vol. 8 (2019)
3. Artese, M.T., Gagliardi, I.: A platform for safeguarding cultural memory: the QueryLab prototype. *Art.* **18** (2019)
4. Artese, M.T., Gagliardi, I.: Multilingual specialist glossaries in a framework for intangible cultural heritage. *Int. J. Herit. Digit. Era* **3**(4), 657–668 (2014)
5. Ciocca, G., Colombo, A., Schettini, R., Artese, M.T., Gagliardi, I.: Intangible heritage management and multimodal navigation. In: *Handbook of Research on Technologies and Cultural Heritage: Applications and Environments*, pp. 85–118. IGIGlobal (2011)
6. Europeana Collections. <https://www.europeana.eu>, Accessed 13 Oct 2020
7. Europeana Data Model. <https://pro.europeana.eu/page/edm-documentation>, Accessed 12 Oct 2020
8. Lopatovska, I., Bierlein, I., Lember, H., Meyer, E.: Exploring requirements for online art collections. *Proc. Am. Soc. Inf. Sci. Technol.* **50**, 1–4 (2013). <https://doi.org/10.1002/meet.14505001109>
9. Lopatovska, I.: Museum website features, aesthetics, and visitors' impressions: a case study of four museums. *Museum Manag. Curatorship.* **30**, 191–207 (2015). <https://doi.org/10.1080/09647775.2015.1042511>
10. Sousa, F.: Map of e-inventories of intangible cultural heritage. *Memoriamedia Rev.* **1**(1) (2017)
11. UNESCO - Identifying and Inventoring Intangible Cultural Heritage. <https://www.unesco.org/culture/ich/doc/src/01856-EN.pdf>, Accessed 13 Oct 2020