LinkedQR: Improving Tourism Experience through Linked Data and QR Codes

Mikel Emaldi, Jon Lázaro, Xabier Laiseca, and Diego López de Ipiña

Deusto Institute of Technology - DeustoTech, University of Deusto
Avda. Universidades 24, 48007, Bilbao, Spain
{m.emaldi, jlazaro, xabier.laiseca, dipina}@deusto.es

Abstract. Since Sir Tim Berners-Lee announced the best practices to publish semantic data through the Web, Linked Data principles have been introduced into a wide variety of application domains. This usage of Linked Data eases the enrichment of offered data in a grade that the semantic data that not applies the Linked Data principles never would achieve. In this paper, we present LinkedQR[1], a tool to improve the collaboration between QR codes and Linked Data; and a case study based on the tourism sector located into an art gallery.

1 Introduction

Since Sir Tim Berners-Lee announced the best practices to publish semantic data through the Web [1], Linked Data principles have been introduced into a wide variety of application domains, e.g. Market Information Systems (MIS) [3], crowdsourcing [13], text processing [12] or documentation management systems [5]. Berners-Lee described these practices as: 1) use URIs as names for things; 2) use HTTP URIs so that people can look up those names; 3) when someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL); 4) include links to other URIs so that they can discover more things. On the other hand, QR codes have demonstrated their usefulness in fields such as logistics, indoor-localization systems [6] and so on. These successful experiences with QR codes suggest that they are a good partner to collaborate with Linked Data technologies. On the other hand, the "tourism experience" term has many interpretations from social, environmental, and activities components of the overall experience, inside some tourism activity. More information about "tourism experience" concept can be found at [15].

In this paper, we present LinkedQR, a tool to improve the collaboration between QR codes and Linked Data, through mobile and Web technologies. LinkedQR allows the data curators of an art gallery or of any kind of tourism installation, to retrieve enriched data from a vague information fragment. This information retrieval process helps them to provide more detailed information about an artwork or about an object of the environment. This information is

[1] This research is founded by project IE10-290 of Etortek 2010 program, Basque Government.

stored into a RDF store, following the Linked Data principles and providing them through a SPARQL endpoint. This data is consumed by a mobile application that uses the QR codes to identify each of the artworks inside the gallery through their unique HTTP URI. Besides, this paper is illustrated with a case study based on the tourism experience located into an art gallery, concretely "Sala Kubo" art gallery from San Sebastián[2].

The remainder of the paper is organized as follows. Section 2 discusses related work. Section 3 presents our solution to improve the tourism experience inside an art gallery. Section 4 refers about the architecture used on this project. Section 5 talks about the performance obtained from the experimentation of the developed system. Finally, Section 6 concludes and outlines the future work.

2 Related Work

2.1 Linked Data and QR Codes

As far as we know, the work related to the collaboration between Linked Data and QR Codes is limited. We can talk about More! [10], a mobile application that enables the exploration of the information about a speaker of an academic event. More! asks to the speaker to fill an application form with all his/him relevant data with the aim of showing them to the attendees. These data is stored into a server, and they can be retrieved through a HTTP URI that has been assigned previously to the speaker data. This URI is stored into a QR code, and it can be retrieved using More! mobile application.

Despite the fact that More! uses unique HTTP URIs, it does not offer the data through standard languages as RDF or SPARQL and it does not link the data with another semantic datasets, so that, we can say that the Linked Data offered by More! only fulfils two of the four principles announced by Berners-Lee. Furthermore, the speaker has to type his data manually into More!, making this process very tiresome.

2.2 Cultural and Artistic Mobile Applications

If we look at the Google Play market[3], we can discover many applications about museums and art galleries. Most of them, like the applications of Orsay's Museum[4] or Museum of Copies of Bilbao[5], only shows the schedule of the expositions and the information of the exposed artworks. The MoMA[6] art gallery adds the audio-guides to its application. In all of these applications, the visitor has to navigate through the different menus to find the concrete art piece he/she is viewing. LinkedQR allows to retrieve this information from a QR code, in

[2] http://www.kutxasocial.net/begira/canarte.nsf/fwhomeKU?openForm
[3] https://play.google.com
[4] http://www.musee-orsay.fr/
[5] http://www.museoreproduccionesbilbao.org/
[6] http://www.moma.org/

an easier way. Directly related to LinkedQR, we can talk about MoMu (Mobile Museum) [8]. The aim of MoMu was to offer additional information about the pieces of art we observe, in a more innovative way than traditional audio-guide systems. Usually, the visitor has to rent these traditional audio-guide systems, and she/he has to type a number to retrieve more information about an artwork. MoMu transforms our particular mobile device into a guiding device. It uses TRIP codes [7] to identify the artwork and retrieves enriched information from a central server. LinkedQR improves the aim of MoMu because the employees of the gallery retrieve this enriched information in a semiautomatic way, with the help of Linked Data techniques.

3 Proposed Solution

Throughout this section the proposed solution to ease the information management inside an art gallery is explained, describing the system architecture and showing how it works. In general terms, our solution improves the traditional guide-systems in three ways: 1) the data curator of the museum obtains invaluable help at data retrieving and merging stage, 2) the visitor has not to rent the audio-guide hardware and 3) the enriched data can be published and re-used by other museums or art galleries.

3.1 System Architecture

As shown at Figure 1, the system has two scenarios: 1) the information enrichment scenario and 2) the information consuming scenario.

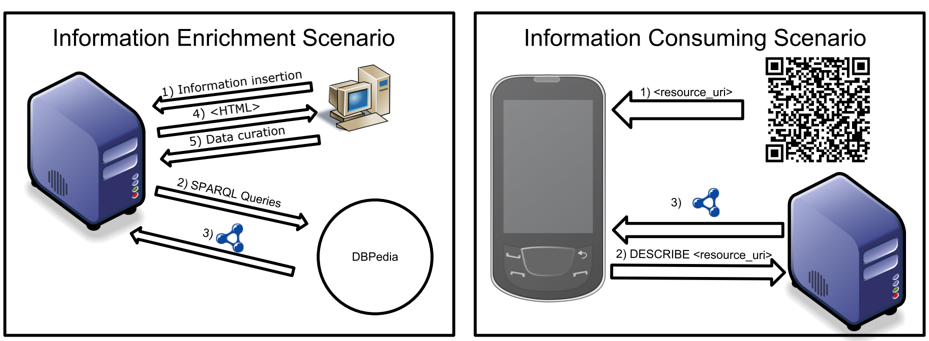

Fig. 1. LinkedQR system architecture.

Information Enrichment Scenario: This scenario takes as first input a set of basic information about artworks and authors exposed in the museum, stored within a traditional relational database. This input is adapted to Semantic Web using D2R, a tool that eases the transformation between relational and RDF based information [2]. Once all input information is described as RDF -using vocabularies like Music Ontology [11] for audioguides, or Dublin Core [16], DBPedia Ontology[7] and Yago [14] for other basic information-, the LinkedQR DBpedia extractor script is executed. This script is a semantic scrapper that retrieves the input RDF data, and makes SPARQL queries based on it over DBpedia SPARQL Endpoint[8]; these queries search for related information about every author, given her/his name as input, and taking into account possible Yago and DBpedia ontologies' classes and subclasses to which an artist could belong. If the author is not found, the script tries a brute-force alternative, building and exploring DBpedia URIs including directly author's name. Once the set of valid DBpedia URIs are found, the script analyzes the retrieved RDF descriptions, covering redirection and desambiguation issues, and searches for every literal in them and in any RDF file linked by them -with a maximum linking exploration depth of 1, for reasons of performance-, saving every result in the RDF store.

As the last step in this scenario, every information about artworks and authors is shown to the employee of the art gallery using a web based interface -entirely based on SPARQL queries and able to directly modify RDF files-. From this interface she/he can view, modify or translate any RDF description, as well as accept or reject any triple retrieved from DBpedia. When any external triple is accepted, LinkedQR enriches the original RDF description with these data, and generates a link between this RDF description and DBpedia one, following Linked Data principles.

Information Consuming Scenario: Once the employee has completed the information enrichment process, the QR codes are ready to be attached to the art pieces. At this scenario, the visitor of the museum has to download the application from Google Play Store or from a QR code attached at the entrance of the gallery, pointing to the URL of the application. When the visitor has the application installed into her/his smartphone, she/he only has to choose the "QR" option from the main menu, as can be seen at Figure 2. Immediately, the QR code reader software is launched to retrieve the URI of the piece of art. This URI is queried against the RDF server and the description of the resource is given. It is important to emphasize that this description is retrieved in the default language of the mobile device, allowing multilingual Linked Data. This semantic data is processed by Sesame, an architecture for storing and processing RDF data [4]. Sesame transforms the data stream given from the HTTP request done, to semantic triples. Finally, all the information about the piece of art, its author and the audio-guide are presented on the user's smartphone, as can be seen at Figure 2.

[7] http://dbpedia.org/ontology/
[8] http://dbpedia.org/sparql

Fig. 2. Application main screen (left) and information presentation screen (right).

4 Experimentation

At this section, the results of the testing of the system are presented. Concretely, this experimentation have been focused on the performance of the system at the following aspects: 1) information retrieval from the local repository by mobile application and its representation; 2) information found about the authors of exposed art pieces. To test this performance, the delay of graphs with different number of triples is presented. We have taken ten different samples, calculating the arithmetic mean, to avoid any distortion of the results. Figure 3 shows the performance of the mobile application[9]. As can be seen at this graph, the time employed by the smartphone to visualize the information grows as the number of triples to be processed grow, just as as we expected; but the time employed to retrieve the information from the server is nearly always 1,500 ms, despite of the increase of the triples. These results show that both server and network delays are not affected by the number of triples retrieved, but the smartphone suffers the consequences of the increase of the triples retrieved. This means that we have to take a lot of effort on optimizing the visualization of the information inside the smartphone. Despite of these delays, the total delay (≈2,000 ms) is acceptable for the user, following the criteria explained at [9], that limits the maximum delay time for an user interaction near 10 seconds.

At the server side of the system, related to the Linked Data retrieval stage, we have obtained these results: from 87 authors of the exposition, the system has recovered information about 57 of them (67.51%), generating 1756 new triples.

[9] The code of the mobile application can be found at https://github.com/memaldi/LinkedQR-android

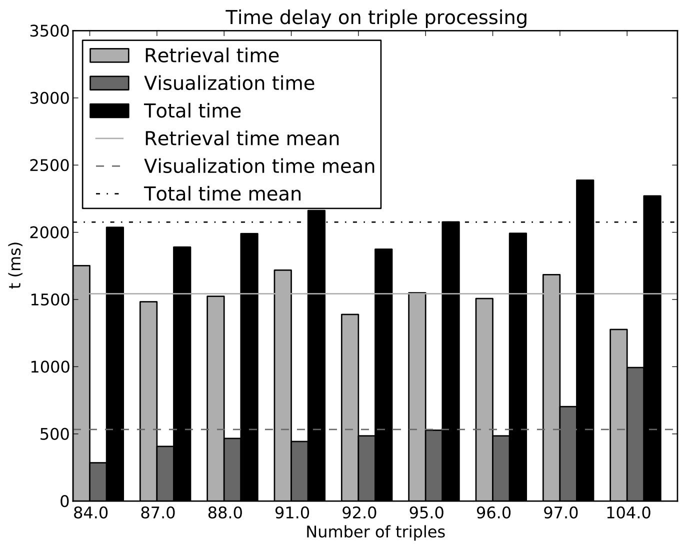

Fig. 3. Mobile application performance graph (Samsung Nexus S).

This data retrieval stage has been done with 2 depth levels: at the first level 885 triples (50.39%) have been retrieved and at the second level, 871 triples (49.61%). The true/false positive/negative rates show that the system has confused 6 persons that they are not authors of the exposition; and one author that she/he really is part of the exposition but the system rejects her/him. We can not talk in terms of retrieved triples, because we can not actually know the amount of total triples related to the artist inside DBPedia. In terms of found authors, the precision (Eq. 1), recall (Eq. 2) and true negative rate (Eq. 3) of the system can be calculated as follows:

$$Precision = \frac{T_{positives}}{T_{positives} + F_{positives}} = \frac{51}{51+6} = 89.47\% \quad (1)$$

$$Recall = \frac{T_{positives}}{T_{positives} + F_{negatives}} = \frac{51}{51+1} = 98.07\% \quad (2)$$

$$True\_negative\_rate = \frac{T_{negatives}}{T_{negatives} + F_{positives}} = \frac{29}{29+1} = 96.66\% \quad (3)$$

With these ratings, the obtained accuracy can be seen at Eq. 4. To measure the accuracy of this test, we have calculate the $F_1 score$ (Eq. 5).

$$Accuracy = \frac{T_{positives} + T_{negatives}}{T_{positives} + T_{negatives} + F_{positives} + F_{negatives}} = \frac{51 + 29}{51 + 29 + 6 + 1} = 91.95\% \quad (4)$$

$$F = 2 * \frac{Precision * Recall}{Precision + Recall} = 2 * \frac{0.8947 * 0.9807}{0.8947 + 0.9807} = 93.57\% \quad (5)$$

Authors				**Triples**			
	Found	*Not found*	*Total*	*Depth 1*	*Depth 2*	*Total*	
	57 (67.51%)	30 (32.59%)	87 (100%)	885 (50.39%)	871 (49.61%)	1756 (100%)	
Authors				**Triples**			
True positives	*False positives*	*True negatives*	*False negatives*	*True positives*	*False positives*	*True negatives*	*False negatives*
51	6	29	1	1612	144	-	-

Table 1. Results of Linked Data retrieval.

As can be seen, the system obtains a good performance when it has to classify the found results. The main issue is the percentage of found authors (67.51%). One of the main reasons of this percentage is the fact that many authors of the exposition are not well known. To solve this issue, at Section 5 we propose to include more specific semantic data sources related to art than DBPedia.

5 Conclusions and Future Work

At this paper, we have proposed a synergy between QR codes and Linked Data to increase visitor's tourism experience inside an art gallery and to ease the work of data curator of this gallery. As can be seen in Section 3, the proposed solution is easily deployable and no special hardware is needed. As can be seen at Section 4, the performance of the system is acceptable, but we consider that the percentage of authors found can be improved, adding more semantic sources about art and cultural heritage into the information enrichment process, like Amsterdam Museum[10] or GoogleArt wrapper[11]. However, the delay suffered at the mobile phone is acceptable in all samples taken into consideration.

We have identified some steps that have to be done in future. First, when the current exposition finishes at "Sala Kubo", the opinion of visitors and workers of the gallery is going to be taken into consideration, with the aim of the improvement of the user interface in both applications (administration panel and

[10] http://thedatahub.org/dataset/amsterdam-museum-as-edm-lod
[11] http://thedatahub.org/dataset/googleart-wrapper

mobile application). Related to this, we are considering the option of including the visitors opinions into the RDF graphs, describing them with SIOC ontology[12]. On the other hand, a wider job related to the performance of SPARQL queries used inside the system is going to be done.

As final reflection, we want to emphasize the significance of the QR codes to identify everyday objects and the enrichment the information about these objects with Linked Data.

References

1. Berners-Lee, T.: Linked data - design issues (2006), http://www.w3.org/DesignIssues/LinkedData.html
2. Bizer, C., Cyganiak, R.: D2r server-publishing relational databases on the semantic web. In: 5th International Semantic Web Conference. p. 26 (2006)
3. de Boer, V., Gyan, N.B., Bon, A., de Leenheer, P., van Aart, C., Akkermans, H.: Voice-based access to linked market data in the sahel. DownScale2012 p. 16 (2012)
4. Broekstra, J., Kampman, A., Van Harmelen, F.: Sesame: A generic architecture for storing and querying rdf and rdf schema. ISWC2002 pp. 54–68 (2002)
5. Emaldi, M., Buján, D., López-de-Ipiña, D.: Towards the integration of a research group website into the web of data. CAEPIA'11 (2011)
6. López-de-Ipiña, D., Klein, B., Guggenmos, C., Pérez, J., Gil, G.: User-Aware semantic location models for service provision. UCAMI2011 (2011)
7. López-de-Ipiña, D., Mendonça, P.R., Hopper, A., Hopper, A.: TRIP: a low-cost vision-based location system for ubiquitous computing. Personal and Ubiquitous Computing 6(3), 206–219 (2002)
8. López-de-Ipiña, D., Vázquez, I., Sainz, D.: Interacting with our environment through sentient mobile phones. pp. 19–28 (2005)
9. Nielsen, J.: Usability engineering. Morgan Kaufmann (1994)
10. Parra, G., Klerkx, J., Duval, E.: More!: Mobile interaction with linked data. vol. 6, p. 37 (2011)
11. Raimond, Y., Abdallah, S., Sandler, M., Giasson, F.: The music ontology. In: Proceedings of ISMIR2007. pp. 417–422. Citeseer (2007)
12. Rusu, D., Fortuna, B., Mladenic, D.: Automatically annotating text with linked open data (2011)
13. Simperl, E., Norton, B., Vrandecic, D.: Crowdsourcing tasks in linked data management. COLD2011 (2011)
14. Suchanek, F., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: Proceedings of WWW2007. pp. 697–706. ACM (2007)
15. Tussyadiah, I.P., Fesenmaier, D.R.: Interpreting tourist experiences from first-person stories: A foundation for mobile guides. In: Proceedings of The 15th European Conference on Information Systems Switzerland (2007)
16. Weibel, S., Kunze, J., Lagoze, C., Wolf, M.: Dublin core metadata for resource discovery. Internet Engineering Task Force RFC 2413, 222 (1998)

[12] http://sioc-project.org/