Semi-Automatic Generation of Quizzes and Learning Artifacts from Linked Data

Guillermo Álvaro Rey
playence, Spain
guillermo.alvaro@playence.com

Irene Celino
CEFRIEL, Italy
irene.celino@cefriel.it

Panos Alexopoulos
iSOCO, Spain
p.alexopoulos@gmail.com

Danica Damljanovic
University of Sheffield, UK
danica.damljanovic@gmail.com

Mariana Damova
Ontotext, Bulgaria
mariana.damova@ontotext.com

Ning Li
Open University, UK
n.li@open.ac.uk

Vladan Devedzic
University of Belgrade, Serbia
devedzic@gmail.com

ABSTRACT

In this position paper, we illustrate how Linked Data can be effectively used in a Technology-enhanced Learning scenario. Specifically, we aim at using structured data to semi-automatically generate artifacts to support learning delivery and assessment: natural language facts, Q&A systems and quizzes, also used with a gaming flavour, can be creatively generated to help teachers and learners to support and improve the learning path. Moreover, those artifacts can in turn be published on the Web as Linked Data, thus directly contributing to make the Web a global data space also for learning purposes.

Keywords

Linked Data, Quizzes, Games with a Purpose

1. INTRODUCTION

The Linked Data publishing paradigm is evolving the World Wide Web into a global data space [8] in which not only documents, but also raw data is exposed and linked through open standards. In the Web of Data, data from very different domains, covering topics as disparate as music, government and biochemistry, are available and semantically interconnected. Publishing data in this manner enables data sharing and reuse on a massive scale which presents great opportunities for developing new added-value services and applications on top of those interconnected datasets.

Technology-enhanced learning (TEL) is definitely an area that can strongly benefit from this trend. Semantic technologies providing structured and interlinked content like in the Linked Open Data (LOD) cloud [5] are a powerful resource that can be exploited naturally in online courseware, as well as being part of enhancements to the main learning resources. Linked Data mash-ups are more and more frequently used to generate charming visualizations (like maps, charts, statistics) to effectively explain and show information to learners in an easier-to-remember way. We argue that it is possible to go beyond existing approaches by exploring new interaction scenarios that build on the original data available on the Web, applying a combination of different knowledge-intensive and semantic techniques and creating new learning and assessment materials.

In particular, we foresee a set of tools that, on top of the available LOD datasets, are able to incrementally and semi-automatically construct different kinds of artifacts: natural language Facts, sets of Questions and Answers and Quizzes, all of which enhance the learning experience of students and professional learners, making it possible to establish innovative learning dynamics. Furthermore, our approach aims to involve the learners in different roles, leveraging their interactions for the detection of inconsistencies on the underlying datasets.

2. RELATED WORK

An important reference and a sort of first realization of the approach suggested in this paper is the Linked Data Movie Quiz (LDMQ [1]), an application that uses the Linked Movie DataBase (LMDB [7]) to generate questions about movies, directors and actors. Another example of generation of educational assessment items from Linked Open Data, in this case from DBpedia, is given by [6]. Those tools illustrate the potential of consuming Linked Data in order to create quizzes on top of it. Still, many are the limitations of such Web applications, which basically consist of queries to a single endpoint from the client side. Our approach foresees different domains by applying the same kind of mechanisms to different datasets, combining them in order to generate more advanced questions. Moreover, we foresee the persistence of both questions and answers, exposing them as Linked Data by making use of a specifically designed ontology for quizzes.

The idea of leveraging the interaction of users with the Web of Data has been recently explored in research trends that include the use of Games With A Purpose (GWAP [12]) approaches to create and curate semantic content [11]. The purpose can be the evaluation of some heuristics that are used to determine a ranking of facts within a knowledge base such as DBpedia or FactForge [4]. Examples of these approaches applied to the quiz area are WhoKnows [13] and

BetterRelations [9] games. By playing, users are improving the quality of the respective Linked Data sources. Other approaches of combining human effort with Linked Data algorithmic techniques have also been explored [10] by using crowdsourcing platforms such as Amazon's Mechanical Turk.

3. GENERATING LEARNING ARTIFACTS FROM LINKED DATA

As depicted in Figure 1, the proposed approach consists mainly of three steps. The main input for the *Knowledge Generation* step consists of structured data available on the Web, under the form of Linked Open Data. Those data are used to generate natural language facts, which are the first artifact that final users (teachers and learners) can creatively use in their learning practice. The *Question Shaping* step includes the definition and specification of question templates, which can then be filled with the data, knowledge and facts generated at the previous step. At this step, final users can use those templates to generate their own questions and knowledge. In the *Question Creation* step, we foresee tools that range from Question-Answering systems, to semi-automatic Quiz generation that can be used to compose Questionnaires (for testing or assessment purpose) and in interactive learning games, to engage learners in innovative and creative knowledge acquisition processes.
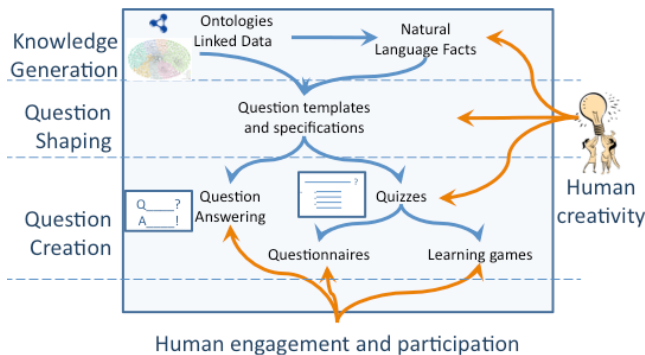

Figure 1: High-level view of our approach

3.1 Linked Data-driven Generation

Three different types of results can be incrementally obtained over the structured data coming from the Linked Open Data cloud. If we take as a basic example a particular piece of data, a statement from DBpedia such as:

`dbp:Napoleon dbpr:placeOfDeath dbp:Saint_Helena.`[1]

it is possible to semi-automatically generate the following kinds of result:

NATURAL LANGUAGE FACTS, i.e. sentences that express in natural language the information contained in the original formal statement(s). In our example, a fact would be something like "*Napoleon died in Saint Helena*".

NATURAL LANGUAGE Q&AS, i.e. a front-end system to answer Natural Language questions with Natural Language answers. For example, upon a question "*Where did Napoleon die?*", the system would be able to respond that "*Napoleon died in Saint Helena*". The answer can be generated using the fact generation from the previous step.

---

[1]RDF triple containing the relation "place of death" of resource `http://dbpedia.org/resource/Napoleon`.

QUIZZES, i.e. questions with one correct answer and multiple possible incorrect ones, generated from Linked Data. In our example, a generated question could be: "*Q: Where did Napoleon die? - a) Ajaccio b) Paris c) Saint Helena d) Waterloo*". An important aspect when generating quizzes is distinguishing which part of the Linked Data is suitable for generating questions of interest to a user, and also, generating a challenging response set with challenging distractors (wrong answers which are close to the correct ones). Therefore, the questions and responses generated can be filtered and ranked based on how appropriate or interesting they are for a given domain or business case. For example, if we start with the node `dbp:Napoleon`, the first question based on the sample DBpedia statement above could be "*Who died in Saint Helena?*". If there were too many answers to this question, then this question could be disregarded and a more complex and specific one could be generated based on other properties of the resource, e.g. "*Which emperor died in Saint Helena?*".

It is worth noting that these results constitute new learning knowledge that could be in turn semantically structured and exposed as Linked Data as a brand new dataset (e.g., instances of the concept "quiz" containing the questions with their right and wrong answers) properly linked to the original data sources, demonstrating once more the advantages of data inter-linkage.

3.2 Human-driven Generation

We give a primary role to the people involved in the learning process, both teachers and learners. We do believe in fact that the automatic tasks described above can be effectively improved and assisted by human check and control; in the meantime, involving people in creative activities better motivate and engage them in the loop.

We foresee a virtuous cycle where users discover problems within the original datasets, e.g. by identifying inconsistencies in the generated facts and quizzes. By crowdsourcing the data curation process, an effective and economical means for detecting errors is put in place, thus enabling the correction of the original datasets and benefiting the whole Linked Data environment. Learners could be involved in quiz generation as a non-linear learning process: it is often the case that, by asking questions and clarifications to teachers, learners better memorize and assimilate knowledge. We follow this idea and foresee a set of tools to collect learners' questions in a structured way, so to create further quizzes and tests for the same learning path.

4. BENEFITS OF OUR APPROACH FOR LEARNING

Quizzes are a powerful and time-tested way to support and assess learning, but generating high-quality quizzes is a knowledge-intensive, time-consuming and therefore expensive part of course production. Applying our approach to the wealth of Linked Data, the quiz generation process can be greatly facilitated to achieve not only enhanced quality of quizzes, but also reduced cost of course production. If the generated data are made accessible through standard formats and protocols for the Semantic Web (RDF, SPARQL and Linked Data), the cost of accessing and integrating it in 3rd party eLearning applications will be minimised.

Moreover, the set of dynamics foreseen by our approach would maximise the participation of students, gathering new knowledge through their interactions, while at the same time

providing them with an experience entertaining enough to keep them engaged with learning.

Besides pure TEL scenarios, the proposed approach would also be relevant in "edutainment" apps on different platforms (Web, phone, TV) in domains such as cinema or music. Existing websites could be enhanced by providing targeted questions which relate to the content of the page, e.g., quizzes for each artist at a music website. The same applies to phone applications that could show questions relevant to the geographical context of the user. In the case of social television, the use of these quizzes will permit the TV channel engage their audience with extra functionality.

5. POSITIVE EFFECTS ON THE LINKED DATA WORLD

Since the formulation of the Linked Data principles by Tim Berners-Lee in 2006 [3], the number of datasets available in the LOD cloud has been steadily growing, and it contains now more than 30 billion triples [5]. Due to its remarkable size and constant growth, it can be argued that Linked Data publication is finally going mainstream. However, for this publication to be effective and useful, there are two very important aspects of the approach that currently need to be considered, which we cover below.

5.1 Adoption and reuse

The consumption side of the paradigm has not been fully exploited, and its economic potential remains unexplored. Generic applications that consume Linked Data have not gone beyond browsers, search engines and standard visualizations, while domain-specific applications have been developed in a completely ad hoc manner by mashing up a few datasets. In any case, querying Linked Data is not possible without in depth knowledge of its structure and the names and semantics of all the properties and classes involved; thus it remains challenging even for the technically literate users.

We argue that the vast amount of available structured factual data can be leveraged through the aforementioned facts, Q&A and quizzes, thus applying a more accessible layer on top of the data. In this way, adoption and reuse of Linked Data can be fostered by providing user-friendly and more generic tools for accessing that knowledge.

5.2 Data Quality

The lack of generic Linked Data applications that fully tap into the potential of the vast available knowledge is, in part, a consequence of insufficient data quality. Linked Data typically suffer from mediocre quality, especially as a large amount of it is generated automatically from the Web-based text (e.g., DBpedia is generated based on Wikipedia). Typical problems consist in data being outdated, inaccurate, incomplete, inconsistent or even vague. Even though these problems are partially addressed [2], much more needs to be done in this respect. Hence, data curation is of utmost importance in many scenarios, while being a topic partly disregarded in the process of publishing Linked Data.

Our approach can contribute to refine the original datasets, thus helping to clean and improve the quality of the data exposed online. This is done (a) by detecting problematic data in the original datasets while creating the quizzes, and (b) by involving the end users in the loop through their participation in the quizzes and the feedback they give on them, either explicitly or implicitly.

6. CONCLUSIONS

In this position paper, we presented our approach to the use of Linked Data to create learning artifacts for a quiz-based assessment of learning paths. We illustrated the artifacts that can be semi-automatically generated and the technical and human means to reach this objective. The Linked Data quizzes could be in turn published on the Web of Data, thus increasing the potential value of the LOD cloud.

7. REFERENCES

[1] G. Álvaro and J. Álvaro. A Linked Data Movie Quiz: the answers are out there, and so are the questions. [blog post]. Retrieved from `http://bit.ly/linkedmovies`, Aug. 2010.

[2] S. Auer and J. Lehmann. Making the Web a Data Washing Machine - Creating Knowledge out of Interlinked Data. *Semantic Web Journal*, 2010.

[3] T. Berners-Lee. Linked Data. *International Journal on Semantic Web and Information Systems*, 4(2), 2006.

[4] B. Bishop, A. Kiryakov, D. Ognyanov, I. Peikov, Z. Tashev, and R. Velkov. FactForge: A fast track to the Web of data. *Semantic Web*, 2(2):157–166, 2011.

[5] C. Bizer, A. Jentzsch, and R. Cyganiak. State of the LOD Cloud, March 2011.

[6] M. Foulonneau. Generating Educational Assessment Items from Linked Open Data: The Case of DBpedia. In *1st Workshop on Linked Learning*, 2011.

[7] O. Hassanzadeh and M. Consens. Linked Movie Data Base. In *Workshop on Linked Data on the Web (LDOW)*, 2009.

[8] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, San Rafael, Calif., 2011.

[9] J. Hees, T. Roth-Berghofer, R. Biedert, B. Adrian, and A. Dengel. BetterRelations: Detailed Evaluation of a Game to Rate Linked Data Triples, October 2011.

[10] E. Simperl, B. Norton, and D. Vrandecic. Crowdsourcing Tasks within Linked Data Management. In *2nd International Workshop on Consuming Linked Data (COLD2011)*, 2011.

[11] K. Siorpaes and M. Hepp. Games with a Purpose for the Semantic Web. *IEEE Intelligent Systems*, 23:50–60, May 2008.

[12] L. von Ahn. Games with a Purpose. *Computer*, 39:92–94, June 2006.

[13] J. Waitelonis, N. Ludwig, M. Knuth, and H. Sack. WhoKnows? - Evaluating Linked Data Heuristics with a Quiz that Cleans Up DBpedia. *International Journal of Interactive Technology and Smart Education (ITSE), Special Issue on Multimedia technologies for e-learning*, 8(4):236–248, 2011.