Exploiting the Web of Data to bridge formal and informal learning experiences

Adolfo Ruiz-Calleja[†]
ETSI Telecomunicación,
Universidad de Valladolid
adolfo@gsic.uva.es

Miguel L. Bote-Lorenzo
ETSI Telecomunicación,
Universidad de Valladolid
migbot@tel.uva.es

Guillermo Vega-Gorgojo
ETSI Telecomunicación,
Universidad de Valladolid
guiveg@tel.uva.es

Sergio Serrano-Iglesias
ETSI Telecomunicación,
Universidad de Valladolid
sergio@gsic.uva.es

Juan I. Asensio-Pérez
ETSI Telecomunicación,
Universidad de Valladolid
juaase@tel.uva.es

Yannis Dimitriadis
ETSI Telecomunicación,
Universidad de Valladolid
yannis@tel.uva.es

Eduardo Gómez-Sánchez
ETSI Telecomunicación, Universidad de Valladolid
edugom@tel.uva.es

ABSTRACT

Smart Education promises personalized learning experiences that bridge formal and informal learning. Our proposal is to exploit the Web of Data to automatically create learning resources that can be, later on, recommended to a learner based on her learning interests and context. For example, a student enrolled in an arts course can get recommendations of learning resources (e.g., a quiz related to a monument she passes by) by exploiting existing geolocalized descriptions of historical buildings in the Web of Data. This paper describes a scenario to illustrate this idea and proposes a software architecture to support it. It also provides some examples of learning resources automatically created with a first prototype of a resource-generator module.

CCS CONCEPTS

• Information systems • Web applications

KEYWORDS

Web of Data, informal learning, smart learning

1 Introduction

Smart Education [1] promises highly personalized learning experiences that connect formal and informal learning. This opens new problems, as informal learning happens in students' daily life. If we want these learning experiences to be related to the topics covered by their formal education, the following question emerges: how can technology help students to detect and take advantage of informal learning experiences that could reinforce ongoing formal education activities?

One possible approach is to build a recommender system [2] to offer learning resources that suggest learning experiences to the students in their daily life. For example, the system may recommend a question to be answered (a resource), so that it indirectly invites the student to find out information related to this question (a learning experience). As such system is meant to be used in informal learning settings, its recommendations should take into account the learners' context [2,3] (e.g. their physical location or the tasks they are doing). However, how to feed such recommender system with contextualized learning resources is still a problem: many resources are needed as the learner's contexts are difficult to foresee and creating them manually would be a very time-demanding task.

Our key idea to overcome this issue is to automatically generate learning resources out of the Web of Data [4]. Other proposals follow a similar approach to create learning resources out of closed datasets [5]. We go beyond them as we exploit the Web of Data, out of which it is possible to create a myriad of meaningful resources from multiple domains. Our previous work [6] showed that it is possible to automatically create these resources, but their educational potential is still to be explored. Our hypothesis is that, as many entities in the Web of Data are related to physical locations or to a syllabus, it is possible to automatically infer contexts in which the educational resources will be meaningful.

This paper proposes a software architecture to automatically create learning resources out of the Web of Data that can be recommended in informal learning scenarios based on the leaner's context. The first prototype of this architecture has already been developed. The prototype has been employed to generate some examples of learning resources that would be meaningful in a

realistic scenario in order to show the potential of the proposed architecture.

The rest of this paper is organized as follows. Section 2 introduces the state of the art. Next, section 3 describes an illustrating scenario that exemplifies our goals and how we overcome current state of the art. Section 4 then introduces the proposed software architecture. Section 5 depicts the first prototype of the architecture and provides some examples of questions that were automatically generated with it. Finally, section 6 underlines the most important conclusions of the paper and describes the future work.

2 State of the Art

Smart Learning Environments (SLEs) are Technology Enhanced Learning (TEL) Environments that make adaptations and provide personalized support to the students considering their individual needs and context [7]. SLEs aim to bring education outside the classroom by exploiting the relationship between formal and informal learning and the personalization of the learning process [1]. With this purpose, SLEs aim to support learning processes adapted to each learner in a continuously evolving environment. Hence, systems that recommend learning tasks or resources become very relevant for this adaptation. These recommendations should be based both on the learner's characteristics and on her context [2,3].

Semantic annotation of learning resources is one of the approaches for context-aware adaptive learning systems [8]. These systems are based on a model of what is meant to be learned by defining a set of concepts and their relationships. Then, contents are annotated using these concepts and related to semantic descriptions of learning contexts in which that content may be relevant. Similarly, learners are modelled according to these same concepts and their contexts are tracked to detect when some resources should be recommended.

One of the main problems of these ontology-based context-aware adaptive systems is the creation and semantic annotation of learning resources. Some authors (e.g., [5]) propose to overcome this problem by automatically creating annotated resources out of a close database. However, the data creation problem persists, as the database should be manually created beforehand. Our approach is to exploit Open Data already published on the Web of Data [9] in order to tackle this issue. Thus, we can leverage the myriad of datasets openly available on the Web for educational purposes. In our previous work [6] we explored a technically-similar approach to create multiple-choice questions for a trivial game. We now face the problem of whether we can exploit third-party open datasets to automatically create learning resources and related them to context descriptions.

3 Illustrating Scenario

As an illustrating scenario, we consider a course of History of Art in a high school of Palencia (Spain). The teacher, Enrique, proposed his students to install an application in their smartphones and use it as a way to relate the content of the subject to their daily life experience. Then, Enrique accessed its web interface to create a new group and configure it. He stated that it is a group of secondary education students. Then he introduced a set of keywords that are related to the first chapter of the course: "Medieval art", "Romanesque architecture", "Gothic art", "Gothic architecture". This will be enough for the application to recommend learning resources to his students, thus helping them to make sense of learning experiences from their daily life.

Beatriz is one of the students in the course who installed the application. She joined the group Enrique created and configured it to get no more than ten recommendations a week. Two days later, she was walking around the city and she passed close to the church of Saint John, an important Romanesque temple. Beatriz got a notification inviting her to approach the temple. As she accepted it, the application showed a description of the church and asked her some multiple-choice questions (e.g. "in which century this temple was built?") that helped her to remember the facts related to it. As Beatriz was answering the questions, the application proposed others related to higher levels of the Bloom's taxonomy. For example, it offered other resources to further understand the characteristics of the temple (e.g. "could you describe the archivolt of Saint John church?") that made Beatriz look at some details of the temple; it also offered other resources to apply the knowledge obtained (e.g. "take a photo to a window that illustrates the characteristics of Romanesque style").

The following day Beatriz was at home and the application recommended her other resources related to medieval art and upper levels of the Bloom's taxonomy. For example, it showed a photograph of Santiago Cathedral (Romanesque style) and Leon Cathedral (Gothic style), asking her to enumerate the main differences. It also suggested Beatriz to look at the Rose window of Leon Cathedral and to make a drawing mimicking its style.

4 Technical Approach

In order to support the scenario described above, the technical system should suggest learning resources -automatically created out of data from the Web- based on the student's interests (defined by the teacher), context (e.g. her location) and her current knowledge (derived from her previous answers). Hence, from the technical perspective, it could be understood as a context-aware recommender system [2] that exploits the Web of Data. Figure 1 represents its software architecture (note that names in *italic* correspond to elements in Figure 1).

The components of the system architecture can be divided into two: those that create learning resources out of the Web of Data, and those related to the context-aware recommender system. Those related to the creation of learning resources are depicted on the left side of Figure 1. We can see that they follow the crawling pattern of Web of Data applications [9, chap. 6]. First, there is a *Web of Data crawler* that automatically retrieves data from different sources of the Web. This crawler merges the data from different sources and relates it to a common vocabulary in order to create an *Internal Dataset*. For example, this crawler may merge the data
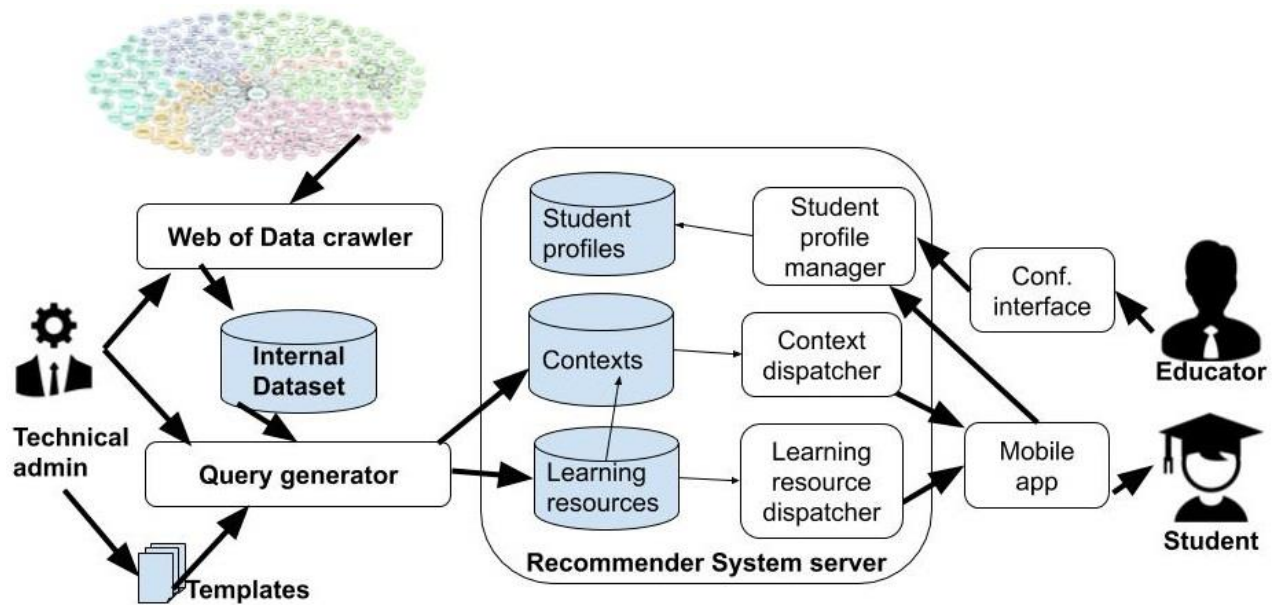
Figure 1: Software architecture of the context-aware recommender system.

from DBpedia, Wikidata, and datasets from the Open Data portal of Castilla y Leon (e.g. the dataset about monuments[1] and the one about municipalities[2]) in order to create an integrated dataset of the monuments of Castilla y Leon.

The data from this dataset is exploited by a *Query generator*. As in our previous work [6] and other proposals [5], the functionality of the crawler is configured by a set of templates -defined by a technical administrator- that include all the needed information about how the learning resources should be constructed out of the data retrieved. In this case, the *Templates* should also state how to relate learning resources to specific contexts. For example, the questions related to Saint John's temple should be related to the physical location of such temple, to a set of topics (e.g., Romanesque architecture), to the type of students the question is intended to, and to their relationship to the Bloom's taxonomy. Note that the physical location and a set of topics related are usually available on the Web of Data, while the administrator should include in the template the relationship to the student's profile and the Bloom's taxonomy.

The components related to the resource recommendations should manage student profiles, contexts and resources, as suggested in context-aware learning systems [3]. Its data layer includes a *Learning resource Dataset*, with the description of the learning resources; a *Context Dataset*, with the description of the contexts of the learning resources; and a *Student Profile Dataset*, which contains profiles of the students to make the resource recommendations. The question generator feeds with data to both the context and the learning resource datasets. Regarding the student profiles, they are built out of information by the educators

-who indicate the topics that are relevant using a configuration interface- and by the students -whose answers are tracked to relate each student to a level of the Bloom's taxonomy for each topic-.

Other elements in the *Recommender System Server* implement the business logic of the system. The *Mobile application* managed by the student will firstly submit to the *Context dispatcher* the context of the student (e.g. its location), and it will retrieve a set of similar contexts related to learning resources. When the student context matches one of the contexts retrieved, the *Mobile application* will request a learning resource to the *Learning resource dispatcher* by submitting the context and the user ID. Later on, the *Mobile application* will notify about the interactions between the student and the system to the *Student profile manager* for it to update its profile.

5   Prototype and Sample Queries

In order to test the technical feasibility of our approach, we developed an initial version of the *Web of Data crawler* and the *Query generator*. This initial version focuses on creating queries about medieval buildings of Castilla y Leon. Thus, these resources could potentially support the scenario described in section 3.

The *Web of Data crawler* is developed using Javascript. In its current version, it collects data from the Spanish[3] and English[4] versions of DBpedia and from Wikidata[5]. It automatically retrieved descriptions of 396 fortifications and 1,056 religious buildings. This data is later enriched with the aforementioned datasets about monuments and municipalities of the Open Data Portal of Castilla

---

[1] https://datosabiertos.jcyl.es/web/jcyl/set/es/cultura-ocio/monumentos/1284325843131

[2] https://datosabiertos.jcyl.es/web/jcyl/set/es/sector-publico/municipios/1284278782067

[3] http://es.dbpedia.org
[4] https://dbpedia.org
[5] http://www.wikidata.org/

y Leon. Thus, we obtained a coherent dataset that integrates data from four different sources.

Once this internal dataset was populated, the *Query generator* (also developed with Javascript) applied a set of templates to automatically create educational queries. As an example, we can take the following example of a *Template*:

"Category": "Fortification"
"Attributes": ["image", "municipality"],
"Question": "In which municipality is this fortification?",
"Answer": "resource:Municipality",
"Image": "resource:Image"
"Bloom": "Remember",
"Age": "12-17"

This *Template* is applied to all the elements in the category "Fortification" that have the "image" and "municipality" attributes. The *Template* also defines some parameters to be related to the query context, such as the bloom taxonomy level "remember" and the age group "12-17". Other information related to the resource, such as its physical location or its Romanesque origin, is also included in the description of the context. Thus, the *Query generator* automatically created a set of queries and a set of contexts, related to these queries, that will be later exploited by the recommender system.

When applied to our initial dataset, this template created 143 questions and their corresponding contexts. As an example, we can see the following question, which is meant to be shown at the same time as the image in Figure 2 (obtained from DBpedia) and whose answer is "Peñafiel".

In which municipality you can find this fortification?

Figure 2: Peñafiel castle.

The context related to this question includes four parameters:

"Age group" : "12-17"
"Bloom" : "Remember"

"Related Topics" : ["Romanesque architecture", "Gothic architecture", "Peñafiel", "Valladolid", "X Century"]

Note that the values of "Age group" and "Bloom" are defined in the *Template*, so these same values are given to all the questions obtained out of this template. On the other hand, the values of "Related topics" are obtained from the description of Peñafiel Castle in the Web of Data[6].

A second example of a template is the following one:

"Category": "Religious Buildings"
"Attributes": [["element" : "archivolt"], "label", "geolocation"],
"Question": "Could you describe the archivolt of {label}?",
"Bloom": "Analyze",
"Age": "15-17"
"Geolocation": "resource:geolocation"

In this case, the *Template* is applied to all the elements in the category "Religious Buildings" whose description include their label, their physical location and "archivolt" as one of their elements. The question is constructed using the attribute "label" of the resource. However, as it is an open question, no correct answer is defined.

As an example, we obtained the following question, which is related to the illustrating scenario described before:

Could you describe the archivolt of Saint John church?

The context associated to this resource is the following one:

"Age group" : "15-17"
 "Bloom" : "Analyze"
"Related Topics" : ["Romanesque architecture", "Palencia", "XI Century"]
"Geolocation" : "42.007637, -4.524489"

Again, we can see that the values of the first two parameters of this context are obtained from the template. The set of related topics and the geolocation are obtained from the description of Saint John Church in the Web of Data[7].

6   Conclusions and Future Work

This paper presented our approach to bridge formal and informal learning context in Smart Learning Environments (SLEs). The main idea is to exploit the Web of Data to automatically generate a big set of learning resources that can potentially be recommended to students based on their context. We described the design of a technical system that would support such approach and we presented a first implementation of a prototype that generates contextualized learning resources out of data from the Web.

These are very promising results as we show that it is not only possible to automatically generate learning resources, but also to relate them with descriptions of contexts where they may be relevant. Right now, we only generated a few contextualized resources related to medieval buildings. However, our previous

---

[6] See http://es.dbpedia.org/resource/Castillo_de_Peñafiel

[7] See http://es.dbpedia.org/page/Ermita_de_San_Juan_Bautista_(Palencia)

work shows that it is possible to automatically generate questions for many other domains (including history, forestry, literature, music, botany, zoology...). Thus, we trust on the potential of our approach to create a significant database of contextualized learning resources.

Our work in the near future will focus on building the system described and on assessing the learning potential of the resources generated. Later on, we plan to integrate this system in a SLE [10] The SLE will manage the student profiles, while our application will offer a set of learning resources and a context-aware recommender system. For example, a teacher may propose some activities in the SLE and the relevant topics for the students will be automatically extracted out of these activities. Later, the recommender system can also propose learning resources to be used in the SLE. Thus, these resources will not only be offered in a mobile application, but also across other spaces, such as web, virtual reality or augmented reality environments.

Even if we are in an initial stage of our research, we can foresee some aspects that will be especially challenging and may lead to relevant research contributions. First, current open datasets may not be as detailed as some applications may require. For example, out of the four datasets considered we can infer that the Cathedral of Leon is gothic, and that it has a rose window. However, it is not explicit that such rose window is especially relevant (so the questions related to the rose window are more relevant than those related to other architectural elements of the temple). We can consider that this problem may be solved in the future as the Open Data community is publishing more and more data in the Web. However, we can also offer an annotation application for teachers and students to collect this kind of information. If the data gathered is published as Linked Open Data, we will also contribute to the Web of Data.

Other problem related to the automatic creation of learning resources is who defines the templates. This definition is critical because it states what are the learning resources, to whom these resources are relevant and in which contexts they should be offered. In our current proposal we consider that a technical administrator would define these templates, as it is done in other similar proposals [5,6]. However, it would be desirable that teachers are the ones who defines the templates. Hence, we will explore the possibility of creating an application for teachers to define these templates.

Another important problem that we should face is how to integrate this system in a SLE. A critical aspect is how to define the interests of the students. Our initial approach is to offer teachers a configuration interface where they can explicitly state these interests. However, as mentioned before, when the system is integrated into a SLE we expect the interests to be automatically collected from the learning design.

ACKNOWLEDGMENTS

REFERENCES

[1] B. Gros "The design of smart educational environments". Smart Learning Environments 1(3):1-11. 2016.

[2] K. Verbert, N. Manouselis, X. Ochoa, M. Wolpers, H. Drachsler, I. Bosnic, and E. Duval. "Context-aware recommender systems for learning: A survey and future challenges". IEEE Transactions on Learning Technologies, 5:318–335, 10 2012.

[3] A. Schmidt. "Impact of context awareness on the architecture of learning support systems". In C. Pahl, editor, Architecture Solutions for e-Learning Systems, pages 306–319. Dublin, Ireland, 2008.

[4] C. Pereira, S. Matsui-Siqueira, B. Pereira-Nunes, and S. Dietze. Linked data in education: A survey and a synthesis of actual research and future challenges. IEEE Transactions on Learning Technologies , 11(3):400–412, 2018.

[5] T. Alsubait, B. Parsia, and U. Sattler. Ontology-based multiple choice question generation. KI-Kˉunstliche Intelligenz, 30(2):183–188, 2016.

[6] G. Vega-Gorgojo. "Clover Quiz: A trivia game powered by DBpedia". Semantic Web 10(4):779-793, 2019.

[7] J.M. Spector. "Conceptualizing the emerging field of smart learning environments". Smart Learning Environments 1(1):2-10, 2014

[8] L. Aroyo, P. Dolog, G. J. Houben, M. Kravcik, A. Naeve, M. Nilsson, and F. Wild. "Interoperability in personalized adaptive learning". Journal of Educational Technology & Society, 9(2), 4-18., 2006.

[9] T. Heath and C. Bizer. Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. 2011.

[10] S. Serrano-Iglesias, M.L. Bote-Lorenzo, E. Gómez-Sánchez, J.I. Asensio-Pérez, G. Vega-Gorgojo. "Towards the enactment of learning situations connecting formal and non-formal learning in SLEs". Proceedings of the International Conference on Smart Learning Environments (ICSLE), 187-190. Denton, USA. 2019.