A Linked-Data-based search system of educational tools for the Web of Data

Adolfo Ruiz-Calleja

School of Telecommunications Engineering, University of Valladolid
Camino del Cementerio s/n, 47011 Valladolid, Spain
adolfo@gsic.uva.es

Abstract. The number of learning situations that can be carried out in a VLE (Virtual Learning Environment) can be improved by the integration of third-party external tools. Before the integration takes place, it is compulsory to retrieve some information, both to select the most appropriate tool to support an specific situation and to be able to integrate it in the VLE. Current tool registries have some drawbacks that make the data retrieval difficult for the educators; among these drawbacks, the most important one is that search engines are not able to automatically import information from external datasets. To overcome this limitation, this paper introduces an ongoing doctorate research that proposes the creation of a search engine based on the Linked-Data principles.

Keywords: Linked Data, educational tools, Semantic Web, integration.

1 State of the art and problem statement

A Virtual Learning Environment (VLE) is a software system used by teachers and students to support the realization and assessment of learning situations [7]. Some examples of VLEs with a wide adoption are Moodle[1] or LAMS[2]. These VLEs typically include a set of general purpose tools, although educators commonly require to use a particular tool to support specific learning situations [5]. For this reason, ongoing research on the field tries to integrate third-party external tools in VLEs, following different approaches, as shown in [2].

A key step to enabling the integration of external tools in VLEs, or other software environments, is facilitating the discovery of tools, providing useful information for the tool selection and its integration in the environment. When educators are searching tools they need to make queries using educational abstractions. In addition, the integration of tools should be semi-automatic, as educators are not expected to manipulate the data needed to integrate the tool. Finally it is convenient to have a big boundle of tools available and thus allowing educators to choose the most appropriate one.

There are some general-purpose search engines, such as Google, that are commonly used by teachers to discover general information. However, they are

[1] http://moodle.org/
[2] http://www.lamsinternational.com/

designed to retrieve text documents and not data about *software* tools; these documents may contain a text description about software tools, but data must be extracted out of them by a human. In order to facilitate the software tool data retrieval, some other search engines, such as Google Gadgets[3] or Yahoo! Widgets[4], give support for searching tools, providing data about thousands of tools. Moreover, they are commonly used by non-expert users. However, these kind of search engines are not specialized in educational domain, so they do not provide educational information related to the tools; in addition they are keyword-based search systems, which are prone to obtain irrelevant results [8, Pag. 91]. These drawbacks are overcome by OntoolSearch[5], an educational tool search engine based on semantic technologies that use the Ontoolcole [10] ontology to describe *software* tools. Nevertheless, OntoolCole only allows to express the functional properties of the tools, and not their non-functional properties. Therefore, OntoolSearch cannot provide the information for integration of tools.

Another important drawback that is common to all the aforementioned search systems is that they are not able to automatically import information from third-party external repositories. These search systems can only take information from their own internal registry, which behaves as a independent data silo. Thus, each search system is only able to provide the information that has been explicitly described in its internal registry, even if there is more relevant information in another dataset freely accessible through the Web. Moreover, a software tool may be described in a search system internal registry but it may have been updated in another external data source, so a teacher that uses this search system will get out-of-date information.

2 Proposed approach

The most important problem that this doctorate research work is trying to solve is how to create an educational tool search engine able to automatically collect information from third-party external data repositories. Obviously, the problem of integrating external data does not only affect to the educational domain; it is a very important problem to solve related to the information retrievement and data management. The Linked Data [3] approach is a recent proposal that is expected to facilitate the automatic access to the information published in external repositories.

Linked Data is a methodology for publishing data in the Semantic Web [4]. Its key idea is to identify concepts (both data and meta-data) with URIs and reuse URIs defined by external data providers. Following this approach, when two repositories have information referred to the same entity, they will use the same URI to define it and thus a software agent could automatically retrieve the information related to a concept published in different data registries. This proposal has been widely adopted and many data providers are linking their

[3] http://desktop.google.com/plugins/
[4] http://widgets.yahoo.com/
[5] http://gsic.uva.es/ontoolsearch

datasets according to these principles, building the Web of Data, which is motivated by the Linking Open Data Project[6].

In the Web of Data there are several interlinked data repositories with information from multiple domains. Even if none of them provide specific information about educational tools, there are some data repositories where useful information of software tools is registered. One of these datasets is DBpedia[7], where up-to-date structured information extracted from Wikipedia is freely available. A key step to solve the proposed research question is to design a search system able to extract structured data from different sources from the Web of Data, such as Dbpedia, and automatically relate the data to the system internal vocabulary. Note that the data extraction from each dataset will not be ad-hoc implemented because, as different data sources are linked, a single software agent can extract information from several datasets.

However, there is no data registry in the Web of Data that describes software tools contemplating their educational capabilities, nor there is any that give details about how to integrate them in a VLE. So that, the support of searching tools using educational abstractions will be limited. In order to solve this problem a search engine is proposed; this search engine will have a collection of adaptors that will allow to automatically extract data from different sources. Moreover, in order to complete the description of software tools , teachers can publish their educational capabilities and technical users can give details of how to integrate them in a learning environment. This way, a new interlinked dataset will be created, containing information with a specific purpose and reusing information from multiple previously available data repositories.

3 Methodology

In order to design, develop and evaluate an educational tool search system, the doctorate research is being iteratively performed according to the methodology in engineering [1], following the next steps:

1. Research problem definition. The first stage consists on the definition of a relevant research problem after exploring the literature. The problem stated is how to retrieve useful information for the selection of learning tools and their integration in VLEs.
2. Solution proposal. The second stage tries to overcome the problem found with a solution proposal. A software that supports the information retrieval from third-party external data sources, as well as the publication of educational information following the Linked Data principles is the current proposal for the problem.
3. Solution design and development. This stage consists on the design of the solution proposed in the previous stage and also implies the development of a prototype. Therefore, a prototype of the architecture, including the search

[6] http://linkeddata.org
[7] http://dbpedia.org/About

engine and some adaptors will be developed. Moreover, the integration of the prototype in at least a VLE will be carried out.

4. Solution assessment. The last stage shows that the solution overcomes the problems that were detected in the first stage of this methodology. In this context, the system should be able to automatically retrieve and integrate information from different sources and to registry educational information about the tools following the Linked Data principles.

The whole picture of the research proposal is represented in Figure 1.

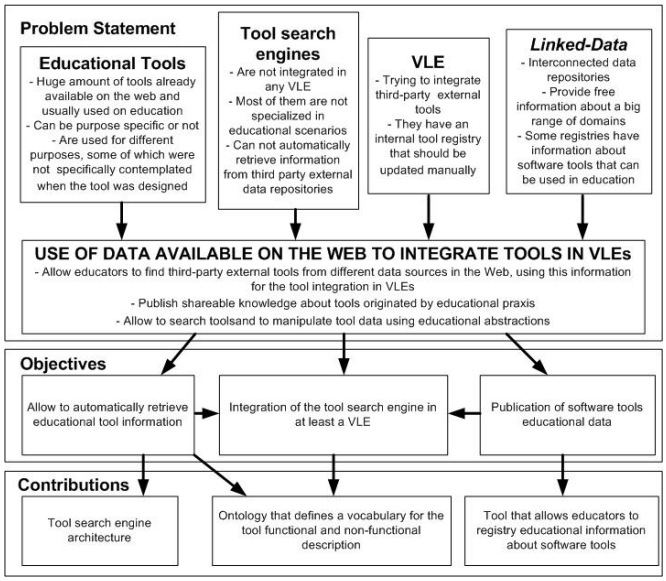

Fig. 1. Doctorate research schema.

4 Results already reached

Doctorate research has been working on some of the partial objectives described in Section 2. Specifically, previous work has focussed on the detection of the information requirements for the tool integration in VLEs and in the design of the search engine architecture.

In order to detect the information that should be provided by the system, some real examples were studied where a tool is published by a tool provider, found and selected by a teacher and later on integrated in a VLE. Furthermore, an analysis of the literature was carried out. Out of this review the tool information requirements were established: the description of the tool should contemplate both functional, technological and administrative parameters of the tool.

Once the information requirements were analyzed, an ontology was design in order to define a vocabulary that can be used by the search system to describe the learning tools. This ontology reuses other conceptualizations that can be found in the literature; for example, Ontoolcole [10] is used to define the educational concepts related to the tools while Dublin Core [6] provides the vocabulary of the administrative domain. Nonetheless, a conceptualization for describing the technical parameters has been specifically defined since there was not found any ontology that defines the needed concepts.

As far as the search engine design concerns, the proposed architecture, which is based on papers such as [9], is shown in Figure 2. The search engine has a central manager that collects the query made by the teacher through its interface and coordinates the data retrievement. The central manager sends the query using a common language, define by the abovementioned ontology, to several adaptors (two in the example of the Figure 2), which mediate in the data exchange between the manager and the external data sources. Finally, there is an educational data registry, which contains educational data about software tools and enriches the information about tools provided by external data repositories.

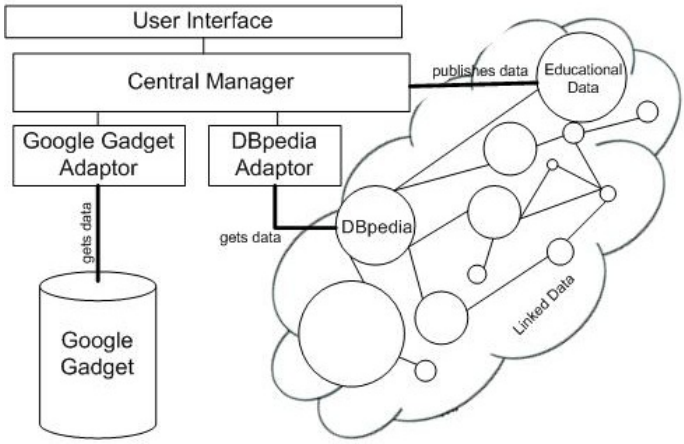

Fig. 2. Current version of the search engine architecture.

5 Conclusions and future work

The present paper shows an ongoing doctorate research work. A research question was detected, which consists on the design of an educational tool search engine able to automatically collect information from external data repositories. The proposed approach to solve this problem is based on the Linked Data prin-

ciples because it is a recent trend that allows software agents to automatically retrieve information from the Web of Data.

This approach overcomes the detected problems in current educational tool search systems, such as OntoolSearch: firstly, it is possible to find educational tools that were not specifically described in the system's data source; secondly, data maintenance will be facilitated, as the system is able to automatically import up-to-date information; finally, the publication of educational tool information will be easier, since it will be possible to reuse the data retrieved from external registries, so it will only be necessary to registry the tools educational aspects. In addition, all the information created by the system will be published and will be freely available on the Web, so it could be reused by other people for some other educational applications.

Future work will focus on the implementation of the search system described in Section 4, taking into account the proposed architecture and the designed ontology. In addition, the integration of the search system in at least a VLE will be an interesting task to be carried out.

References

1. W.R. Adrion. Research methodology in software engineering: summary of the Dagstuhl workshop on future directions on software engineering. *SIGSoft Software Engineering Notes*, 18(1):36–37, January 1993.
2. C. Alario Hoyos, J.I. Asensio Pérez, M.L. Bote Lorenzo, E. Gómez Sánchez, G. Vega Gorgojo, and A. Ruiz Calleja. Integration of external tools in virtual learning environments: main design issues and alternatives. In *Proceedings of 10th International Conference on Advanced Learning Technologies (ICALT)*, Sousse, Túnez, July 2010.
3. T. Berners-Lee. Linked Data - Design Issues, July 2007. URL = http://www.w3.org/DesignIssues/LinkedData.html, última visita junio de 2010.
4. T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):34–43, 2001.
5. Miguel L. Bote-Lorenzo, Eduardo Gómez-Sánchez, Guillermo Vega-Gorgojo, Yannis A. Dimitriadis, Juan I. Asensio-Pérez, and Iván M. Jorrín-Abellán. Gridcole: A tailorable grid service based system that supports scripted collaborative learning. *Computers & Education*, 51(1):155 – 172, 2008.
6. DCMI Usage Board. DCMI metadata terms. Specification, DCMI, January 2008. URL = http://dublincore.org/documents/dcmi-terms/, last visited July 2010.
7. P. Dillenbourg. Virtual learning environments. In *Learning in the New Millennium: Building New Education Strategies for Schools. EUN Conference Workshop on Virtual Learning Environments*, Rhodes, Greece, 2000.
8. D. Fensel. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag, Berlin, Germany, second edition, 2004.
9. W. Meng, C. Yu, and K.L. Liu. Building efficient and effective metasearch engines. *ACM Computing Surveys (CSUR)*, 34(1):48–89, 2002.
10. G. Vega-Gorgojo, M. L. Bote-Lorenzo, E. Gómez-Sánchez, J. I. Asensio-Pérez, Y. A. Dimitriadis, and I. M. Jorrín-Abellán. Ontoolcole: Supporting educators in the semantic search of CSCL tools. *Journal of Universal Computer Science (JUCS)*, 14(1):27–58, 2008.