Exploiting the Web of Data to provide descriptions of ICT tools: a preliminary report about SEEK-AT-WD*

Adolfo Ruiz-Calleja
University of Valladolid
Paseo Belén, 15
Valladolid, Spain
adolfo@gsic.uva.es

Thanassis Tiropanis
University of Southampton
Faculty of Physical and
Applied Sciences
Southampton, UK
tt2@ecs.soton.ac.uk

Guillermo Vega-Gorgojo
University of Valladolid
Paseo Belén, 15
Valladolid, Spain
guiveg@tel.uva.es

Juan I. Asensio-Pérez
University of Valladolid
Paseo Belén, 15
Valladolid, Spain
juaase@tel.uva.es

ABSTRACT

In order to support educators when selecting ICT tools, several educational organizations provide ICT tool registries whose functionality is limited by the data they contain. These registries could reduce the cost of creating and updating their datasets obtaining ICT tool descriptions from the Web of Data. However, some problems hinder the educational consumption of these descriptions: they are not described using an appropriate vocabulary for this specific domain and registries do not discriminate which tools are useful for education. SEEK-AT-WD is proposed as an infrastructure that overcomes these problems and consumes ICT tool descriptions from the Web of Data, publishing them back once they are related to educational concepts. This paper discusses the challenges found when developing SEEK-AT-WD, and how they were solved to take advantage of data from the Web in the educational applications.

Categories and Subject Descriptors

H.3.5 [**Information Systems**]: Information Storage and Retrieval Online Information Services

Keywords

Web of Data, educational ICT tools, Linked dataset

1. INTRODUCTION

The use of Information and Communication Technologies (ICTs) is widespread in education. One of their main advantages is that educators can choose among different available ICT tools so as to support a wide range of learning

---

*Linked Learning 2012, 17 April 2012, Lyon, France.

situations. However, when selecting an appropriate tool for a specific learning situation, educators need to be informed about the educational capabilities of ICT tools [8]. For this reason, many educational organizations offer registries of ICT tool descriptions that support educators when searching, discovering and selecting tools (i.e. CoolTools-ForSchools[1] or SchoolForge[2]).

When developing these registries, it is required to describe a collection of ICT tools in a way that facilitates their discovery by educators. Additionally, the organization should maintain updated the collection of descriptions, adding new elements or modifying already existing ones when necessary. These tasks are especially challenging: on the one hand, the organization should put some resources to create and update the educational descriptions; on the other hand, tools are convenient to be described according to an educational-specific vocabulary that fit educators' questions [8].

Linked Data [1] can well be seen as a promising approach to reduce the cost of creating and updating registries of educational ICT tools, since it allows to consume open data available on the Web that may be useful in this specific context. In fact, several cross-domain repositories of the Web of Data [5, sec. 3.2] (i.e. DBpedia[3]) provide updated descriptions of ICT tools that are typically employed in education. However, these repositories do not describe ICT tools using educational vocabularies, nor they discriminate which tools can be useful to support learning situations. These problems hinder the creation of a registry of educational tools that take advantage of the data published on the Web.

In this context, SEEK-AT-WD (Support for Educational External Knowledge About Tools in the Web of Data) is proposed as an infrastructure that aims to provide a specialized Linked Data registry of educational ICT tools. SEEK-AT-WD retrieves ICT tool descriptions from different sources of the Web of Data and publishes them aligned to an edu-

---

[1]http://cooltoolsforschools.wikispaces.com/
[2]http://www.schoolforge.net/
[3]http://dbpedia.org/

cational vocabulary. Thus, these descriptions are enriched with an educational value that is useful in the domain. In addition, SEEK-AT-WD expects to motivate the federation of ICT tool registries, envisioning a situation where educators enrich tool descriptions and share their experiences.

This paper presents SEEK-AT-WD as an ongoing research work. Specifically, this paper focuses on describing SEEK-AT-WD requirements and the challenges associated to take advantage of the data from the Web (on section 2). Then, its development is reported and it is discussed how these problems are overcome (section 3). Finally, conclusions and outlook are provided (section 4).

2. SEEK-AT-WD REQUIREMENTS

SEEK-AT-WD is understood as an infrastructure that provides educational descriptions of ICT tools. In this regard, SEEK-AT-WD defines a read-only API that allows third parties to consume its data.

SEEK-AT-WD is expected to contain a large number of ICT tool descriptions in order to allow educators to choose among different options the one that best suits their needs. These descriptions should be retrieved from different sources of the Web of Data. Concerning this, it should be possible to automatically discover data sources that provide useful information, as well as to obtain data from useful sources that may appear as the Web of Data grows [5, chap. 3]. Finally, SEEK-AT-WD should update the tool descriptions provided if they are updated in other sources of the Web of Data. In this regard, SEEK-AT-WD is expected to periodically retrieve new data from external sources in an automatic way.

Regarding the data acquisition, when searching for educational tools, educators make complex queries and expect a reasonable response time [7]. On the other hand, it is convenient to provide a vocabulary that define educational abstractions [8]. Thus, it will be easier for educators to discover and understand tool descriptions, as well as to select an appropriate one. In addition, commonly used vocabularies on the Web of Data [2], and more specifically on the educational domain, should be used in order to make easier the retrieval of data by third parties.

3. SEEK-AT-WD DESIGN AND IMPLEMENTATION

Taking into account the aforementioned requirements, SEEK-AT-WD has been designed and developed as an infrastructure that both consumes and publishes Linked Data.

3.1 SEEK-AT-WD data API

The data API provided by SEEK-AT-WD follows the Linked Data principles [1] and several best practices [5, chap. 5] were followed for its definition. This decision has two main advantages: firstly, data consumption is facilitated since a standardized format and publication methodology is followed; secondly, it enables the federation of ICT tool datasets, trying to reach an scenario where open data about ICT tools is shared between educational organizations.

SEEK-AT-WD is expected to publish several thousands of tool descriptions, which will be updated frequently. For

these reasons, an RDF store was considered to be the best option to store the data, as well as a Linked Data interface and a SPARQL endpoint to allow third parties to consume it. Concerning to the vocabulary used to describe ICT tools, it is recommended to reused already existing ontologies since it reduces the effort of creating the vocabulary and makes easier to share the data between third parties [3, chap. 1]. After analyzing the vocabularies used in the Web of Data [2] and in the learning domain, Ontoolcole [8] was the only ontology found specifically develop to describe educational ICT tools. Ontoolcole defines three taxonomies: one of educational tools, another one of educational task that can be supported by these tools, and finally another taxonomy of artifacts that can be managed by the tools. In addition, Ontoolcole formally defines the relationships of these taxonomies to allow the inference of some properties of ICT tools. However, this inference implies high complexity when describing tools, as well as a reasoner to make the inferences [7], which slows down the data retrieval.

Experience shows that most of the time, this high expressiveness is not necessary [7]. So that, for the sake of simplicity when publishing and consuming ICT tool descriptions, SEEK ontology was developed as a new vocabulary based on Ontoolcole. Specifically, it defines the same taxonomies of tools, tasks and artifacts, as Ontoolcole does, but there is no formal relationship between them. Furthermore, following Linked Data principles, each concept in SEEK ontology is explicitly related to its Ontoolcole synonym to allow interoperation. This way, describing a tool is much simpler while the educational tasks it supports and the artifacts it handles can still be stated. Finally, some other characteristics of the tools, such as their descriptions or some administrative data, are defined using Ontoolcole or other vocabularies commonly used on the Web of Data, such as Dublin Core[4], FOAF[5], DBpedia Ontology[6] or RDFS[7].

3.2 SEEK-AT-WD architecture and implementation

Once SEEK-AT-WD data API was defined, it was necessary to develop an infrastructure that retrieves educational tool descriptions from the Web, aligns them to SEEK vocabulary and stores them in SEEK-AT-WD data repository, allowing its consumption by third parties, as it was previously stated. The crawling pattern [4][5, chap. 6] was followed to design SEEK-AT-WD architecture, since it separates the tasks of creating a cache of data from the Web and using this data, which enables applications to query data from different sources with a high performance [4] and a reasonable response time. As a disadvantage, it is possible that SEEK-AT-WD may contain stale data, so the crawler should periodically crawl the Web of Data.

Figure 1 shows the architecture of SEEK-AT-WD. It can be seen that a `crawler` retrieves data from different sources of the Web of Data and stores them in SEEK-AT-WD data repository (`SEEK-KB`). Then, several `educational applications` can consume the data provided by SEEK-KB either

---

[4]http://dublincore.org/2010/10/11/dcelements.rdf

[5]http://xmlns.com/foaf/spec/

[6]http://wiki.dbpedia.org/Ontology

[7]http://www.w3.org/1999/02/22-rdf-syntax-ns
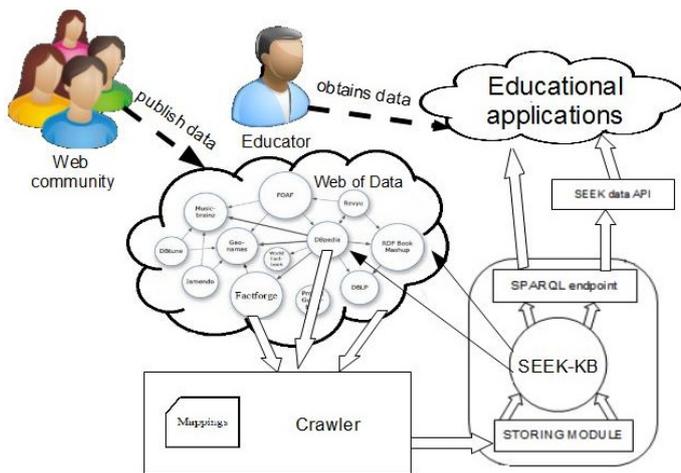
Figure 1: SEEK-AT-WD architecture and roles involved.

posing queries to its `SPARQL endpoint` or obtaining it from `SEEK Data API`. These applications are expected to be used by `educators`, and thus, they can consume ICT tool descriptions that are published by the `Web community`.

Note that in this infrastructure the crawler plays a central role. It should access several sources in the Web of Data, extract ICT tool descriptions and create descriptions compliant to SEEK vocabulary out of the data extracted [5, chap. 6]. The crawler creates the descriptions with the aid of some ontology `mappings` that formally define the relationships between the ontologies used in cross-domain datasets and SEEK vocabulary. Details about these mappings can be seen at [6]. Regarding on how the data is obtained, two important decisions should be noted. First, the crawler uses the well-known *follow-your-nose* principle [5, chap. 2]. However, in order to guarantee the quality of data it is configured to retrieve tool descriptions from two reliable sources (DBpedia and Factforge[8]). Then, the crawler only *follows its nose* one step, meaning that it only retrieves data from the documents that are directly linked by these two sources. Thus, automatic discovery of sources is possible without the need of implementing a complex heuristic to evaluate the trustworthiness of each data source. Second, a simple heuristic is implemented to resolve identities: it is considered that any two tools linked with `owl:sameAs` or with the same `rdf:label` property are the same tool. Thus, the crawler can infer that two descriptions refer to the same tool even if they are described by different sources.

Once the crawler was designed, it was implemented using the Jena Ontology API[9]. In January 2012 this crawler obtained 3556 descriptions of ICT tools. On the other hand, SEEK-KB was deployed using RKBExplorer[10], which is an infrastructure whose robustness and reliability had already been proved. SEEK-KB can be accessed at `http://seek.rkbexplorer.com/`, while an example of a tool description is found at `http://seek.rkbexplorer.com/id/tool/GNU_Paint`.

---

4. CONCLUSIONS AND FUTURE WORK

The Web of Data represents an opportunity to reduce the cost of creating and updating educational ICT tool registries. However, educational applications need to overcome several problems in order to take advantage of ICT tool descriptions published on the Web of Data. This paper briefly reported the design and development of SEEK-AT-WD, focusing on how this infrastructure is able to obtain educational ICT tool descriptions from third-parties. SEEK-AT-WD requires a vocabulary related to others commonly used in the Web of Data, while using educational concepts to describe tools. Then, a crawler was developed to obtain tool descriptions, align them to SEEK vocabulary and store them in SEEK-KB. The data contained in SEEK-KB is publicly available on the Web and can be accessed using a web browser. In order to improve its accessibility, authors are developing a search client that will retrieve data from SEEK-KB. Authors also envision to develop an educational data publishing tool. Thus, educators will be able to consume data from the Web of Data and to contribute to its grow.

Acknowledgment

5. REFERENCES

[1] T. Berners-Lee. Linked Data - Design Issues, 2006. http://www.w3.org/DesignIssues/LinkedData.html, last visited February 2012.

[2] C. Bizer, A. Jentzsch, and R. Cyganiak. State of the LOD cloud, 2011. http://www4.wiwiss.fu-berlin.de/lodcloud/state/, last visited February 2012.

[3] A. Gómez-Pérez, M. Fernández-López, and O. Corcho. *Ontological Engineering*. Springer, London, UK, 2004.

[4] O. Hartig and A. Langegger. A database perspective on consuming Linked Data on the Web. *Datenbank-Spektrum*, 10(2):1 – 10, 2010.

[5] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool, first edition, 2011.

[6] A. Ruiz-Calleja, G. Vega-Gorgojo, E. Gómez-Sánchez, C. Alario-Hoyos, J. I. Asensio-Pérez, and M. L. Bote-Lorenzo. Automatic retrieval of educational ICT tool descriptions from the Web of Data. In *Proceedings of the Twelfth International Conference on Advanced Learning Technologies*, 2012. In press.

[7] G. Vega-Gorgojo, M. L. Bote-Lorenzo, J. I. Asensio-Pérez, E. Gómez-Sánchez, Y. A. Dimitriadis, and I. M. Jorrín-Abellán. Semantic search of tools for collaborative learning with the Ontoolsearch system. *Computers & Education*, 54(4):835 – 848, May 2010.

[8] 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.