International Journal of Engineering Education Vol. 31, No. 3, pp. 893–900, 2015
Printed in Great Britain
0949-149X/91 $3.00+0.00

Interlinking Educational Data: An Experiment with Engineering-related Resources in GLOBE*

ENAYAT RAJABI, MIGUEL-ANGEL SICILIA and SALVADOR SANCHEZ-ALONSO
Information Engineering Research Unit, Computer Science Department, University of Alcalá, 28805, Alcalá de Henares, Spain.
E-mail: {enayat.rajabi, msicilia, salvador.sanchez}@uah.es

JUAN MANUEL DODERO
Computer Engineering Department, University of Cádiz, Spain. E-mail: juanma.dodero@uca.es

Linking different kinds of engineering learning resources on the Web of Data enables enrichment, ease of navigation, casual discovery and improves resource seeking. This is performed by many tools and approaches built to discover similarities between the entities on the Web. In this paper we present a report primarily focused on evaluating the interlinking of engineering-related resources of a significant educational repository (GLOBE) to one of the most important datasets (DBpedia) on the Linked Open Data (LOD) cloud. After considering various interlinking approaches for link discovery, the paper focuses on the use of one of the interlinking tools (LIMES) and outlines the number of resources linked to the DBpedia dataset. In this empirical study, we report that almost 40,000 engineering resources were matched to the DBpedia concepts. Our findings are also examined as well as classified in various categories by human experts.

Keywords: engineering resources; linked data; interlinking; educational data; GLOBE; LIMES

1. Introduction

For years, a significant decline in the number of students graduating in Engineering fields has been observed [1]. It has been suggested that students who study science and engineering concepts experience a higher workload because this knowledge has a richer, more complex structure [2]. Of the diverse attempts to understand the problem, Felder [3] argues that more effective teaching methods in introductory courses will result in a higher retention rate. One of these methods is known as Problem-Based Learning, which requires students to identify and research, based on a poorly structured problem proposal, a set of resources useful for acquiring the knowledge needed to solve the problem [4]. However, if students lack the generic skills needed to undertake self-directed study, then it is likely that the goals of the problem-based strategy will not be achieved [5].

Making engineering students aware of the learning approach necessary can increase their involvement in courses [6]. Pedagogical approaches have been applied to initially structure open-ended problem-based learning approaches and gradually move towards open-ended problems. Nevertheless, relevant discussions on the suitability of problem-based strategies for teaching engineering conclude that it has certain limitations, which make it less suitable as an overall strategy for engineering education [2]. Another pedagogy that is usually applied in individual courses or throughout a curriculum is the project-based learning strategy [7].

In project-based learning, students access learning content when required, but the teacher prepares much of it. In problem-based learning, students control the content and delivery while the lecturer usually determines the problem. Eventually, either the students or the teacher must identify and control the content that is relevant to successfully solve the problem or achieve the project goals. The appropriate structuring of a project or an open-ended problem requires finding a handful of relevant resources, which can be found in open learning repositories as long as they are available.

From the point of view of content, various kinds of e-learning resources have motivated data providers to publish their educational documents on the Web of Data [8]. Linking engineering learning resources isolated in different repositories to valuable datasets facilitates resource seeking on the Web and pushes forward the exploitation of the large amounts of open data available on the Web [9]. Furthermore, it enriches the source information by connecting them to various targets of knowledge [8]. In particular, discovery of learning resources about engineering can be facilitated when they are interlinked with public domain datasets, statistics sources, and governmental data. Linked Data (LD) [10], as a recent approach for interlinking data, allows digital resources to be shared, reused, and accessed by students. Using LD, repository owners can publish structured data and establish categorized links between their repositories and from other sources. Furthermore, the LD approach and tools provide some solutions for intelligent

* Accepted 24 October 2014.

linking, as well as for integration and consumption of experiment data [11]. Many educational institutions, universities and libraries have embraced LD principles and released educational resources as part of the LD cloud. DBpedia [12], one of the most used datasets [13] is an LD version of Wikipedia that makes it possible to link data items to general information on the Web. In particular, the advantages of linking of engineering content to DBpedia is to make public information usable for other datasets and to enrich datasets by linking to valuable resources on the Web of Data [14]. However, interlinking educational data is still largely unexplored. Dietze et al. [15] presented a general approach to exploit the wealth of existing technology-enhanced learning (TEL) data on the Web by allowing its exposure as LD.

In this paper, we evaluate existing approaches for interlinking objects on the Web of Data and select LIMES [16] for linking a large collection of data to the LOD cloud. As a result, we expose the GLOBE (Global Learning Objects Brokered Exchange) metadata as LD and discover the similarities between its metadata elements and DBpedia. Finally we evaluate the results and list the advantages of this interlinking process.

The rest of the paper is structured as follows. Section 2 describes briefly how educational data is nowadays exposed as LD and discusses different existing approaches for interlinking. In Section 3, we discuss the dataset used for examining the interlinking framework as our experimental setting. Section 4 provides the methods and results for our evaluation. Conclusions and outlook are provided in Section 5.

2. Background

In the last decade, the existing teaching and learning strategies in engineering education have been improved so that faculty members in academe are recommended to make enhanced design pedagogy their highest priority in future resource allocation decisions [17]. Engineering graduates also need to have a broader knowledge of fundamental engineering science and computer literacy [7]. Given that engineering students' demands are unlikely to be satisfied by a traditional engineering curriculum, they are expected to find their learning resources on the Web. On the other hand, the majority of e-learning materials, which are engineering-related, can be enriched when they are conjoined to useful information on the Web of Data. In the following sub-sections we will explain how several educational institutions have exploited their learning materials on the Web and what the current linking approaches for connecting various learning resources are. Finally, we will select our approach for interlinking.

2.1 Exposing educational resources as Linked Data

Several educational institutions e.g., the University of Muenster (DE) [18], the Open University (UK) [19], the National Research Council (CNR, Italy) [20], and the Southampton University (UK) [21] embraced the LD approach by exposing their learning resources as LD formats. Notably, we outline two educational datasets which have exploited their learning (meta)data in RDF format:

Organic.Edunet [22] is a learning portal that provides access to digital learning resources on Organic Agriculture and Agroecology—it facilitates access, usage and exploitation of such content. This collection, which currently contains the metadata of almost 11,000 resources, has exposed its content as LD [23] and published these resources as a dataset in the LOD cloud [24]. This dataset is also linked to other datasets such as DBpedia through its metadata elements.

Europeana [25], the European Union's flagship digital library project, enables search and discovery in more than 17 million items by collecting metadata from approximately 1,500 cultural data providers across Europe [25]. Europeana published a first subset of the Europeana dataset [26] after enriching existing metadata records via a SPARQL endpoint and data dump. It exposes data based on the Europeana Data Model (EDM), which is for publishing and linking Europeana metadata. It also links the data provider's metadata to other datasets such as DBpedia, Geonames [27] and GEMET [28].

In particular, one of the approaches for representing any kind of data as LD is mapping the collection to RDF triples [29] which has been applied for our interlinking purpose in the following steps:

(a) Storing the metadata in a repository that will be accessible via the web.
(b) Converting them to RDF using semantic web tools.
(c) Making the educational data accessible via a SPARQL endpoint or RDF dump

2.2 Approaches for interlinking

Several tools and approaches exist for interlinking data in the LOD datasets. Simperl et al. [30] provided a comparison of interlinking tools based on a set of criteria such as use cases, annotation, input and output. Similarly, we explain some of the related tools, but focusing on their need for human contribution (to what extent users have to contribute to interlinking), their automation (to what extent the tool needs human input), and area (in which environment the tool can be applied).

From a human contribution perspective, User Contributed Interlinking (UCI) [31] is an interlinking tool that creates different types of semantic links such as “owl:sameas” and “rdf:seeAlso” between two datasets relying on user contributions. In this Wiki-style approach, users can add, view or delete links between data items in a dataset by using a UCI interface. “Games With A Purpose” (GWAP) [32] is another software which provides incentives for users to interlink datasets using games and pictures. In this tool, the user distinguishes different pictures with the same name. “Linkage Query Writer” (LinQuer) [33] is also a software for semantic link discovery between different datasets, based upon a framework that consists of APIs that allow users to write their queries in an interface.

Semi-automatic interlinking [34], as another approach for interlinking, provides a type of analysis technique to assign multimedia data to users using multimedia metadata. “Interlinking Multimedia” (iM) [35] is also a pragmatic way for applying the LD to fragments of multimedia items and presents methods for enabling a widespread use of interlinking multimedia. RDF-IA [36] is another linking application that carries out matching, fusion and interlinking of RDF datasets according to the user configuration, and generates several outputs such as interlink files including “owl:sameAs” statements between the data items. Another semi-automatic approach for interlinking is the Silk Link Discovery Framework [37] which finds the similarities by specifying the types of RDF links. Some similarity metrics are combined based on the link conditions within different LD sources. LIMES is also a link discovery software in the LOD which implements a time-efficient approach for large datasets in metric spaces [16]. This approach presents a command-line tool and a graphic user interface for finding similarities between two datasets and automatically suggests the results based on the metrics. GoogleRefine [38] is software for cleaning, transforming, and interlinking any kind of data with a web user interface. It has also the benefit of reconciling data to the LOD datasets (e.g., Freebase or DBpedia) [39]. The following table briefly summarizes the described tools and mentions the area of application for each one.

2.3 Selected approach for interlinking large datasets

As our approach is to interlink the datasets via a SPARQL endpoint, Silk and LIMES were selected as our final candidates. Besides that, they were well-documented, updated frequently and used rich as well as diverse matching algorithms for interlinking [16, 37]. In both approaches, the user specifies the SPARQL endpoints of the datasets, comparable elements and thresholds of acceptance of output. Eventually, the tools report the results based upon the user configuration and similarities between two datasets. In a study, Ngonga et al. [16] examined both LIMES and Silk, from a time-efficiency perspective and showed that LIMES is more time-efficient than Silk for link discovery between two LOD datasets. They evaluated LIMES using synthetic as well as real data and it outperformed other approaches with respect to the number of comparisons and runtime. They also showed that the speed of this tool improves with the complexity of the mapping task and makes it especially suitable for handling large-scale matching tasks. Moreover, in a study Rajabi et al. [40] evaluated several interlinking tools on the Web of Data and identified LIMES as one of the most promising tools for linking datasets, and thus we selected the tool to carry out the interlinking.

3. Experimental setting

GLOBE is a large repository with almost 1.2 million learning resources [41]. Including various kinds of educational data encouraged us to assess the possibility of interlinking GLOBE to the LOD datasets. GLOBE is a federated repository that consists of several other repositories, such as OER Commons [42], which has manually created metadata as well as aggregated metadata from different sources. Current research on the use of GLOBE learning resource metadata [41] shows that 20 out of 50 of the metadata elements, which are based upon the

Table 1. Existing interlinking tools description

Tool	User contribution	Area
UCI	Reviewing the semantic links	General data source
GWAP	Matching of objects through playing a game	Web pages, e-commerce offerings, Flickr images, and YouTube
LinQuer	Writing LinQL queries	LOD datasets
IM	Matches multimedia by annotating and linking	Multimedia
RDF-IA	Configuring the input	LOD datasets
Silk	Configuring the input file, reviewing the result	LOD datasets
LIMES	Configuring the input file, reviewing the result	LOD datasets
GoogleRefine	Importing data, reviewing the result	General data ,LOD datasets

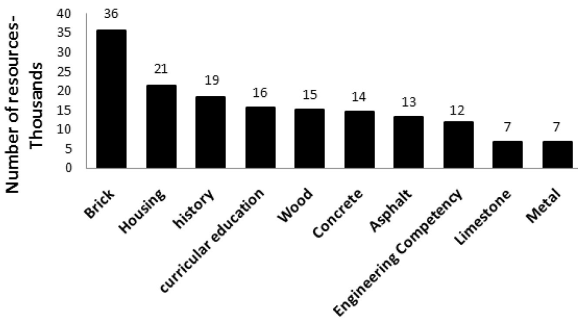

Fig. 1. Top ten taxonomies in the GLOBE metadata.

IEEE LOM schema [43], are used consistently in the repository and thus can be considered for interlinking. After analyzing the GLOBE metadata, we realized that several metadata elements such as "General.Identifier" or "Technical.Location" are mostly included local values provided by each repository and thus cannot be considered for interlinking. Additionally, constant values (e.g., dates and times) or controlled vocabularies (e.g., "Contribute.Role" and "Aggregation.Level") were not suitable for interlinking, as the user could not obtain useful information by linking these elements. In our previous study [44] we showed that the following metadata elements had the greatest possibility of interlinking to the LOD datasets:

- The time, culture, geography, or region to which the learning resource applies ("General.Coverage")
- The taxonomy given to a learning resource ("Classification.Taxon")
- A keyword or phrase describing the topic of learning objects ("General.Keyword").

As a consequence of the interlinking process, around 815,000 metadata files were harvested through the OAI-PMH [45] protocol from the GLOBE repository. Some GLOBE metadata could not be harvested due to validation errors (e.g., LOM extension errors). Particularly, several repositories in GLOBE extended the IEEE LOM by adding new elements without using namespaces, which caused a number of errors detected by the ARIADNE validation service. Analyzing the harvested records, more than half of the resources (55%) were in English language and almost all of them were free (99% without cost). From a technical point of view, around 256,898 resources (31%) with more than one million repetitions in GLOBE provided taxonomies in the metadata, of which 162,203 records (20%) with almost 524,000 repetitions were in English language. Figure 1 illustrates the top taxonomies of the metadata categorized according to their string values. The Y-axis in the diagram indicates the number of resources in thousands.

Table 2. Engineering related data in GLOBE

Title	Number
Resources including taxonomies in English language	162,203
Engineering-related resources (ERR)	39,801(25%)
Total ERR resources provided including *coverage* values	5,221
Total ERR resources including *taxon* values	39,801
Total ERR resources including *keyword* values	17,006

In order to identify the engineering resources within the GLOBE metadata, we carried out a comparative study between the "classification" category in the metadata and the latest version of the hierarchical ACM classification [46] system for some information about computing as the world's largest educational and scientific computing society. As a result, we found 39,801 records that were matched either to the ACM taxonomies or contained engineering values in the classification element. As shown in Table 2, there were almost 5,200 engineering-related resources (ERR) in GLOBE that included a "coverage" element in the metadata, while the number of ER records including the keyword element was 17,006.

To expose the former elements as RDF, we installed a D2RQ service [47], which is a mapping service for exploiting relational database as LD format (e.g., RDF, N3). To this end, we converted the harvested metadata files, which were in XML format, into a relational database. As a result, the GLOBE engineering data was accessible through a local SPARQL endpoint in order to be interlinked to the DBpedia dataset.

4. Interlinking results and discussion

As we discussed earlier in this paper, LIMES was selected for link discovery between GLOBE and DBpedia. This tool generates links between items contained in two datasets via a SPARQL endpoint or RDF dump. Users can set a threshold in LIMES for the metric above in which two entities are considered to match one another, and another threshold (e.g., 50%) for manual evaluation of the results. Interlinking can be performed either via a SPARQL endpoint or through an RDF dump. In the case of GLOBE, we set up a SPARQL endpoint for interlinking, as an RDF dump of a huge collection would have been too large and hard to parse. The SPARQL endpoints of datasets, similarity measurements, and acceptance or review conditions are set up by the user as software configurations. After running the tool, the result of interlinking obtained in two separate log files (matched concepts, and concepts for user review) is presented to

Table 3. Matches found between GLOBE and DBpedia

Element	Number of matches	Total records
Coverage	1,468 (28%)	5,221
Keyword	10,341 (60%)	17,006
Taxon	27,099 (68%)	39,801

Table 4. Similarity between GLOBE and DBpedia for manual review

Element	Number of similarities	Total records
Coverage	3,422 (65%)	5,221
Keyword	9,414 (55%)	17,006
Taxon	17,477 (43%)	39,801

the user. In the case of GLOBE, we set the matching threshold as 98% and the review threshold as 50% for manual evaluation of the results. In the following sub-sections, we will outline the interlinking results along with the human evaluation of discovered links.

4.1 Semi-Automatic Interlinking

The LOD cloud includes a wide variety of datasets that can be applied for linking entities. In this paper, we used the DBpedia dataset, which includes structured information about persons, places, and organizations. The full DBpedia dataset features labels and abstracts for 10.3 million unique topics in 111 different languages [48]. Hence, this dataset was selected for linking keywords and taxonomies of metadata. This dataset also fits the coverage element of GLOBE metadata, including places, countries and regions applicable to the learning objects. We will discuss the interlinking results to this dataset in detail.

As a result (see Table 3), values in the "Coverage" element of ERR in GLOBE have been exactly matched to 1,468 (out of 5,221) regions and places of the DBpedia dataset. Keyword and taxonomy are two elements of the LOM metadata frequently used to classify learning objects. To this end, we focused on the DBpedia classification [49] and

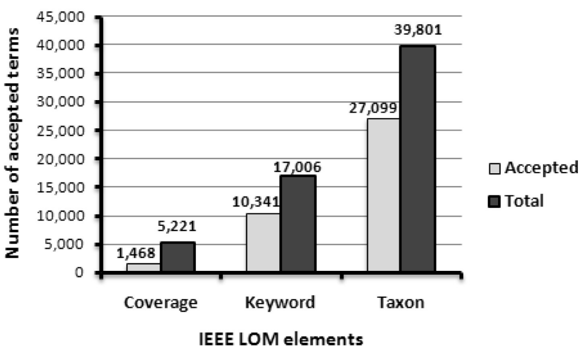

Fig. 2. Total number of similarities between GLOBE and DBpedia

10,341 (out of 17,006) keywords were found by LIMES as matching the DBpedia category, while the number of matched taxonomies was around 27,099 (68%) concepts.

As it can be seen from Table 4, there exist a wide variety of records in the GLOBE repository that had similarities (not exactly matched to the target dataset) to the DBpedia concepts and were recommended to the user for review. As the records did not fully match (with more than 50% similarity) the terms, they have been manually reviewed. Some examples of the results (matched and similar terms) are presented in the Appendix 1. Fig. 2 also depicts total accepted terms between GLOBE and DBpedia (exactly and nearly matched).

With respect to the similar concepts identified by the LIMES tool, the values of the "Coverage" element in both the Dbpedia and GLOBE datasets had a 1-n relationship, as each country could be assigned to many regions of the GLOBE metadata. However, some of the similar concepts did not seem to be semantically accurate. For instance, "History of Portugal" and "History of Science" were two terms identified as similar concepts, but they point to different data. "Keyword" values in both the GLOBE and Dbpedia datasets, as another example, had an m-n relationship, as each "Keyword" was connected to several resources of the Dbpedia (and vice versa). Appendix 1 illustrates some samples of these similarities.

4.2 Manual evaluation

As we discussed earlier, most interlinking tools present two types of results, i.e. matched and similar concepts. When analyzing the matched concepts outlined by the tools, undoubtedly both terms in GLOBE and DBpedia were the same from a string pattern-matching criterion (consider Appendix 1). As a consequence of evaluating similar terms by human expert, we presented hundred records of each result to three domain experts to assess as well as classify each one in a specific category. In the examination phase, the following possibilities might occur for each term:

- Matched: Two concepts are exactly the same (e.g., "Italy" in the GLOBE metadata and "Italy" (country) in the DBpedia dataset)
- Related: Two terms are not the same, but they have a relationship with each other as follows:
  - isPartOf: The source concept in GLOBE is a physical or logical part of the target concept. (e.g., "Mexico City" and "Mexico")
  - isParentOf: The DBpedia concept is a physical or logical part of the GLOBE concept. (e.g., "Nuclear Energy Companies" in DBpedia and "Nuclear energy" in GLOBE)

- isRelatedTo: The source and target concepts have various kinds of relationships (except isPartOf and isParentOf) with each other. (e.g., “criticism” and “Criticism of journalism”)
- isnotRelated: The similar source and target terms have string similarities, but they are not conceptually the same. (e.g., “Pacific Islands” with “Cayman Islands”)

The following table illustrates the results of the manual evaluation from the domain experts. The number in the table depicts the average number of concepts examined by the experts. Some of the similar concepts that had more than 80% string similarity were detected as “exactly matched” by experts. This was mostly correct particularly for those elements that have a close relationship to the target dataset (e.g., “Urban Studies and Planning” with “Planning and Urban Research”). The “Coverage” element, as an example, had 32 concepts exactly matched to the target dataset by LIMES (e.g., “Niger, Africa” with “Niger”). In “Keyword” and “Taxon” elements, the isParentOf relationship had the most similarities among other kinds of measurements, while the isNotRelated relationship was not found among them. In particular, a majority of concepts in DBpedia were physically or logically part of the GLOBE keywords. To take an example for the Keyword element, we found around 24 concepts in DBpedia which all were part of University of Cambridge as one of the keywords in the GLOBE dataset (e.g., “University of Cambridge examinations” or “Alumni of Cambridge University”) and this means that the term in GLOBE was the parent of those concepts in DBpedia (isParentOf relationship). In regard with “Taxon” element, there were found 10 concepts in DBpedia as part of martial arts (e.g., “Hybrid martial art” or “German martial arts”). On the contrary, 31 cases were found with isNotRelated relationships in the “Coverage” element in both DBpedia and GLOBE. Given that the “Coverage” element of learning resources in GLOBE mostly point to geographical places, the interlinking tool identified some concepts like “North America” and “Korea North” as similar concepts, while conceptually they are different and thus we categorized them as isNotRelated relationships. On the other hand, most of countries in DBpedia and GLOBE were exactly matched with each other, as the context of these two datasets on this element was very close.

Table 5. Average number of concepts (out of 100 sample records) reported by LIMES evaluated by three experts for LOM elements

Element	Similarity type	Average	%
Keyword	Exactly matched	4	4%
	GLOBE **is part of** DBpedia	8	8%
	GLOBE **is parent of** DBpedia	82	82%
	GLOBE **is related to** DBpedia	6	6%
	Is not related	0	0%
Taxon	Exactly matched	3	3%
	GLOBE **is part of** DBpedia	9	9%
	GLOBE **is parent of** DBpedia	84	84%
	GLOBE **is related to** DBpedia	4	4%
	Is not related	0	0%
Coverage	Exactly matched	32	32%
	GLOBE **is part of** DBpedia	17	17%
	GLOBE **is parent of** DBpedia	14	14%
	GLOBE **is related to** DBpedia	6	6%
	Is not related	31	31%

The foregoing discussion shows that it seems fair to conclude that when the context of the metadata element is more related to the target dataset (e.g., we consider “Coverage” element to DBpedia places), the result will include a greater frequency of matched results.

5. Conclusion

In this paper we evaluated the interlinking of engineering resources in the GLOBE repository to the DBpedia dataset. Considering various interlinking tools, we chose LIMES as our linking approach. After exposing the GLOBE metadata as LD, we analyzed the similarities of many entities in this collection and other existing datasets in the LOD cloud, such as DBpedia.

The GLOBE resources include valuable educational metadata that can be enriched when they are applied in linkable ways. By linking to the related datasets on the Web, the GLOBE users can get more valuable information about the learning resources. The “Coverage” that applies to learning resources can be linked, for example, to DBpedia places or other datasets such as Eurostat as long as it includes places (e.g. countries, cities). The more data provided in the DBpedia dataset (e.g., population, statistics data), the better help for users to obtain useful and enriched information. Furthermore, when the GLOBE resources are linked to the SKOS classification of the DBpedia, they can be discovered by any LD application, particularly those that use the SKOS classification for their search process. There exist 11 million triples in the DBpedia dataset, out of which 1.7 million triples include the SKOS category, which was conjoined with the GLOBE metadata.

Manual evaluation of the interlinking outcome by domain experts also showed that the GLOBE resources definitely have the potential to be linked to the related datasets, as we found special relationships among the results (e.g., isParentOf and isPartOf) that can be used for linking the terms in the GLOBE repository and other datasets. Based on our analysis, other datasets also exist in the LOD that can be interlinked to the GLOBE materials, when they include learning content. Linking more related and linkable datasets in the LOD cloud to huge educational repositories provides users with more flexibility to expand their knowledge regarding the source collections.

Acknowledgments—The work presented in this paper has been part-funded by the European Commission under the ICT Policy Support Programme with project No. 297229 "Open Discovery Space (ODS)" and project No. 27099 "Organic.Lingua".

References

1. E. Godfrey, T. Aubrey and R. King, Who leaves and who stays? Retention and attrition in engineering education, *Engineering Education*, **5**(2), Dec. 2010, pp. 26–40.
2. J. C. Perrenet, P. A. J. Bouhuijs and J. G. M. M. Smits, The Suitability of Problem-based Learning for Engineering Education: theory and practice, *Teaching in Higher Education*, **5**(3), July 2000, pp. 345–358.
3. R. M. Felder, D. O. C. Engineering, P. H. Mohr, C. O. Engineering and E. J. Dietz, A longitudinal study of engineering student performance and retention. I. Success and failure in the introductory course, *Journal of Engineering Education*, 1993, pp. 15–21.
4. C. E. Hmelo-Silver, Problem-Based Learning: What and How Do Students Learn?, *Educational Psychology Review*, **16**(3), Sept. 2004, pp. 235–266.
5. E. Montero and M. J. Gonzalez, Student Engagement in a Structured Problem-Based Approach to Learning: A First-Year Electronic Engineering Study Module on Heat Transfer, *IEEE Transactions on Education*, **52**(2), May 2009, pp. 214–221.
6. D. Hernandez-Leo, V. M. Oliver, J. M. Dodero, A. Pardo, M. Romero-Ternero, Y. Dimitriadis and J. I. Asensio-Perez, Applying Recommendations to Align Competences, Methodology, and Assessment in Telematics, Computing, and Electronic Engineering Courses, *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, **8**(1), Feb. 2013, pp. 15–22.
7. J. Mills and D. Treagust, Engineering Education—is problem-based or project-based learning the answer?, *Australasian Journal of Engineering Education*, 2003.
8. H. Q. Yu, C. Pedrinaci, S. Dietze and J. Domingue, Using Linked Data to Annotate and Search Educational Video Resources for Supporting Distance Learning, *IEEE Transactions on Learning Technologies*, **5**(2), pp. 130–142, 2012.
9. M. Fernandez, M. d'Aquin and E. Motta, Linking data across universities: an integrated video lectures dataset, *Proceedings of the 10th international conference on The semantic web*—Volume Part II, Berlin, Heidelberg, 2011, pp. 49–64.
10. C. Bizer, T. Heath and T. Berners-Lee, Linked Data—The Story So Far, *International Journal on Semantic Web and Information Systems*, **5**(3), 2009, pp. 1–22.
11. S. Dietze, S. Sanchez-Alonso, H. Ebner, H. Q. Yu, D. Giordano, I. Marenzi and B. P. Nunes, Interlinking educational resources and the web of data: A survey of challenges and approaches, *Program: electronic library and information systems*, **47**(1), Feb. 2013, pp. 60–91.
12. "DBpedia dataset", Available at: http://dbpedia.org [Accessed: 22 February 2014].
13. E. Rajabi, S. Sanchez-Alonso and M.-A. Sicilia, Analyzing Broken Links on the Web of Data: an Experiment with DBpedia, *Journal of the Association for Information Science and Technology*, **65**, pp. 1721–1727, doi: 10.1002/asi.23109.
14. M. D'Aquin, Linked Data for Open and Distance Learning, Common Wealth of Learning, 2012. [Online]. Available: http://elearningeuropa.info/en/directory/Linked-Data-for-Open-and-Distance-Learning. [Accessed: 22 February 2014].
15. S. Dietze, H. Q. Yu, D. Giordano, E. Kaldoudi, N. Dovrolis, and D. Taibi, Linked education: interlinking educational resources and the Web of data, *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, New York, NY, USA, 2012, pp. 366–371.
16. A. Ngonga and A. Sören, LIMES—A Time-Efficient Approach for Large-Scale Link Discovery on the Web of Data, *presented at the IJCAI*, 2011.
17. C. L. Dym, A. M. Agogino, O. Eris, D. D. Frey and L. J. Leifer, Engineering Design Thinking, Teaching, and Learning, *Journal of Engineering Education*, **94**(1), 2005, pp. 103–120.
18. University of Muenster Linked Data, Available at: http://education.data.gov.uk/ [Accessed: 22 February 2014].
19. Open University (UK) dataset, Available at: http://data.open.ac.uk/ [Accessed: 22 February 2014].
20. National Research Council (CNR, Italy) dataset, Available at: http://data.cnr.it [Accessed: 22 February 2014].
21. Southampton University dataset, Available at: http://data.southampton.ac.uk/ [Accessed: 22 February 2014].
22. Organic.Edunet Project, Available at: http://wiki.organic-edunet.eu/index.php/Main_Page [Accessed: 22 February 2014].
23. Organic.Edunet dataset, Available at: http://data.organic-edunet.eu [Accessed: 22-February-2014].
24. M.-A. Sicilia, H. Ebner, S. Sanchez-Alonso, F. Alvarez, A. Abian, and E. Garcia-Barriocanal, Navigating learning resources through linked data: A preliminary report on the re-design of Organic.Edunet, *1st International Workshop on eLearning Approaches for the Linked Data Age*, 2011.
25. B. Haslhofer and A. Isaac, data.europeana.eu—The Europeana Linked Open Data Pilot, *DCMI International Conference on Dublin Core and Metadata Applications*, The Hague, The Netherlands, 2011.
26. Europeana dataset, Available at: http://data.europeana.eu [Accessed: 22 February 2014].
27. GeoNames, Available at: http://www.geonames.org/. [Accessed: 22 February 2014].
28. GEMET Dataset, Available at: http://www.eionet.europa.eu/gemet/. [Accessed: 22 February 2014].
29. Turtle—Terse RDF Triple Language, W3C Team Submission 28 March 2011, Available at: http://www.w3.org/TeamSubmission/turtle/. [Accessed: 22 February 2014].
30. E. Simperl, S. Wölger, S. Thaler, B. Norton and T. Bürger, Combining human and computation intelligence: the case of data interlinking tools, *International Journal of Metadata, Semantics and Ontologies*, **7**(2), Oct. 2012, pp. 77–92.
31. M. Hausenblas, W. Halb and Y. Raimond, Scripting User Contributed Interlinking, *Proceedings of the 4th workshop on Scripting for the Semantic Web (SFSW2008)*, co-located with ESWC2008, 2008.
32. K. Siorpaes and M. Hepp, Games with a Purpose for the Semantic Web, *IEEE Intelligent Systems*, **23**(3), May 2008, pp. 50–60.
33. O. Hassanzadeh, R. Xin, R. J. Miller, A. Kementsietsidis, L. Lim and M. Wang, Linkage Query Writer , *Proceedings of the VLDB Endowment*, 2009.
34. M. Hausenblas, Interlinking Multimedia: How to Apply Linked Data Principles to Multimedia Fragments, *LDOW*, 2009.
35. T. Bürger and M. Hausenblas, Interlinking Multimedia—Principles and Requirements, *Proceedings of the International Workshop on Interacting with Multimedia Content in the Social Semantic Web (IMC-SSW'08)*, 2008.
36. F. Scharffe, Y. Liu and C. Zhou, RDF-AI: an architecture for RDF datasets matching, fusion and interlink, *Proceeding of IJCAI*, IR-KR Work, 2009.
37. O. Hassanzadeh, A. Kementsietsidis, L. Lim, R. J. Miller, and M. Wang, Semantic Link Discovery over Relational Data, *Semantic Search over the Web*, 2012, pp. 193–223.
38. Google-refine—Google Refine, a power tool for working with messy data (formerly Freebase Gridworks), Available at: https://code.google.com/p/google-refine/. [Accessed: 22 February 2014].
39. M. Verlic, LODGrefine-LOD-enabled Google Refine in Action, *presented at the I-SEMANTICS (Posters & Demos)*, 2012.
40. E. Rajabi, M.-A. Sicilia and S. Sanchez-Alonso, An empiri-

cal study on the evaluation of interlinking tools on the Web of Data, *Journal of Information Science*, first published on June 11, 2014 doi:10.1177/0165551514538151.
41. X. Ochoa, J. Klerkx, B. Vandeputte and E. Duval, On the use of learning object metadata: the GLOBE experience, *Proceedings of the 6th European conference on Technology enhanced learning: towards ubiquitous learning,* Berlin, Heidelberg, 2011, pp. 271–284.
42. OER Commons, Available at: http://www.oercommons.org/. [Accessed: 22 February 2014].
43. IEEE LTSC, IEEE Standard for Learning Object Metadata 1484.12.1-2002, Final Draft version, [Online]. Available: http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html. [Accessed: 22 February 2014].
44. E. Rajabi, M.-A. Sicilia and S. Sanchez-Alonso, Interlinking Educational Data: an Experiment with GLOBE Resources, *presented at the First International Conference on Technological Ecosystem for Enhancing Multiculturality*, Salamanca, Spain, 2013.
45. Open Archives Initiative Protocol for Metadata Harvesting, Available at: http://www.openarchives.org/pmh. [Accessed: 22 February 2014].
46. The 2012 ACM Computing Classification System. Available at: https://www.acm.org/about/class/2012 [Accessed: 22 February 2014].
47. C. Bizer, D2RQ—treating non-RDF databases as virtual RDF graphs, *Proceedings of the 3rd International Semantic Web Conference (ISWC2004)*, 2004.
48. DBpedia blog, Available at: http://blog.dbpedia.org, [Accessed: 22 February 2014].
49. DBpedia classification, Available at: http://dbpedia.org/page/Category, [Accessed: 22 February 2014].

Appendix

Sample interlinking results

Tool	GLOBE	DBpedia	Element
LIMES matches	Montreal Copenhagen Victoria (Australia)	http://dbpedia.org/resource/Montreal http://dbpedia.org/resource/Copenhagen http://dbpedia.org/resource/Victoria (Australia)	Coverage
LIMES review	Tibet Rickenbach Medel	http://dbpedia.org/resource/Taibet http://dbpedia.org/resource/Krickenbach http://dbpedia.org/resource/Medeo	Coverage
LIMES matches	Mechanical engineering biometrics Addition reactions	http://dbpedia.org/resource/Category:Mechanical_engineering http://dbpedia.org/resource/Category:Biometrics http://dbpedia.org/resource/Category:Addition_reactions	Keywords
LIMES review	networks Phoenician revenue	http://dbpedia.org/resource/Category:Social_networks http://dbpedia.org/resource/Category:Phoenicia http://dbpedia.org/resource/Category:Revenge	Keywords
LIMES matches	Teleconferencing Linear algebra Project Management	http://dbpedia.org/resource/Category:Teleconferencing http://dbpedia.org/resource/Category:Linear_algebra http://dbpedia.org/resource/Category:Project_Management	Taxonomy
LIMES review	economic system Brics Queueingtheory	http://dbpedia.org/resource/Category:Economic_Systems http://dbpedia.org/resource/Category:Brick http://dbpedia.org/resource/Category:Queueing_theory	Taxonomy

Enayat Rajabi is a PhD researcher in the Computer Science Department of the University of Alcalá since 2012. He holds his MSc (2004) degree in Computer Engineering from the department of Ferdowsi University of Mashhad (Mashhad, Iran). He is presently a full time researcher in the Information Engineering research unit at the University of Alcalá (Alcalá de Henares, Madrid, Spain). The aim of his PhD is to interlink educational data using Semantic Web and Linked Data.

Miguel-Angel Sicilia is currently full professor at the Computer Science Department of the University of Alcalá and the head of the Information Engineering research unit, where he leads several European and national research projects in the topics of learning technology and Semantic Web. He is the Editor-in-Chief of the International Journal of Metadata, Semantics and Ontologies, published by Inderscience, and serves as a member of editorial board of many other scientific journals in the area of Semantic Web, Computational Intelligence and Information Systems. He has been active in metadata research and was awarded the 2006 Cyc prize for the best research paper.

Salvador Sanchez-Alonso is an associate professor in the Computer Science Department of the University of Alcalá, Spain. He has participated in and coordinated several EU-funded projects over the last few years related to TEL, learning object repositories and federations, notably eContent+ Organic.Edunet, CIP-PSP Organic.Lingua and VOA3R,. His current research interests include TEL, Learning repositories, Semantic Web and Computer Science education.

Juan Manuel Dodero is currently an associate professor of the University of Cádiz, Spain and the head of the Office of Libre Software and Open Knowledge. He has a PhD in Computer Science, and his research interests include web engineering, web science and computer-assisted learning. He is the leader of a number of research projects, such as asceta.uca.es, related with his interests. He is a national coordinator of the OpenDiscoverySpace.eu project.