

Linked Data as facilitator for TEL recommender systems in research & practice

Stefan Dietze¹

¹L3S Research Center, Leibniz University, Hannover, Germany
dietze@l3s.de

Abstract. Personalisation, adaptation and recommendation are central features of TEL environments. In this context, information retrieval techniques are applied as part of TEL recommender systems to filter and deliver learning resources according to user preferences and requirements. However, the suitability and scope of possible recommendations is fundamentally dependent on the quality and quantity of available data, for instance, metadata about TEL resources as well as users. On the other hand, throughout the last years, the Linked Data (LD) movement has succeeded to provide a vast body of well-interlinked and publicly accessible Web data. This in particular includes Linked Data of explicit or implicit educational nature. The potential of LD to facilitate TEL recommender systems research and practice is discussed in this paper. In particular, an overview of most relevant LD sources and techniques is provided, together with a discussion of their potential for the TEL domain in general and TEL recommender systems in particular based on insights from highly related European projects, mEducator and LinkedUp.

Keywords. Linked Data, Education, Semantic Web, Technology-enhanced Learning, Data Consolidation, Data Integration

1 Introduction

As personalisation, adaptation and recommendation are central features of TEL environments, TEL recommender systems apply information retrieval techniques to filter and deliver learning resources according to user preferences and requirements. While the suitability and scope of possible recommendations is fundamentally dependent on the quality and quantity of available data, data about learners, and in particular metadata about TEL resources, the landscape of standards and approaches currently exploited to share and reuse educational data is highly fragmented.

The latter includes, for instance, competing metadata schemas, i.e., general-purpose ones such as Dublin Core¹ or schemas specific to the educational field, like IEEE Learning Object Metadata (LOM) or ADL SCORM² but also interface mechanisms such as OAI-PMH³ or SQI⁴. These technologies are exploited by educational

¹ <http://dublincore.org/documents/dces/>

² Advanced Distributed Learning (ADL) SCORM: <http://www.adlnet.org>

³ Open Archives Protocol for Metadata Harvesting
<http://www.openarchives.org/OAI/openarchivesprotocol.html>

resource repository providers to support interoperability. To this end, although a vast amount of educational content and data is shared on the Web in an open way, the integration process is still costly as different learning repositories are isolated from each other and based on different implementation standards [3].

In the past years, TEL research has already widely attempted to exploit Semantic Web technologies in order to solve interoperability issues. However, while the Linked Data (LD) [1] approach has widely established itself as the de-facto standard for sharing data on the Semantic Web, it is still not widely adopted by the TEL community. Linked Data is based on a set of well-established principles and (W3C) standards, e.g. RDF, SPARQL [5] and use of URIs, and aims at facilitating Web-scale data interoperability. Despite the fact that the LD approach has produced an ever growing amount of data sets, schemas and tools available on the Web, its take-up in the area of TEL is still very limited. Thus, LD opens up opportunities to substantially alleviate interoperability issues and to substantially improve quality, quantity and accessibility of TEL data.

2 Challenges

While there is already a large amount of educational data available on the Web via proprietary and/or competing schemas and interface mechanisms, the main challenge for improving impact of TEL recommender systems is to (a) start adopting LD principles and vocabularies while (b) leveraging on existing educational data available on the Web by non-LD compliant means. Following such an approach, major research challenges need to be taken into consideration towards Web-scale interoperability [3]:

- *Integrating distributed data from heterogeneous educational repositories*: educational data and content is usually exposed by heterogeneous services/APIs such as OAI-PMH or SQL. Therefore, interoperability is limited and Web-scale sharing of resources is not widely supported yet.
- *Metadata mediation and transformation*: educational resources and the services exposing those resources are usually described by using distinct, often XML-based schemas and by making use of largely unstructured text and heterogeneous taxonomies. Therefore, schema and data transformation (into RDF) and mapping are important requirements in order to leverage on already existing TEL data.
- *Enrichment and interlinking of unstructured metadata*: existing educational resource metadata is usually provided based on informal and poorly structured data. That is, free text is still widely used for describing educational resources while use of controlled vocabularies is limited and fragmented. Therefore, to allow machine-processing and Web-scale interoperability, educational metadata needs to be enriched, that is transformed into structured and formal descriptions by linking it to widely established LD vocabularies and datasets on the Web.

Our work builds on the hypotheses that Linked Data offers high potential to improve take-up and impact of TEL recommender systems and introduces key past and future

⁴ Simple Query Interface: <http://www.cen-itsa.net/main.aspx?put=859>

projects which serve as building blocks towards *Linked Education*⁵, i.e. educational data sharing enabled by adoption of Linked Data principles.

3 Towards TEL data integration and exploitation

In particular, we focus on two projects which address the aforementioned challenges by providing innovative approaches towards (a) integration of heterogeneous TEL data (as part of the *mEducator*⁶ project) and (b) exploitation of educational open data addressed by the *LinkedUp*⁷ project. With respect to (a) we identify a set of principles (see [2][6]) to address the above challenges:

- (P1) **Linked Data-principles:** are applied to model and expose metadata of both educational resources and educational services and APIs. In this way, resources are interlinked but also services' description and resources are exposed in a standardized and accessible way.
- (P2) **Services integration:** Existing heterogeneous and distributed learning repositories, i.e. their Web interfaces (services) are integrated on the fly by reasoning and processing of LD-based service semantics (see P1).
- (P3) **Schema matching:** metadata retrieved from heterogeneous Web repositories, for instance is automatically lifted into RDF, aligned with competing metadata schemas and exposed as LD accessible via de-referenceable URIs.
- (P4) **Data interlinking, clustering and enrichment:** Automated enrichment and clustering mechanisms are exploited in order to interlink data produced by (P3) with existing datasets as part of the LD cloud.

While this work aims at increasing the quantity, quality and accessibility of available educational data on the Web, LinkedUp addresses (b) by aiming to push forward the exploitation of the vast amounts of public, open data available on the Web, in particular by educational institutions and organizations. This will be achieved by identifying and supporting highly innovative large-scale Web information management applications through an open competition (the *LinkedUp Challenge*) and dedicated evaluation framework. The vision of the LinkedUp Challenge is to realise personalised university degree-level education of global impact based on open Web data and information. Drawing on the diversity of Web information relevant to education, ranging from OER metadata to the vast body of knowledge offered by the LD approach, this aim requires overcoming substantial challenges related to Web-scale data and information management involving Big Data, such as performance and scalability, interoperability, multilinguality and heterogeneity problems, to offer personalised and accessible education services. Therefore, the LinkedUp Challenge provides a focused scenario to derive challenging requirements, evaluation criteria, benchmarks and thresholds which are reflected in the LinkedUp evaluation

⁵ <http://linkededucation.org>: an open platform to share results focused on educational LD. Long-term goal is to establish links and unified APIs and endpoints to educational datasets.

⁶ <http://www.meducator.net>

⁷ LinkedUp: Linking Web Data for Education Project – Open Challenge in Web-scale Data Integration (<http://www.linkedup-project.eu>)

framework. Information management solutions have to apply data and learning analytics methods to provide highly personalised and context-aware views on heterogeneous Web data. Building on the strong alliance of institutions with expertise in areas such as open Web data management, data integration and Web-based education, key outcomes of LinkedUp include a general-purpose evaluation framework for Web-data driven applications, a set of quality-assured educational datasets, innovative applications of large-scale Web information management, community-building and clustering crossing public and private sectors and substantial technology transfer of highly innovative Web information management technologies.

4 Conclusions

We provided an overview of two efforts both aiming at the overall goal of fostering the reuse of open educational data on the Web. While the accessibility of large-scale amounts of data is a foundation for TEL recommender systems, both efforts contribute to improvements in scope, quantity and quality of recommendations in TEL environments. This includes both, TEL recommender systems in research, where data is required for evaluation and benchmarking, as well as in practice, where data is a core requirement for offering suitable recommendations to users.

Acknowledgments

This work is partly funded by the European Union under FP7 Grant Agreement No 317620 (LinkedUp).

References

- [1] Bizer, C., T. Heath, Berners-Lee, T. (2009). Linked data - The Story So Far. Special Issue on Linked data, International Journal on Semantic Web and Information Systems.
- [2] Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N. and Taibi, D. (2012), "Linked Education: interlinking educational Resources and the Web of Data", *Proceedings of the 27th ACM Symposium On Applied Computing (SAC-2012), Special Track on Semantic Web and Applications*, Riva del Garda (Trento), Italy, 2012.
- [3] Dietze, S., Sanchez-Alonso, S., Ebner, H., Yu, H., Giordano, D., Marenzi, I., Pereira Nunes, B. (2013) Interlinking educational Resources and the Web of Data – a Survey of Challenges and Approaches, accepted for publication in Emerald Program: electronic Library and Information Systems, Volume 47, Issue 1 (2013).
- [4] IEEE (2002), "IEEE Standard for Learning Object Metadata", *IEEE Std 1484.12.1-2002*, pp.i-32. doi: 10.1109/IEEESTD.2002.94128.
- [5] World Wide Web Consortium (2008). *W3C Recommendation, SPARQL query language for RDF*, 2008, available at: www.w3.org/TR/rdf-sparql-query/
- [6] Yu, H. Q., Dietze, S., Li, N., Pedrinaci, C., Taibi, D., Dovrolis, N., Stefanut, T., Kaldoudi, E. and Domingue, J. (2011), "A linked data-driven & service-oriented architecture for sharing educational resources", in *Linked Learning 2011*, Proceedings of the 1st International Workshop on eLearning Approaches for Linked Data Age, May 29, 2011, Heraklion, Greece.