Exploring the Linked University Data With Visualization Tools

Miika Alonen, Tomi Kauppinen, Osma Suominen, Eero Hyvönen

Semantic Computing Research Group (SeCo)
Aalto University, Department of Media Technology
`firstname.lastname@aalto.fi, http://www.seco.tkk.fi/`

Abstract. University data is typically stored in separate data silos even though the data is implicitly richly related together. Such data has a large and diverse user base, including faculty members, students, industrial partners, alumnis, collaborating universities, and media. In this paper, we demonstrate two tools for understanding and using the contents of linked university data. The first tool, Visualization Playground (VISU), supports querying and visualizing the data for example for illustrating emerging trends in universities (e.g., about publications) and for comparing differences. The second tool, Vocabulary Visualizer (V^2), demonstrates the usage of vocabularies in the Linked University Data Cloud. It reveals what kinds of data different universities have published, and what terms are used to describe the contents. Such analysis is a basis for facilitating design of Linked Data applications across university data boundaries.

1 Towards Linked University Data

Data production and knowledge publication in universities are traditionally based on separate data silos for different data types and domains. Such silos include data such as publication information, course and event descriptions, educational materials, web pages and news feeds. University information systems have traditionally been implemented without considering opening the data stored in there and how it could be done. Another big challenge with separated data silos is the wide diversity of data models and practices in use. Linked Open Data (LOD) principles and technologies enable universities to publish their legacy data with shared open standards, and offer a variety of approaches for integrating university contents with the existing Web of Data[1]. The promise is that the use of LOD technologies supports academic organizations to be more transparent, comparable, and even more open for new ideas.

Linked Universities[1] is a collaboration alliance and application scenario where open datasets from universities are published and linked together using the 5-star methodology[2]. Several universities[3] have already published SPARQL endpoints

---

[1] `http://linkeduniversities.org`

[2] `http://5stardata.info/`

[3] `http://linkeduniversities.org/lu/index.php/datasets-and-endpoints/`

for accessing their contents as Linked Data. These efforts enable applications to use enriched data in novel ways[2,4].

Our Linked Open Aalto (LOA) project[4] develops a Linked Data infrastucture for the Aalto University. One of the results so far is the Linked Open Aalto Data Service portal[5], an ongoing collaborative effort with different schools and service providers of Aalto. The bigger goal is that the portal connects data from the Aalto University to data from other Linked Universities, for example from The Open University[6], the University of Southampton[7], the University of Münster[8] and the University of Bristol[9]. The data in the Linked University Cloud consists of publications, profiles of people, course and event descriptions, educational materials, project information, and service descriptions.

Use cases and advantages of using the Linked University Data have been shown in projects like Lucero [6] and Linked Open Data University of Münster (LODUM) [4] by demonstrating the use of Linked University Data in novel ways. The main focus of the projects so far has been to show the usefulness of linked data within one university. However, the full potential of Linked Open Universities can be achieved by interlinking resources across the universities. Use cases for the web of university data include at least the following:

1. **Linking scientific assets** Connecting publications, patents, projects, people and other scientific assets created by universities and other research communities could foster the reuse of limited resources, bring out new ideas, and facilitate the creation of new projects.
2. **Linking educational materials** Linking related courses, study materials, and other educational assets across universities help students and teachers to find relevant educational material [3].
3. **Finding funding and career offers** Linking researchers with related opportunities could help in increasing mobility and for building social networks among researchers.
4. **Analyzing the web of university data** Linked Data can be analyzed statistically in order to get an insight of activities, trends, and other phenomena in universities. For these purposes, methods such as combining the R statistical computing environment with SPARQL are available[10].

2 Tools for Exploring the Linked University Data

Shared open vocabularies facilitate the semantic interoperability between datasets. However, this happens only if either the same terms are used in the datasets or if

---

[4] `http://www.seco.tkk.fi/projects/loa/`

[5] `http://data.aalto.fi`

[6] `http://data.open.ac.uk`

[7] `http://data.southampton.ac.uk/`

[8] `http://data.uni-muenster.de`

[9] `http://resrev.ilrt.bris.ac.uk/research-revealed-hub/`

[10] See our accompanying contribution to analyze university contents with R `http://linkedscience.org/tutorials/analyzing-and-visualizing-linked-data-from-the-aalto-university-with-r`

mappings relate the terms. Well-defined vocabularies and their mappings make it possible to aggregate data from different universities. In practice this is still very challenging without the knowledge of the vocabularies used in the datasets. Our contribution for this problem are two visualization tools: VISUalization playground and Vocabulary Visualizer (V^2). They explicate 1) what datasets and vocabularies are used in separate endpoints 2) and to show similarities and differences between the use of classes and properties between given endpoints.

The tools were implemented using a set of JavaScript libraries[11] for visualizing the SPARQL-query results with standard Scalable Vector Graphics (SVG). This enables the processing of the queries on the client side without server-side help other than a proxy. The proxy is in use to overcome cross-domain issues with some of the endpoints.

2.1 Visualization Playground VISU

The visualization playground VISU[12] is a novel SPARQL interface for creating data visualizations. The goal of our work was to create a flexible and easy-to-use tool for exploring and visualizing data from Linked Universities. VISU can aggregate query results from multiple endpoints, manipulate the results, customize the visualization, and export the data in several formats.

Visualizations are created by using a query that is sent to the selected SPARQL endpoints. The SPARQL-query responses are tabular variable bindings, which are processed in a way similar to Sgvizler[5]. Here VISU utilizes the Google Chart editor that can be used create the visualizations on the fly. Query-reponses from multiple endpoints are processed and the resources using the same URI-identifiers are aggregated in the resulting visualization. For example, the comparison of vocabulary usage (see Figure 1) in separate endpoints can be easily implemented with the playground. Using VISU is further demonstrated in our accompanying tutorial for data exploration and information visualization[13].

Fig. 1. Vocabulary usage in multiple endpoints

2.2 Vocabulary Visualizer

Designing queries to visualize and explore relationships between multiple endpoints can be very time-consuming. Vocabulary Visualizer[14] (V^2) supports exploring and comparing the vocabulary usage in multiple SPARQL endpoints. V^2

---

[11] D3, see http://d3js.org/ and Google Charts, see https://developers.google.com/chart/

[12] http://data.aalto.fi/visu

[13] http://linkedscience.org/tutorials/sparql-tutorial-for-data-exploration-and-information-visualization/

[14] http://data.aalto.fi/V2

extracts metadata in different SPARQL endpoints using concurrent SPARQL queries, and then visualizes the joint use of the vocabularies.

V^2 consists of two views. 1) *Class usage view* provides a high level abstraction of used classes in the endpoints. Class definitions that are used by the given endpoints are visualized using set diagrams and pie charts. In other words, they represent the joint use of classes and the number of class instances. 2) *Property usage view* (see Figure 2) visualizes the usage of properties in the instances of a certain class. The idea is that this analysis supports for designing the use of terms for existing and new datasets and mappings between the existing terms.

Fig. 2. Usage of properties in separate endpoints

3   Conclusions

We argued that the usefulness of the Linked University Data is based on solid standards and reusability of the data by different user groups and universities. A key requirement for this is to use shared vocabularies. The number of vocabularies used in the linked datasets is likely to increase as new data from different domains are being opened. This leads to interoperability problems in linking data across datasets.

The two tools introduced in the paper address these fundamental problems by proving an end-user new means for analyzing vocabulary usage, and for comparing and contrasting Linked Data across different endpoints. 1) VISUalization Playground is an interactive tool for specifying and creating visualizations for exploring and comparing the linked data. 2) Vocabulary Visualizer (V^2) enables the comparison of Linked Data by revealing the metadata stuctures, number of instances and the actual usage of vocabularies across the endpoints in novel ways. It supports endpoint developers to take account other relevant datasets and to foster reuse of shared vocabularies to enable better interoperability between the endpoints.

Future versions of the tool should also use the relationships between terms for the evaluation of the vocabulary usage. Another future task is to add functionality to suggest term mappings and point out issues like range and domain violations. The tools presented were developed in the context of Linked University Data. However, the tools and ideas behind them are not domain specific and can thus be adapted to virtually any SPARQL endpoints in the Linked Open Data Cloud.

References

1. Dietze, S., Yu, H.Q., Giordano, D., Kaldoudi, E., Dovrolis, N., Taibi, D.: Linked Education: Interlinking Educational Resources and the Web of Data. In: Proceedings

of the 27th Annual ACM Symposium on Applied Computing. pp. 366–371. ACM (2012)

2. d'Aquin, M.: Putting Linked Data To Use In a Large Higher-Education Organisation. Interacting with Linked Data (ILD 2012) p. 9 (2012)

3. Fernandez, M., d'Aquin, M., Motta, E.: Linking Data Across Universities: An Integrated Video Lectures Dataset. The Semantic Web–ISWC 2011 pp. 49–64 (2011)

4. Kessler, C., Kauppinen, T.: Linked Open Data University of Muenster—Infrastructure and Applications. In: Demos the Extended Semantic Web Conference 2012 (ESWC2012). Heraklion, Crete, Greece (May 2012)

5. Skjæveland, M.: Sgvizler: A JavaScript Wrapper for Easy Visualization of SPARQL Result Sets. In: Extended Semantic Web Conference 2012 (ESWC2012). Heraklion, Crete, Greece (May 2012)

6. Zablith, Fouad and Fernandez, Miriam and Rowe, Matthew: The OU Linked Open Data: Production and Consumption. In: The Semantic Web: ESWC 2011 Workshops. pp. 35–49. Springer (2012)