Vocabulary Services Based on SPARQL Endpoints: ONKI Light on SPARQL

Osma Suominen, Alex Johansson, Henri Ylikotila,
Jouni Tuominen, and Eero Hyvönen

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

1 Introduction

Vocabularies are an integral part of Linked Data (LD). They are published as browsable services for humans, as data dumps, as LD services, and as SPARQL endpoints for machines. Centralized vocabulary services, such as Bioportal and ONKI [3], are emerging [1], and the SKOS standard is heavily used for representing controlled vocabularies as LD.

Vocabulary services, such as concept lookup via term searches and hierarchical browsing, are used by indexers describing documents and searchers looking for suitable keywords. Basing vocabulary services on SPARQL endpoints brings several benefits: 1) any vocabulary published by, e.g., a library as an endpoint can be provided easily as a service by anyone; 2) the vocabulary service is always up to date w.r.t. to the data published in the endpoint, allowing fast update cycles in vocabulary development; 3) federated vocabulary services can be created for accessing several vocabularies simultaneously.

There are challenges in realizing this dream. The research question we address is: *Can vocabulary services be based on SPARQL endpoints alone?* As a promising solution, we present the **ONKI Light on SPARQL** system.

2 Challenges

Using SPARQL endpoints for vocabulary services presents several challenges: 1) *Quality and Heterogeneity*: SKOS vocabularies have varying structure. This makes it difficult to develop SPARQL queries that can handle all special cases. Some of these issues can be mitigated by using Skosify [2], a tool which enforces best practices in expressing SKOS vocabulary structures. 2) *Text Search*: The current SPARQL 1.1 draft specification does not include functions for efficient text search. Search functionality can be implemented using filters, but these are often very slow. Many SPARQL engines provide proprietary text search functions. 3) *Global Search*: A vocabulary service may provide access to several vocabularies at once. A federated search approach could be used to search several SPARQL engines simultaneously. Vocabularies stored in distinct named graphs on a single endpoint may all be accessed in one query.

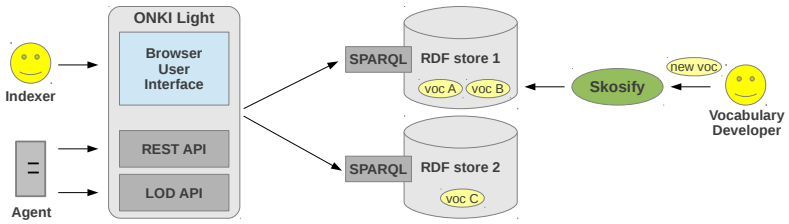

Fig. 1. ONKI Light on SPARQL Architecture.

3 Prototype Implementation

We implemented ONKI Light on SPARQL[1], a prototype demonstrating the SPARQL-only approach. The system consists of the ONKI Light application, including a multilingual browser interface and APIs for machine access, as well as SPARQL endpoints containing SKOS vocabularies (Figure 1). The system provides access to more than 15 vocabularies published using the ONKI infrastructure. These are automatically kept up to date and preprocessed with Skosify. The global search functionality currently relies on a single endpoint containing all the vocabularies. The system was implemented in PHP and currently uses a Jena Fuseki SPARQL endpoint and a TDB triple store with a LARQ text index. The code is available as open source under a MIT-style license[2].

Based on experiences with the current prototype, we can already state that the performance of the SPARQL-only approach is sufficient for routine use, with most response times well below 1 second. In future, we plan to support more SPARQL text search dialects, more vocabularies, and richer API access.

Acknowledgments This work is part of the National Semantic Web Ontology project in Finland FinnONTO[3] (2003–2012), funded mainly by Tekes. We thank Eetu Mäkelä and Kim Viljanen for their input into the design of the system.

References

1. Noy, N.F., d'Aquin, M.: Where to publish and find ontologies? A survey of ontology libraries. Web Semantics: Science, Services and Agents on the World Wide Web 11(0) (2011)
2. Suominen, O., Hyvönen, E.: Improving the Quality of SKOS Vocabularies with Skosify. In: Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012). Springer-Verlag (2012)
3. Tuominen, J., Frosterus, M., Viljanen, K., Hyvönen, E.: ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. In: Proceedings of the 6th European Semantic Web Conference (ESWC 2009). Springer-Verlag (2009)

[1] `http://light.onki.fi/`
[2] `http://code.google.com/p/onki-light/`
[3] `http://www.seco.tkk.fi/projects/finnonto/`