SKOS as a Key Element in Enterprise Linked Data Strategies

Andreas Blumauer

Semantic Web Company GmbH, Mariahilfer Straße 70/8,
1070 Vienna, Austria
a.blumauer@semantic-web.at

Abstract. The challenges in implementing linked data technologies in enterprises are not limited to technical issues only. Projects like these deal also with organisational hurdles to be crossed, for instance the development of employee skills in the area of knowledge modelling and the implementation of a linked data strategy which foresees a cost-effective and sustainable infrastructure of high-quality and linked knowledge graphs. SKOS is able to play a key role in enterprise linked data strategies due to its relative simplicity in parallel with its ability to be mapped and extended by other controlled vocabularies, ontologies, entity extraction services and linked open data.

1 Introduction

The use of semantic web methodologies and technologies has been perceived to be an appealing solution approach for various issues in enterprise information management and data integration [1][2]. Amongst others, the following application scenarios are typically discussed in the context of enterprise linked data: content enrichment and content augmentation, integrated views on distributed data (enterprise mashups), knowledge visualisation, smart assistants and search-driven applications. In all cases, linked data graphs build the basis for such applications, thus the following question is most central for a linked data strategy: How can an enterprise create and maintain knowledge graphs in a sustainable way, whereas the corresponding processes should be as cost-effective as possible and the resulting graphs should be of quality grades which are acceptable for enterprise information services.

1.1 Creating Knowledge Graphs with the Simple Knowledge Organization System (SKOS)

Since the Simple Knowledge Organization System (SKOS) has become a W3C recommendation in 2009 [3], several scenarios to make use of SKOS ontologies have been described [4][5] and many discussions around key design principles such as "minimal ontological commitment" have been led [6]. The increasing use of SKOS can be documented by two key facts:

1. SKOS concept is among over 108,000 classes the most used RDF class in the Linked Open Data cloud[1]
2. When NISO has published ISO 25964 – the international standard for thesauri and interoperability with other vocabularies, one of the main efforts was to reach the goal of interoperability with SKOS and other schemas[2]

The usage of SKOS as a starting point to create knowledge graphs in enterprises has in parallel to its relative simplicity (in contrast to other ontology languages like OWL) one other main advantage: It has been accepted broadly as a standard and is well understood by various stakeholders (database engineers, information professionals, knowledge managers), thus little force is needed to overcome the resistance to the introduction of something new like SKOS based vocabularies.

1.2 SKOS as a nucleus of large enterprise knowledge graphs

The scope of a full-blown enterprise knowledge graph is much broader than a taxonomy would be able to cover. When taking a closer look on it, we will find all kinds of categorized and annotated legacy data and documents, additional schemas which describe various business objects, their specific relations and attributes, and linked data graphs from third-party sources, especially from the linked open data cloud [7]. Additionally, a large knowledge graph will contain a lot of mappings between resources from different (named) graphs. What role can SKOS based graphs play in this complex information system?

When starting with SKOS thesauri to describe all kinds of 'things' (or 'business objects'), their names and relations to each other, we don't have to think about classes or any kind of restrictions or axioms yet. This makes it easy to build a first robust layer of business semantics on top of distributed and heterogeneous information sources. SKOS is based on RDF, thus an extension by additional schemas (classes and properties) is feasible out-of-the-box, at least from a technical point of view. Either custom schemas or already existing ones like FOAF, ORG or schema.org can be used to put additional semantics on top of a SKOS based knowledge graph.

For example, a video game which has been created initially as a SKOS concept with the preferred label 'SimCity' and as a narrower concept of another SKOS concept labeled with 'Video Game' will be classified in a second step by `http://schema.org/SoftwareApplication` which is a subclass of `http://schema.org/CreativeWork`, whereas both are an RDF class derived from schema.org [8]. By using additional schemas, we can express more specific semantics around SKOS graphs and we can map them more accurately with already existing database schemas.

[1] http://stats.lod2.eu/rdf_classes?sort=overall (accessed on September 15, 2014)
[2] http://www.niso.org/schemas/iso25964/#skos (accessed on September 15, 2014)

In addition, when linking SKOS concepts with resources from linked data graphs like Geonames or DBpedia, we can harvest vast amounts of facts around concepts, e.g. birth dates, number of employees, longitude, latitude, etc.

When asking the principal question, why haven't we started the modelling process with a more complex schema from the very beginning, we should consider the following two aspects:

- Domain experts are one of the most valuable resources when creating enterprise knowledge graphs. They most often have no or little expertise with ontology modelling. Thus, they feel more comfortable with a bottom-up approach which starts with concrete instances of classes, and not with a rather abstract schema.
- Although the ontology of SKOS offers only a few ways to express semantics explicitly, the implicit semantics of a SKOS thesaurus is rich enough in many cases to be made available explicitly and machine-readable by applying additional ontologies. One of the most principal design patterns of a linked data architect should be, to convert implicit semantics of existing information sources into explicit semantic models based on standards, not the other way around!

2 Integrating Knowledge Graphs in Enterprise Information Systems

Enterprise information systems benefit from using knowledge graphs in different ways. The following three scenarios shall illustrate some options:

1. Knowledge graphs are browsed by end-users and serve as a knowledge base. In order to make company knowledge better accessible, interfaces should be integrated in popular platforms like Microsoft SharePoint or Atlassian Confluence. As an example for this scenario serves Semantic Knowledge Base for SharePoint[3] which is based on standards-compliant semantic knowledge graphs providing a user interface seamlessly integrated in SharePoint.
2. Knowledge graphs are used to link and to index information from various sources. In many cases it will be used for automatic tagging of enterprise information. A knowledge graph can also be used for concept-based search and to generate more complex queries than usually used by simple full-text search. Examples and online-demos for this approach can be found at the PoolParty Semantic Integrator[4] site.

[3] http://www.semantic-sharepoint.com/?page_id=11 (accessed on September 16, 2014)

[4] http://www.poolparty.biz/portfolio-item/semantic-integrator/ (accessed on September 16, 2014)

3. Knowledge graphs can also be used in enterprises to analyse and to visualise complex contexts and correlations between business objects[5]. As part of a linked data strategy, components like these should be standard-compliant to increase the chance of being reused all over the places which increases usability. In this concrete example mentioned above, only a SPARQL-endpoint [9] is required to deploy the user application.

3 Conclusion

Knowledge graphs play a central role when establishing linked data based enterprise information systems. Application scenarios are manifold, but creation, maintenance and the actual use of it can be a tedious process. In this paper we described some strategies to develop linked data infrastructures which have turned out to be practically applicable. The use of SKOS to get started with knowledge graphs is one of the key elements in an enterprise linked data strategy.

References

1. Frischmuth, P., Auer, S., Tramp, S., Unbehauen, J., Holzweißig, K., & Marquardt, C. M. (2013). Towards Linked Data based Enterprise Information Integration. In WaSABi@ ISWC.
2. Mezaour, A. D., Van Nuffelen, B., & Blaschke, C. (2014). Building Enterprise Ready Applications Using Linked Open Data. In Linked Open Data--Creating Knowledge Out of Interlinked Data (pp. 155-174). Springer International Publishing.
3. Miles, A., & Bechhofer, S. (2009). SKOS simple knowledge organization system reference. W3C recommendation, 18, W3C.
4. Schandl, T., & Blumauer, A. (2010). Poolparty: SKOS thesaurus management utilizing linked data. In The Semantic Web: Research and Applications (pp. 421-425). Springer Berlin Heidelberg.
5. Nagy, H., Pellegrini, T., & Mader, C. (2011). Exploring structural differences in thesauri for SKOS-based applications. In Proceedings of the 7th International Conference on Semantic Systems (pp. 187-190). ACM.
6. Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G., & Summers, E. (2013). Key choices in the design of Simple Knowledge Organization System (SKOS). Web Semantics: Science, Services and Agents on the World Wide Web, 20, 35-49.
7. Igata, N., Nishino, F., Kume, T., & Matsutsuka, T. (2014). Information Integration and Utilization Technology using Linked Data. FUJITSU Sci. Tech. J, 50(1), 3-8.
8. Nogales, A., Sicilia, M. A., García-Barriocanal, E., & Sánchez-Alonso, S. (2013). Exploring the Potential for Mapping Schema.org Microdata and the Web of Linked Data. In Metadata and Semantics Research (pp. 266-276). Springer International Publishing.
9. Prud'Hommeaux, E., & Seaborne, A. (2008). SPARQL query language for RDF. W3C recommendation, 15.

[5] For example, visit http://vocabulary.semantic-web.at/semweb/2184.visual to find a visual representation of 'Linked Data'