Exploring and Exploiting(?) the Awkward Connections Between SKOS and OWL

Stefan Belk, Gerhard Wohlgenannt, Axel Polleres

Vienna Univ. of Economics and Business, Welthandelsplatz 1, 1200 Wien, Austria
{stefan.belk,gerhard.wohlgenannt,axel.polleres}@wu.ac.at
http://www.wu.ac.at

Abstract. In the Semantic Web, the Web Ontology Language (OWL) vocabulary is used for the representation of formal ontologies, while the Simple Knowledge Organisation System (SKOS) is a vocabulary designed for thesauri or concept taxonomies without formal semantics. Despite their different nature, on the Web these two vocabularies are often used together. Here, we try to explore and exploit the joint usage of OWL and SKOS. More precisely, we first define usage patterns to detect problematic modeling from connections between SKOS and OWL. Next, we also investigate if additional information can be inferred from joint usage with SKOS in order to enrich semantic inferences through OWL alone – although SKOS was designed without formal semantics, we argue for this heretic approach by applicability "in the wild": the patterns for modeling errors and inference of new information are transformed to SPARQL queries and applied to real world data from the Billion Triple Challenge 2014; we manually evaluate this corpus and assess the quality of the defined patterns empirically.

Keywords: OWL, SKOS, Data Quality, Linked Data

1 Introduction

While each language, SKOS and OWL, and their fragments serve different use cases in theory, in the Linked Data world, ie., "in the wild", SKOS and OWL are often used together. Two of the questions arising from joint usage are: (i) how can the interplay of SKOS and OWL help to find modeling problems and quality issues within the data, and (ii) can we find recipes to infer additional formal information from joint usage patterns? In order to address these questions we define patterns – in the form of SPARQL queries – to detect cases of modeling problems and potential additional entailments. While the detection of modeling errors is rather straightforward, inferring new information may be seen as heretic, because the design of SKOS is very informal and open on purpose.

We would like to emphasize that, while our attempt to suggest inferences from joint usage from OWL and SKOS may be viewed as contradicting the often stressed informal nature of SKOS, previous works on OWL reasoning for Linked Data have likewise shown that OWL reasoning on Linked Data itself has to be taken "with a grain of salt", i.e. also OWL inferences themselves may often be misleading when applied naively on Web data, cf.[5]. We therefore view our "recipes" to combine SKOS and OWL as a useful starting point to gain additional knowledge out of real, published data.

The **reasons** to visit the posters are: (i) Find out about SKOS-modeling errors appearing in the wild, (ii) learn about potential modeling errors which were observed in joint usage of SKOS and OWL, and (iii) discuss our ideas of inferring additional statements from joint usage of SKOS and OWL.

2 Related Work

Previous work on the implications of combined usage of SKOS and OWL is very limited. Related work can be grouped accordingly: (i) Methods to measure and improve the quality of SKOS (which does not consider joint usage with OWL), eg. Suominen and Mader [6] who identify 26 quality issues that point to potential quality problems in SKOS thesauri, or Cohen [1], who provides formal modeling guidelines for the mapping relations in SKOS. (ii) Methods to analyse the effects on computational properties of using OWL together with SKOS. Jupp et al. [3] which discusses some implications and problems when combining OWL and SKOS. (iii) Works about SKOS schema mapping using OWL statements. Hoekstra [2] presents an approach for mapping SKOS vocabularies using OWL2 semantics. On top of that, work on applying OWL inference to Linked Data "in the wild" may also be viewed as related to our effort.

3 Patterns

Error Patterns: Error patterns detect potentially problematic modeling for (linked) data that uses both OWL and SKOS. The patterns are represented by generic SPARQL queries, which can be applied to any triple store / SPARQL endpoint. The patterns are kept simple, and combine SKOS vocabulary with OWL primitives. We applied a number of error patterns, of which ten patterns occurred in the data. Due to space restrictions we only list one example below, at our web site[1] we present all patterns for error checking and inference.

EP4: `skos:narrowerTransitive` **and** `rdfs:subClassOf`

```
SELECT DISTINCT ?sub ?super WHERE {
?sub rdfs:subClassOf ?super.
?sub skos:narrowerTransitive ?super.
}
```

EP4 states that at the same time there is a *rdfs:subClassOf* and a *skos:narrowerTransitive* between two resources. The SKOS Reference [4] defines that ?*a* *skos:narrowerTransitive* ?*b* asserts that ?*b*, the object of the triple, is a narrower "descendant" of ?*a*, i.e. ?*a* is a broader concept than ?*b*. This hints at an incorrect understanding of the direction of *skos:narrower*.

[1] `http://owl_skos_lod.ai.wu.ac.at`

Inference Patterns: Inferring new formal information from joint usage of OWL and SKOS is the most daring part of our research, because SKOS is deliberately informal by design. The SPARQL *CONSTRUCT* queries presented in the example below (and many other queries on our web site) suggest new triples to be added to the data set.

IP1: owl:equivalentClass when skos:closeMatch and superClass

```
CONSTRUCT {?a owl:equivalentClass ?b.} WHERE {
    ?a skos:closeMatch ?b.
    ?a rdfs:subClassOf ?c.
    ?b rdfs:subClassOf ?c.
}
```

The motivation behind this pattern is that two *owl:Class* classes having the exact same super-class and being linked via *skos:closeMatch* are likely to be equivalent classes. This pattern adds a *owl:equivalentClass* relation between those classes to the triple store. We know that the added relation will not always be correct, all inference patterns are manually checked in the evaluation.

4 Evaluation

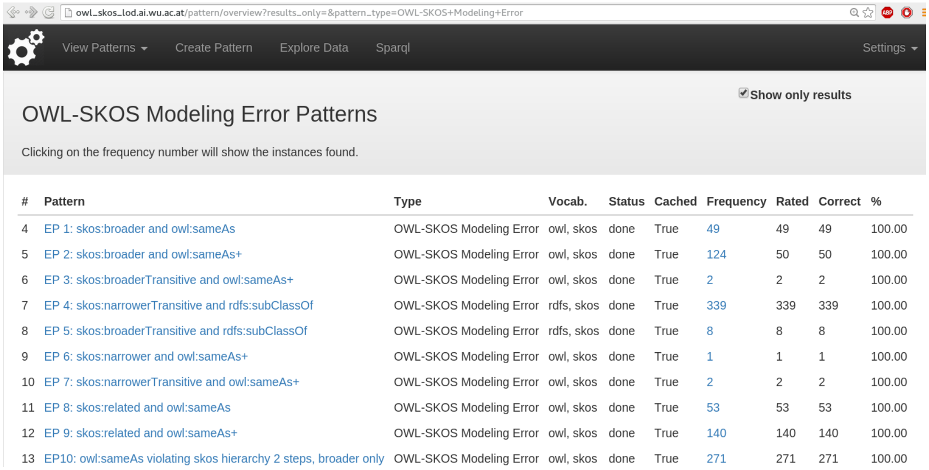
To evaluate the patterns, we aimed to use a real world data set published by a variety of different sources on the Internet. We decided to apply the patterns to the data from the Billion Triple Challenge 2014[2].

First, we applied the patterns for detecting modeling problems to the evaluation dataset. Some of the ten patterns (EP1–EP10) occurred more frequently in the dataset, others only a few times. We manually examined the data, all occurrences indeed pointed to modeling problems.

Figure 1 present a screenshot of the Web Frontend, which lets users not only see the SPARQL queries and evaluation results for all patterns, but also access the underlying data matching the patterns.

Secondly, the inference patterns were applied to the evaluation dataset. In summary, all of the presented patterns in the Web Frontend correctly inferred new relations. However, obviously the quality and the amount of the these relations depends heavily on the data set that the patterns are applied to. Some patterns work well with a particular data set while others do not. In the end, our goal was to find out which connections between SKOS and OWL are used in practice and divide them into ones which arguably look like modeling errors and others which could suggest additional inferences: in cases where we could not detect a clear trend, it should be decided on a per-dataset basis which of these 'recipes' to apply or not, which probably remains a manual task or, resp., in the control of the dataset owner.

[2] `http://km.aifb.kit.edu/projects/btc-2014`



#	Pattern	Type	Vocab.	Status	Cached	Frequency	Rated	Correct	%
4	EP 1: skos:broader and owl:sameAs	OWL-SKOS Modeling Error	owl, skos	done	True	49	49	49	100.00
5	EP 2: skos:broader and owl:sameAs+	OWL-SKOS Modeling Error	owl, skos	done	True	124	50	50	100.00
6	EP 3: skos:broaderTransitive and owl:sameAs+	OWL-SKOS Modeling Error	owl, skos	done	True	2	2	2	100.00
7	EP 4: skos:narrowerTransitive and rdfs:subClassOf	OWL-SKOS Modeling Error	rdfs, skos	done	True	339	339	339	100.00
8	EP 5: skos:broaderTransitive and rdfs:subClassOf	OWL-SKOS Modeling Error	rdfs, skos	done	True	8	8	8	100.00
9	EP 6: skos:narrower and owl:sameAs+	OWL-SKOS Modeling Error	owl, skos	done	True	1	1	1	100.00
10	EP 7: skos:narrowerTransitive and owl:sameAs+	OWL-SKOS Modeling Error	owl, skos	done	True	2	2	2	100.00
11	EP 8: skos:related and owl:sameAs	OWL-SKOS Modeling Error	owl, skos	done	True	53	53	53	100.00
12	EP 9: skos:related and owl:sameAs+	OWL-SKOS Modeling Error	owl, skos	done	True	140	140	140	100.00
13	EP10: owl:sameAs violating skos hierarchy 2 steps, broader only	OWL-SKOS Modeling Error	owl, skos	done	True	271	271	271	100.00

Fig. 1. A screenshot of a part of the Web frontend – listing the patterns for detecting OWL-SKOS modeling errors.

5 Conclusions

Our contributions are as follows: (i) Presenting a number of usage patterns for OWL and SKOS usage which help to identify modeling errors in data sets, (ii) provide patterns to infer additional information, and (iii) evaluate the patterns defined in (i) and (ii) with regards to the Billion Triple Challenge data set to determine their quality. There are many directions for future work, eg. increasing the number and complexity of patterns, and applying them to different datasets.

Acknowledgments. This work has been supported by project uComp, which receives the funding support of EPSRC EP/K017896/1, FWF 1097-N23, and ANR-12-CHRI-0003-03, in the CHIST-ERA ERA-NET programme.

References

1. Cohen, M.: Semantics for mapping relations in skos. In: Faber, W., Lembo, D. (eds.) Web Reasoning and Rule Systems, LNCS, vol. 7994, pp. 223–228. Springer (2013)
2. Rinke Hoekstra (Dept. of CS, V.U.A.: Bestmap: Context-aware skos vocabulary mappings in owl 2. Proceedings of OWL: Experiences and Directions 2009 (2009)
3. Jupp, S., Bechhofer, S., Stevens, R.: Skos with owl: Don't be full-ish! In: Dolbear, C., Ruttenberg, A., Sattler, U. (eds.) OWLED. CEUR WS Proc., vol. 432 (2008)
4. Miles, A., Bechhofer, S.: Simple knowledge organization system reference. Recommendation, W3C (August 18 2009)
5. Polleres, A., Hogan, A., Delbru, R., Umbrich, J.: RDFS & OWL reasoning for linked data. In: Rudolph, S., Gottlob, G., Horrocks, I., van Harmelen, F. (eds.) Reasoning Web 2013, LNCS, vol. 8067, pp. 91–149. Springer, Mannheim, Germany (Jul 2013)
6. Suominen, O., Mader, C.: Assessing and improving the quality of skos vocabularies. Journal on Data Semantics 3(1), 47–73 (2014)