

The current state of SKOS vocabularies on the Web

Nor Azlinayati Abdul Manaf^{1,2}, Sean Bechhofer¹, and Robert Stevens¹

¹School of Computer Science, The University of Manchester, United Kingdom

²MIMOS Berhad, Technology Park Malaysia, 57000 Kuala Lumpur, Malaysia
{abdulman, seanb, robert.stevens}@cs.man.ac.uk

Abstract. We survey the current state of SKOS vocabularies on the Web. We identified 478 SKOS vocabularies, which were gathered through collections and web crawling. Analyses were then conducted that included investigation of the use of SKOS constructs; the use of SKOS semantic relations and lexical labels; and the structure of vocabularies in terms of the hierarchical and associative relations, branching factors and the depth of the vocabularies. Almost one-third of the SKOS vocabularies collected fall into the term lists category, with no use of any SKOS semantic relations. As concept labelling is core to SKOS vocabularies, a surprising find is that not all SKOS vocabularies use SKOS lexical labels, whether `skos:prefLabel` or `skos:altLabel`, for their concepts. The survey results can serve to provide a better understanding of the modelling styles of the SKOS vocabularies published on the Web, especially when considering the creation of applications that utilize these vocabularies.

1 Introduction

We have surveyed the Simple Knowledge Organization System (SKOS) vocabularies on the Web. The aim of this survey is to understand what a typical SKOS vocabulary looks like, especially in terms of the shape, size and depth of the vocabulary structure. We wish to find how many and what types of SKOS vocabularies are publicly available on the Web. We are interested in understanding which of the SKOS constructs listed in the SKOS Reference document [1] are actually being used in the SKOS vocabularies and how often these constructs are used. Additionally, we are interested in learning what the SKOS vocabularies look like in terms of size, shape and depth of the vocabulary structure.

The result of this survey will equip us with a better understanding of the modelling styles used in the SKOS vocabularies. This may be important when considering creating an application that utilizes these vocabularies, for example, when converting artefacts such as OWL ontologies into SKOS, where some typicality of SKOS vocabularies may be useful to guide the conversion.

For the purposes of this survey, we used the following definition of a SKOS vocabulary.

Definition 1. *A SKOS vocabulary is a vocabulary that at the very least contains SKOS concept(s) used directly, or SKOS constructs that indirectly infer the use of a SKOS concept, such as use of SKOS semantic relations.*

2 Analysis of SKOS vocabularies

We briefly summarize the main conclusions of our survey; a full report is in [2].

SKOS vocabulary discovery on the Web. In this survey, we collected 478 vocabularies that according to our definition are SKOS vocabularies. Even three years since becoming a W3C Recommendation, the use of SKOS appears to remain low. However, our total of SKOS vocabularies may be artificially low, with some being hidden from our collection method. The reasons for some of these vocabularies not being publicly accessible by an automated process could be due to factors such as proprietary issue, vocabularies stored within SVN, etc.

SKOS construct usage. Of all the SKOS constructs that are available in the SKOS standards, only 3 out of 35 SKOS constructs are used in more than 60% of the vocabularies. These constructs are `skos:Concept`, `skos:prefLabel` and `skos:broader`. The SKOS lexical labelling properties `skos:prefLabel` or `skos:altLabel` are not used in 100% of the vocabularies. Further analysis revealed that some vocabularies use `rdfs:label` for labelling their concepts.

SKOS classification. We found out that all but two of the SKOS vocabularies that we collected from the Web fell into one of the categories listed by the traditional KOS; flat vocabularies (*Glossary*), multi-level vocabularies (*Taxonomy*) and relational vocabularies (*Thesaurus*). 61% or 293 of the vocabularies are categorised as *Taxonomy*. The second largest type is *Glossary* with 27% or 129 vocabularies. 11% or 54 vocabularies fell into the *Thesaurus* category.

Typicality of a SKOS vocabulary. Based on the results of this survey, a typical taxonomy looks like a polyhierarchy that is 2 levels deep, with a *hierarchical forward branching factor*, $\mathcal{FBF}_{\mathcal{H}}$ of 10 concepts and a *hierarchical backward branching factor*, $\mathcal{BBF}_{\mathcal{H}}$ of 3 concepts. A typical thesaurus also looks like a polyhierarchy that is 6 level deep, with a *hierarchical forward branching factor*, $\mathcal{FBF}_{\mathcal{H}}$ of 3 concepts and a *hierarchical backward branching factor*, $\mathcal{BBF}_{\mathcal{H}}$ of 2 concepts, additionally having associative relationships, *associative forward branching factor*, $\mathcal{FBF}_{\mathcal{A}}$ of 1 concept.

The result of this survey provides a better understanding of the discovery, usages of SKOS constructs, types and modelling styles of SKOS vocabularies published on the Web. This understanding is important, especially when considering the creation of applications that utilize these vocabularies.

References

1. Miles, A., Bechhofer, S.: SKOS simple knowledge organization system reference. W3C recommendation, W3C (August 2009)
2. Abdul-Manaf, N.A., Bechhofer, S., Stevens, R.: The current state of SKOS vocabularies on the Web. In: Proceedings of the 9th Extended Semantic Web Conference (ESWC2012). (May 2012)