Assessing and Improving SKOS Mapping Quality

Hong Sun, Jos De Roo, Giovanni Mels, Kristof Depraetere,

Marc Twagirumukiza, Boris De Vloed, Dirk Colaert

Advanced Clinical Applications Research Group, Agfa HealthCare, Gent, Belgium
{hong.sun, jos.deroo, giovanni.mels, kristof.depraetere,
marc.twagirumukiza, boris.devloed, dirk.colaert}@agfa.com

Abstract. The Simple Knowledge Organization System (SKOS) is popular for expressing controlled vocabularies for their use in Semantic Web applications. Using SKOS, concepts can be linked to other concepts and organized into hierarchies inside a single terminology system. Meanwhile, expressing mappings between concepts in different terminology systems is also possible with SKOS mapping properties. This poster discusses potential quality issues in using SKOS to express terminology mappings. Problematic patterns are defined and corresponding rules are developed to automatically detect situations where the mappings either result in 'SKOS Vocabulary Hijacking' or cause conflicts. The validation rules, expressed in N3 format, are available as open source.

Keywords: SKOS, Terminology Mapping, Clinical Terminology, N3 rules.

1 Introduction

The Simple Knowledge Organization System (SKOS) [1] provides a data model and vocabulary for expressing controlled vocabularies. Many organizations have published their controlled vocabularies using SKOS for their use in Semantic Web applications [2]. In order to further facilitate the knowledge transfer where data is represented with different terminologies, terminology mapping is urgently required. In the clinical domain where many clinical terminology systems are in use, mappings between these different terminologies become a prerequisite for clinical data sharing.

Mappings between different terminology systems do exist in many domains. For example, in the clinical domain, the International Health Terminology Standards Development Organisation, together with the World Health Organization, developed the SNOMED CT to ICD-10 mapping [3] , where the mappings are expressed in an Excel sheet. However, most of the existing mappings are expressed in a non-semantic format, which prevents their direct use in semantic web applications.

The SKOS specification defines five mapping properties which can be used to specify mappings in different situations (e.g. broader, narrow, etc.). However, we discovered that SKOS mappings may assert unintended semantic relations to source vocabularies. Such injected relations are not developed by the owners of source vocabularies and therefore are considered as 'SKOS Vocabulary Hijacking' [2]. Furthermore, as

such assertions are imperceptible, the mapping creators might not be aware of such side effects brought by their mappings, which may lead to serious consequences, especially in the clinical domain. Besides the consequence of resulting in vocabulary hijacking, the mappings may also cause conflicts with existing relations.

This poster first analyzes the cause of the above mentioned issues, and then introduces rules to detect problematic patterns. A set of validation rules that detect problematic patterns is developed in N3[1] format and published as open source [4].

2 SKOS Vocabulary Hijacking Caused by SKOS Mapping

In [2], the concept 'SKOS vocabulary hijacking – the assertion of facts about vocabularies published by others' is coined. The SKOS authors permit such assertion and considered it as 'asserting semantic links within someone else's concept scheme'[2]. However, we deem the unintentional assertions brought by SKOS mapping as a different scenario which may bring serious consequence. The mapping creators should be notified when their mappings are resulting in such inferred relation assertions.

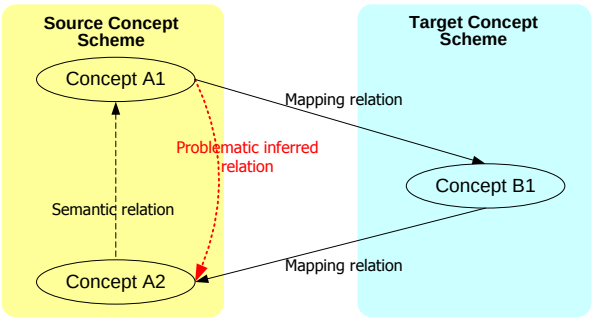

Fig. 1. Basic vocabulary hijacking pattern

Fig. 1 shows the basic pattern of SKOS vocabulary hijacking caused by SKOS mappings. As most of the SKOS mapping relations can be considered as transitive and bidirectional after certain inference, it is possible to infer a relation (the red dotted line in Fig. 1) between two concepts of the same concept scheme. If such an inferred relation is not stated in (or cannot be inferred from) that concept scheme, then inferring such a relation via SKOS mapping relations is considered as vocabulary hijacking.

In addition, if such an inferred relation is contradictory to any existing relation, e.g. the semantic relations displayed as black dashed line in Fig. 1, it would be considered as a conflict, and the related mappings need to be corrected.

Furthermore, problematic relations can also be inferred from the combination of semantic relations and mapping relations. For example, in Fig. 1, it is possible to infer a relation between A2 and B1 based on the semantic relation between A2 and A1 and the mapping relation between A1 and B1. If this inferred relation is contradictory to the mapping relation between A2 and B1, it would be considered as a conflict as well.

---

[1] Notation 3: http://www.w3.org/TeamSubmission/n3/

[2] http://www.w3.org/2006/07/SWD/wiki/WashingtonAgenda/MappingIssues

3 Detecting Problematic Patterns in SKOS Mapping

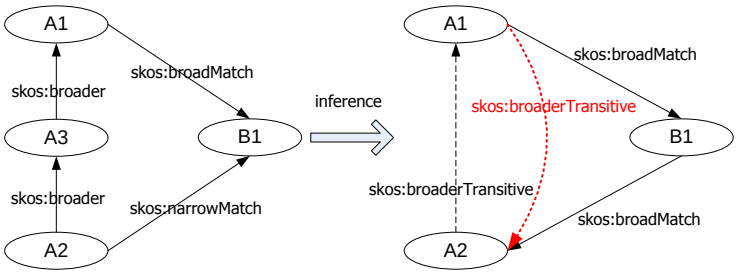

Fig. 2. An example of inferring a problematic pattern

This chapter shows an example of inferring a problematic pattern caused by SKOS mapping. A set of other problematic patterns are provided in [4].

In Fig. 2, the left part of the figure shows the original SKOS concepts and relations where upon the validation rules are to be applied. The listed mappings appear to be unidirectional and there is no obvious conflict.

A minimal set of SKOS inference rules [4] are developed in order to simplify the semantic and mapping relations. After the inference, the pattern displayed on the left side of Fig. 2 is translated into the pattern displayed on the right side, except for the *skos:broaderTransitive* relation (red dotted line). The skos:broader relations stated in the left side patterns are simplified into the *skos:broaderTransitive* relation between A1 and A2 (black dashed line). It can also be observed that the unidirectional mapping relations represented in the left pattern turns to bidirectional.

Furthermore, the *skos:broaderTransitive* relation (red dotted line) can be inferred from the *skos:broadMatch* relations. If the inferred relation is not stated in (or cannot be inferred from) concept scheme A, then it is considered as vocabulary hijacking. In addition, in Fig. 2, such an inferred relation also results in a non-consistency with the *skos:broaderTransitive* relation represented in black dashed line.

Listing 1. Detecting Pattern1VocabularyHijacking

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@prefix validation: <http://eulersharp.sourceforge.net/2003/03swap/skos-mapping-validation-rules#>.
@prefix e: <http://eulersharp.sourceforge.net/2003/03swap/log-rules#>.

1 {
2       ?A1 skos:broadMatch ?B1.
3       ?B1 skos:broadMatch ?A2.
4       ?SCOPE e:findall ( ?A1 { ?A1 skos:broaderTransitive ?A2. } () ).
5       {       ?A1 skos:broadMatch ?B1.
6               ?B1 skos:broadMatch ?A2. } e:graphCopy ?pattern.
7 } => {
8       ?pattern a validation:Pattern1VocabularyHijacking.
9 }.
```

The rule displayed in Listing 1 detects a problematic vocabulary hijacking pattern and classifies this pattern in the output of the rule, so that the detected problematic patterns can be checked. The rule is expressed in N3 format and executed by Euler YAP Engine (EYE) [5], an open source reasoning engine. Line 2-4 detects the problematic pattern of vocabulary hijacking in Fig. 2. Line 4 states the fact that the triple

'?A1 skos:broaderTransitive ?A2' does not exist in all the input graphs. Line 5 and 6 copies the detected problematic mappings to a graph (?pattern), which is passed to result in Line 8. The detected pattern in Listing 1 is considered an instance of the validation:Pattern1VocabularyHijacking class. A set of classes are defined to reflect different problematic patterns [4]. Rules that detect those problematic patterns are also presented in [4]. Detailed explanations to those patterns and rules are provided in a separate document [6].

4 Conclusions

This poster analyzes quality issues in SKOS mapping, where unintended assertions may occur and conflicts may exist. A problematic pattern which causes such issues is discussed as an example. As such problematic patterns are only visible after inferences, the mapping creators may not be aware of them, which makes it hard to detect those problems. In addition, established mappings still require updates when related terminologies evolve [7]. It is therefore important that the quality of the mappings can be assessed continuously.

We have developed a set of SKOS mapping validation rules [4] [6] to automatically detect problematic patterns in terminology mappings that are expressed with SKOS mapping properties. Mappings that cause problematic patterns are reported and classified as different classes, so that a mapping creator could validate these problematic patterns, and eventually improve the quality of their mappings. The validation rules have been tested in assessing SKOS mappings between medical terminologies. Detailed analysis of the validation results of public terminology mappings in the medical domain will be presented in the future.

Acknowledgement: This work was supported by funding from the SALUS project (http://www.salusproject.eu/). Grant agreement N° 287800.

5 References

1. Miles A, Bechhofer S, SKOS simple knowledge organization system reference, W3C Recommendation, World Wide Web Consortium, 18 August 2009.
2. Suominen O, Mader C. Assessing and Improving the Quality of SKOS Vocabularies. Journal on Data Semantics, 2013, 2(2).
3. J. Campbell, et al. Mapping SNOMED CT to ICD-10, Technical Specifications. http://www.who.int/classifications/icd/snomedCTToICD10Maps
4. SKOS Mapping Validation Rules. http://eulersharp.sourceforge.net/2003/03swap/skos-mapping-validation-rules Last visited on October 2013.
5. Euler Yap Engine, http://eulersharp.sourceforge.net/ Last visited on October 2013.
6. Sun H, De Roo J, Twagirumukiza M, Mels G, Depraetere K, De Vloed B, Colaert D. Validation rules for assessing and improving SKOS mapping quality. arXiv:1310.4156 (2013).
7. Saitwal H, Qing D, Jones S, et al. Cross-terminology mapping challenges: A demonstration using medication terminological systems. Journal of Biomedical Informatics, 2012.