Represent Changes of Knowledge Organization Systems on the Semantic Web

Qing Zou

Abstract:

Traditional knowledge organization systems (KOS) including thesauri, classification schemes, taxonomies, subject heading systems, name authorities, and other lists of terms and codes have been playing important roles in indexing, information organization, and retrieval. With the advent of the semantic web, a large number of them have been converted into Linked Open Data (LOD) datasets. Since the Simple Knowledge Organization Systems (SKOS) and SKOS eXtension for Labels (SKOS-XL) became formal World Wide Web Consortium (W3C) recommendations, they have been applied to knowledge organization systems. In this article, the issues surrounding changes, versioning control, and evolution of KOS are investigated. From KOS services providers and consumers perspectives, this study focuses on the representation of changes on the semantic web.

To submit your article to this journal:

Go to http://ojs.calaijol.org/index.php/ijol/about/submissions

Represent Changes of Knowledge Organization Systems on the Semantic Web

Qing Zou

Lakehead University, Canada

ABSTRACT

Traditional knowledge organization systems (KOS) including thesauri, classification schemes, taxonomies, subject heading systems, name authorities, and other lists of terms and codes have been playing important roles in indexing, information organization, and retrieval. With the advent of the semantic web, a large number of them have been converted into Linked Open Data (LOD) datasets. Since the Simple Knowledge Organization Systems (SKOS) and SKOS eXtension for Labels (SKOS-XL) became formal World Wide Web Consortium (W3C) recommendations, they have been applied to knowledge organization systems. In this article, the issues surrounding changes, versioning control, and evolution of KOS are investigated. From KOS services providers and consumers perspectives, this study focuses on the representation of changes on the semantic web.

Keywords: KOS, Change Representation, Linked Data, SKOS, SKOS-XL

INTRODUCTION

The Simple Knowledge Organization System (SKOS) provides a common data model for organizing knowledge organization systems such as thesauri, classification schemes, subject headings, and taxonomies (W3C, 2012). Since the Simple Knowledge Organization Systems (SKOS) and SKOS eXtension for Labels (SKOS-XL) became formal World Wide Web Consortium (W3C) recommendations, they have been applied to knowledge organization systems. SKOS and SKOS-XL capture the common relations of KOS in a formal and explicit way. Moreover, it potentially improves the interoperability between KOS. SKOS and SKOS-XL can not only bring paper format KOSs but electronic format KOSs into an open and highly connected

linked data world.

However, an actively used KOS needs to be regularly updated to reflect the development of human knowledge. Therefore, changes are inevitable for KOS. Changes need to be properly managed by KOS services providers and KOS consumers because an out-of-date or poorly managed KOS may provide obsolete information that may hinder information retrieval. KOS services providers often actively maintain their KOSs. For example, a thesaurus such as Arts and Architecture Thesaurus (AAT)[1], subject heading systems such as Library of Congress Subject Headings (LCSH)[2], Canadian Subject Headings (CSH)[3] and RVM (Répertoire de vedettes-matière)[4] publish revisions regularly. However, KOS consumers have difficulty to keep up with KOS services providers. For instance, researchers have identified obsolete subject headings in library automation systems (Buckland, 2012; Chan & Vizine-Goetz, 1997). To actively manage KOS changes at the consumers' side, KOS changes need to be explicitly expressed and properly propagated. With more and more KOSs being published on the semantic web, the issue of KOS changes needs to be investigated in this new context.

LITERATURE REVIEW

This section is organized into two subsections. Related studies about KOS changes are reviewed in the first subsection. Studies related to change representation are summarized in the second subsection.

Changes in KOS

The obsolete problem caused by KOS changes especially in subject heading systems has long been identified (Ashman, 2006; Buckland, 2012; Chan & Vizine-Goetz, 1997). Changes in KOS are caused by new topics, names, concepts, and cultural changes (Buckland, 2012). If changes in KOS are not propagated to applications, the old terms, concepts, and names become obsolete.

Since SKOS has been widely applied in publishing KOS on the semantic web, the capabilities of handling KOS changes in SKOS have been examined (Tennis, 2005). The two mechanisms (i.e., notes and web ontology language (OWL) versioning) for concept schema revision provided by SKOS do not "account for … refinement, lumping and other transformations of concepts" (Tennis, 2005, p. 276). Although Tennis (2005) proposes an approach to extend SKOS Core to track vocabularies changes over their lifetime through capturing three different changes including lumping, refining, and relationship changes, there are no detailed explanations on how the

---

[1] http://www.getty.edu/research/tools/vocabularies/aat/

[2] http://id.loc.gov/authorities/subjects.html

[3] https://www.bac-lac.gc.ca/eng/services/canadian-subject-headings/Pages/canadian-subject-headings.aspx

[4] https://rvmweb.bibl.ulaval.ca/

approach works. For example, regarding relationship changes, it is not clear that the proposed "skos:wasRelated" will stick to the concept along its lifetime or not. If one concept has associative relations with two concepts and two association relations are changed to hierarchical relations more than one time, there will be more than two "skos:wasRelated" of one concept in the KOS. There is no way to tell the temporal differences between the two "skos:wasRelated". In addition, it is not clear what the differences between the proposed "skos:ConceptLump" and "skos:Concept" and between the proposed "skos:ConceptRefinement" and "skos:Concept". From the perspectives of services providers and consumers, the approach may not be applicable.

Tennis (2007) categorized characteristics of scheme changes into three major changes including "structural change, word-use change and textual change" (p.90). Traditionally, term records have been used to manage values of thesaurus entries by thesaurus management manuals and standards (Aitchison, Bawden, & Gilchrist, 2000; National Information Standards Organization, 2005; Soergel, 1974). Tennis (2007) proposed values records, an expanded version of term records, for managing changes. However, although the approach was claimed for scheme versioning in the semantic web, it has not been empirically tested.

Tennis and Sutton (2008) differentiate the abstract concept and "concept instances" to capture concept evolution in vocabulary development applications. However, this approach is bound to vocabulary development applications as shown in the iterative workflow described in the article. No doubt, there are some advantages to link concept instances to their abstract concepts. This approach also brings in extra work on maintaining abstract concepts. It is not clear how this approach can couple with situations when any changes happen to abstract concepts including addition, modification, and deletion.

Change Representation

Change representation is one of the six phases (i.e., change capturing, change representation, semantics of change, change implementation, change propagation, and change validation) of ontology evolution in a cyclic loop (Stojanovic, Maedche, Motik, & Stojanovic, 2002). Ontology can be defined as an "explicit specification of a conceptualization" (Gruber, 1993, p. 199). In a broader sense, traditional knowledge organization systems include terms list, subject headings, classification schemes, taxonomies, and thesauri are ontologies. Uschold and Gruninger (2004) categorized taxonomies and thesauri as "lightweight-ontologies". Since ontology change refers to "the generic problem of changing an ontology in response to a certain need" (Flouris, Manakanatas, Kondylakis, Plexousakis, & Antoniou, 2008, p. 117), studies focusing on ontology evolution, versioning and change management are relevant to KOS changes.

Ontology Changes have been categorized into 1) atomic (basic) change; and 2) composite changes (Javed, Abgaz, & Pahl, 2013). Noy, Chugh, Liu, and Musen (2006) introduced a Change and Annotation Ontology (CHAO) to explicitly express the changes between different versions of ontologies. A change history ontology is used to represent changes (Khattak, Batool, Pervez, Khan, & Lee, 2013). A layered change history log has been introduced to log ontology changes (Javed,

2013).

Taking into consideration of syntactic and semantics of ontologies, Zeginis, Tzitzikas, and Christophides (2011) proposed "delta functions" to compare RDFs. A temporal logic approach is used to manage and reason in different versions of an ontology (Huang & Stuckenschmidt, 2005). Palma, Corcho, & Haase (2009) proposed OWL 2 change ontology for managing changes following a temporal ordering.

Types of KOS changes have been identified. Changes happened in a thesaurus fall into six categories: "amendment of existing terms, status of existing terms, deletion or demotion of existing terms, addition of new, or deletion of old relationships, addition of new terms, amendment of existing structure" (Aitchison et al., 2000, p. 170). Changes can be categorized into three general groups including "structural change, word-use change and textual change" (Tennis, 2007, p. 90). With more and more KOS published on the semantic web, KOS changes need to be discussed in this new context.

In summary, change representation in ontology needs to take features of ontologies into consideration. Since KOSs and ontologies are different, some methods (such as reasoning) used in ontologies cannot apply to KOSs. In this study, methods such as change log and temporal order are utilized in the investigation of KOS changes.

OBJECTIVE OF THE STUDY

This study aims to explore how to represent KOS changes on the semantic web.

METHODOLOGY

In order to address the issue, this study takes a three steps approach: 1) formally define changes in KOS, 2) to model change sets; 3) to apply the change sets model to use cases which are taken from Tennis's Studies.

Changes Representation

In the context of the semantic web, Resource Description Framework (RDF) is the fundamental tool and model for representing resources (Manola & Miller, 2004). Knowledge organization systems can be encoded in RDF statements using SKOS and SKOS-XL. KOS changes can be defined as any modification of a KOS.

Definition 1. A knowledge organization system is a set of concepts interrelated by relationships. A KOS K can be denoted as:
$K = (C, A, R)$
Where C is a set of concepts, A is a set of attributes, and R is a set of relations. For each concept

$c_i \; C,\; A(c_i) = \{a_1, a_2, \ldots, a_m\}$, and $R(c_i) = \{r_1, r_2, \ldots, r_n\}$.

In other words, a KOS can be considered as a set of RDF triples of form <subject, predicate, object> with semantics defined in RDF, SKOS and SKOS-XL, if the KOS only uses SKOS and SKOS-XL, RDF.

Definition 2. Operator + is an addition operation and +c is an operation to add a concept, +a(c) is an operation to add an attribute to concept c, and +r(c) is an operation to add a relation to concept c. More specifically, the three operations are denoted as follows:

$K^{old} = (C^{old}, A^{old}, R^{old})$ and $\qquad K^{new} = (C^{new}, A^{new}, R^{new})$

$K^{old} + c = K^{new},\; c \in C^{new} \; C^{old})$

$K^{old} + a(c) = K^{new},\; c \in C^{old} c \in C^{new} a(c) \in A^{new} \; A^{old})$

$K^{old} + r(c) = K^{new},\; c \in C^{old} c \in C^{new} r(c) \in R^{new} \; R^{old})$

Definition 3. Operator - is a deletion operation and -c is an operation to delete a concept, -a(c) is an operation to delete an attribute of concept c, and -r(c) is an operation to delete a relation of concept c. More specifically, the three operations are denoted as follows:

$K^{old} = (C^{old}, A^{old}, R^{old})$ and $K^{new} = (C^{new}, A^{new}, R^{new})$

$K^{old} - c = K^{new},\; c \in C^{old} \; C^{new})$

$K^{old} - a(c) = K^{new},\; c \in C^{old} c \in C^{new} a(c) \in A^{old} \; A^{new})$

$K^{old} - r(c) = K^{new},\; c \in C^{old} c \in C^{new} r(c) \in R^{old} \; R^{new})$

Definition 4. Given two knowledge organization systems $K^{old} = (C^{old}, A^{old}, R^{old})$ and $K^{new} = (C^{new}, A^{new}, R^{new})$, the changes are $\Delta(K^{old} -> K^{new}) = \Delta C$ and

$\Delta C = C^{new} - C^{old} = \{ C^{new} (c C^{old})\}$,

$\Delta A = C^{new} | A(c) A^{new} A^{old}) \}$,

$\Delta R = C^{new} | R(c) R^{new} R^{old}) \}$

Two basic operations add and delete are defined above. In terms of changes in KOS, Tennis (2007) identified structural change, word-use change and textual change. More specifically, KOS changes include concept splitting and lumping (Tennis, 2005). Take the example used in Tennis (2005), the changes can be represented using the definitions:

Table 1. Change example 1

DC2002 Terms (a) Applications (b) Web services	skos:Concept "Applications" skos:Concept "Web services"
DC2003 Metadata Thesaurus (a) Applications NT Web services	skos:Concept "Applications" skos:narrower "Web services" skos:Concept "Web services"

The change in Table 1 can be considered as one relation addition operation. The concept "Applications" has a "skos:narrower" relation with the concept "Web services" after the change:

$\Delta = \Delta R = R(Concept(\text{"Applications"})) = \text{"skos:narrower"}$ and
$\text{"skos:narrower}(Concept(\text{"Application"})) \rightarrow Concept(\text{"Web services"})$

The change can be represented by Turtle (a terse RDF triple language (Beckett, Berners-Lee, Prud'hommeaux, & Carothers, 2014)) as follows:

changelog:operation01 a ch:RelationAdd ;
ch:hasTimeStamp "2018-03-01 15:34:45+5" ;
    ch:subject skos:Concept "Application" ;
    ch:relation "skos:narrower";
    ch:target skos:Concept "Web services" .

Changes are a set of operations. Therefore, a general model was built on Changeset (Tunnicliffe & Davis, 2009), change log history (Khattak, Latif, & Lee, 2013), and layered change log (Javed et al., 2013).

Figure 1. A KOS change set model
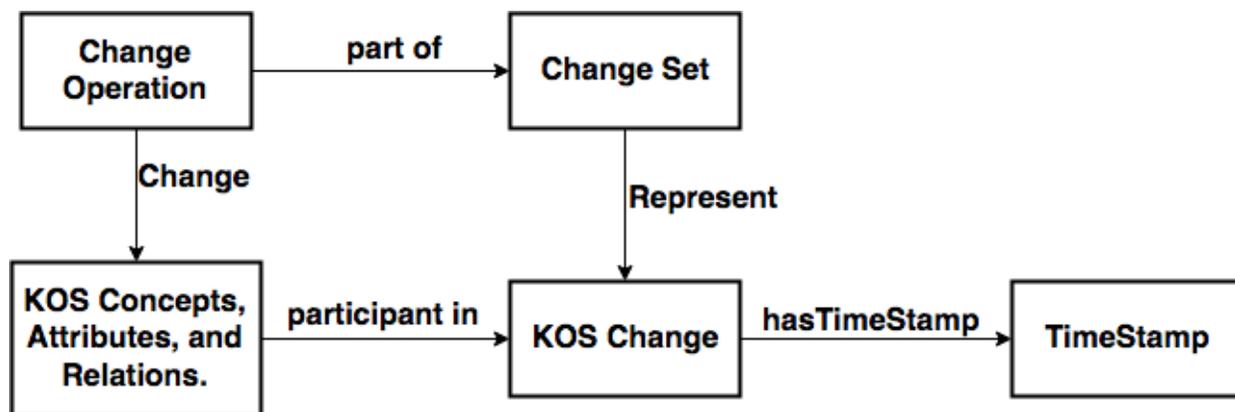
In this model, a change set contains a set of change operations. A change set represents a series of KOS changes. KOS concepts, attributes, and relations participant in KOS changes and are changed by change operations.

To express KOS changes, an ontology[5] was developed based on the model. At the core of this ontology, changes are formally defined. Figure 2 shows that the hierarchical structure of the AtomicChange class.

---

[5]   http://nkos.info/ontology/cs.owl

Figure 2. The AtomicChange class and subclasses
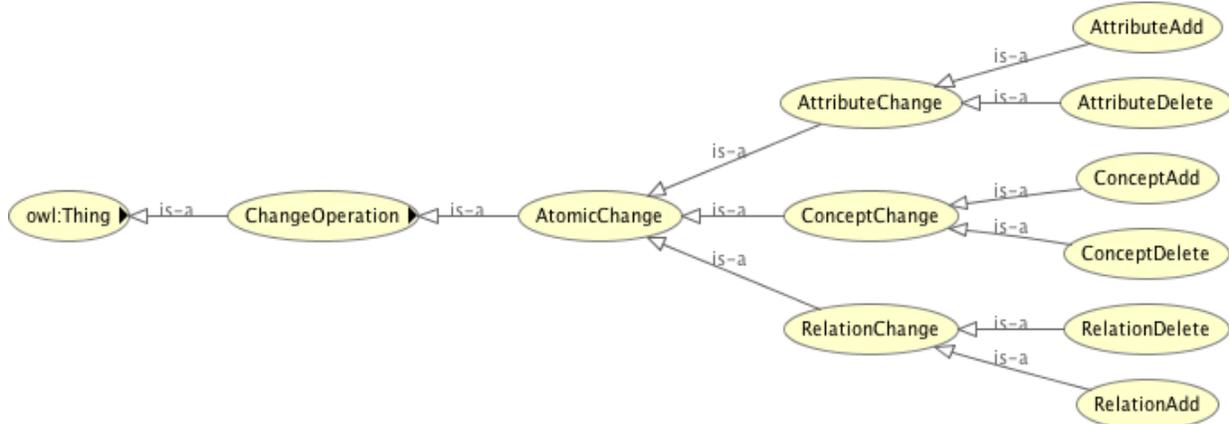
The example 1 is complete with the following statements that link the operation to a change set using the developed ontology. Suppose that "ch" is the prefix for the ontology. The following is in a Turtle format:

changelog: ChangeSet01 a ch:ChangeSet
    ch:hasChangeBeginTime "2017-12-01 15:34:45+5";
    ch:hasChangeEndTime "2017-12-01 15:35:00+5".
    ch:operator   changelog:agent01 ;
    ch:hasOperation changelog:operation01.
changelog:operation01 a ch:RelationAdd ;
    ch:hasTimeStamp "2018-03-01 15:34:45+5" ;
        ch:subject skos:Concept "Application" ;
        ch:relation "skos:narrower";
        ch:target skos:Concept "Web services" .

The statements indicate when the change starts and ends, who is the operator, and to which operation it links. In this example, the operation (changelog:operation01) is a RelationAdd operation that includes subject, relation, and target.

The second example is listed in Table 2 (Tennis, 2005). In the example, two concepts lumped together.

Table 2. Change example 2

DC2002 Terms (a) Metadata harvesting (b) Open Archives Initiative	skos:Concept "Metadata harvesting" skos:Concept "Open Archives Initiative"
DC2003 Metadata Thesaurus	skos:Concept "Open Archives Initiative Protocol for Metadata Harvesting"

(a) Open Archives Initiative Protocol for Metadata Harvesting	

With the change set model, the change can be represented in Turtle format as follows:

changelog: ChangeSet02 a ch:ChangeSet; // **A Change Set**
    ch:hasReason "reason 2";
    ch:hasChangeBeginTime "2017-12-02 15:34:45+5";
    ch:hasChangeEndTime "2017-12-02 15:35:00+5".
    ch:hasOperation changelog:orderedOperations.

changelog:conceptDelete01 a ch:ConceptDelete ; // **A change operation**
    ch:operator   changelog:agent01 ;
    ch:subject skos:Concept "Metadata harvesting" .

changelog:conceptDelete02 a ch:ConceptDelete ; // **A change operation**
    ch:hasTimeStamp "2017-12-02 15:34:45+5" ;
    ch:operator   changelog:agent01 ;
    ch:subject skos:Concept "Open Archives Initiative" .

changelog:conceptAdd02 a ch:ConceptAdd ; // **A change operation**
    ch:hasTimeStamp "2017-12-02 15:34:46+5" ;
    ch:operator   changelog:agent01 ;
    ch:subject skos:Concept "Open Archives Initiative Protocol for Metadata Harvesting" .

changelog:orderedOperations
    ch:hasOrderedOperationList changlog:operationList02.

changelog:operationList02 a ch:OpertionList;
    ch:hasContents changelog:conceptDelete01;
    ch:hasNext changelog:opertionList03.

changelog:operationList03 a ch:OpertionList;
    ch:hasContents changelog:conceptDelete03;
    ch:hasNext changelog:opertionList04.

changelog:operationList04 a ch:OpertionList;
    ch:hasContents changelog:conceptAdd02;

ch:hasNext changelog:endOpertionList.

changelog: endOpertionList a ch:EmptyList.

In this example, there are three operations that should be executed in order. An ordered operation list is defined. The refinement changes in Tennis (2007, p.15) is listed as follows:

Table 3. Change example 3

DC2003 Metadata Thesaurus (a) Cultural heritage [no other concepts]	skos:Concept "Cultural heritage"

DC2004 Metadata Thesaurus
 	
(a) Cultural heritage
   NT Sekisui-zu
	skos:Concept "Sekisui-zu"
	skos:Concept "Cultural heritage"
	  skos:narrower skos:Concept "Sekisui-zu"

The change can be represented as a ConceptAdd and RelationAdd operations. For the sake of simplicity, the detailed statements are not listed here. However, they are similar to the examples above.

In the ontology, there are only add and delete operations. A modification operation can be expressed by one delete and add operations.

DISCUSSION AND CONCLUSION

The goal of this study is to represent KOS changes. Definitions are given to express KOS changes. In addition, a change set model was proposed along with types of changes. As a result, an ontology for representing KOS changes was proposed. This study took the examples from other studies as use cases. Through the examples, it is clear that KOS changes can be expressed by the proposed approach.

Using this approach, KOS changes are explicitly expressed. The explicitly RDF statements can be queried by using SPARQL Protocol and RDF Query Language (SPARQL). Moreover, changes can be separated from the KOS that makes easier to manage KOS and changes. Formally expressed changes and the separation of changes with KOS make it possible to deal with KOS changes on the semantic web.

Five groups of stakeholders (i.e., LOD Dataset producer group, vocabulary producer group, researcher group, web site/tool developer group, and KOS service provider group) of KOS have been identified (Zeng & Mayr, 2018, p. 5). KOS changes need to propagate not only from

producer groups to KOS consumers end but within producer groups. For example, a LOD dataset producer may use more than one KOSs as indicated by Zeng and Mayr (2018). This study is the first step to connect KOS producer groups to consumer groups. The following steps including generating change sets and change propagation from KOS sources to applications need to be investigated. The ultimate goal is to keep KOS current at the application level through change sets without downloading the full KOS every time.

References

Aitchison, J., Bawden, D., & Gilchrist, A. (2000). *Thesaurus Construction and Use: A Practical Manual* (4th ed.). London, United Kingdom: ASLIB.

Ashman, A. B. (2006). The Persistence of Superseded Subject Headings in Online Catalogs. *Technical Services Quarterly*, *24*(2), 27–34. https://doi.org/10.1300/J124v24n02_03

Beckett, D., Berners-Lee, T., Prud'hommeaux, E., & Carothers, G. (2014). *RDF 1.1 Turtle*. W3C. Retrieved from https://www.w3.org/TR/turtle/#language-features

Buckland, M. K. (2012). Obsolescence in subject description. *Journal of Documentation*, *68*(2), 154–161. https://doi.org/10.1108/00220411211209168

Chan, L. M., & Vizine-Goetz, D. (1997). Errors and obsolete elements in assigned Library of Congress subject headings: implications for subject cataloging and subject authority control. *Library Resources & Technical Services*, *41*(4), 295–322.

Flouris, G., Manakanatas, D., Kondylakis, H., Plexousakis, D., & Antoniou, G. (2008). Ontology change: classification and survey. *The Knowledge Engineering Review*, *23*(02). https://doi.org/10.1017/S0269888908001367

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199–220.

Huang, Z., & Stuckenschmidt, H. (2005). Reasoning with Multi-version Ontologies: A Temporal Logic Approach. In Y. Gil, E. Motta, V. R. Benjamins, & M. A. Musen (Eds.), *The Semantic Web — ISWC 2005* (pp. 398–412). Galway, Ireland: Springer.

Javed, M. (2013). *Operational change management and change pattern identification for ontology evolution* (Ph.D. Thesis). Dublin City University.

Javed, M., Abgaz, Y. M., & Pahl, C. (2013). Ontology Change Management and Identification of Change Patterns. *Journal on Data Semantics*, *2*(2–3), 119–143. https://doi.org/10.1007/s13740-013-0024-2

Khattak, A. M., Batool, R., Pervez, Z., Khan, A. M., & Lee, S. (2013). Ontology Evolution and Challenges. *J. Inf. Sci. Eng.*, *29*(5), 851–871.

Khattak, A. M., Latif, K., & Lee, S. (2013). Change management in evolving web ontologies. *Knowledge-Based Systems*, *37*, 1–18. https://doi.org/10.1016/j.knosys.2012.05.005

Manola, F., & Miller, E. (2004). *RDF Primer*. Retrieved from http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

Noy, N. F., Chugh, A., Liu, W., & Musen, M. A. (2006). A framework for ontology evolution in

collaborative environments. In *International semantic web conference* (pp. 544–558). Springer.

Organization, N. I. S. (2005). *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. NISO Press.

Palma, R., Haase, P., Corcho, O., & Gomez-Perez, A. (2009). Change Representation For OWL 2 Ontologies. In Rinke Hoekstra & Peter F. Patel-Schneiderz (Eds.), *Proceedings of the Sixth OWLED Workshop on OWL: Experiences and Directions* (p. 10). Chantilly, VA, United States.

Soergel, D. (1974). *Indexing languages and thesauri: construction and maintenance* (Los Angeles, CA). Melville Pub. Co.

Stojanovic, L., Maedche, A., Motik, B., & Stojanovic, N. (2002). User-Driven Ontology Evolution Management. In A. Gómez-Pérez & V. R. Benjamins (Eds.), *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web* (Vol. 2473, pp. 285–300). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45810-7_27

Tennis, J. T. (2005). SKOS and the Ontogenesis of Vocabularies. In *Proceedings of the International Conference on Dublin Core and Metadata Applications, Madrid, Spain*.

Tennis, J. T. (2007). Scheme Versioning in the Semantic Web. *Cataloging & Classification Quarterly*, *43*(3–4), 85–104. https://doi.org/10.1300/J104v43n03_05

Tennis, J. T., & Sutton, S. A. (2008). Extending the simple knowledge organization system for concept management in vocabulary development applications. *Journal of the American Society for Information Science and Technology*, *59*(1), 25–37. https://doi.org/10.1002/asi.20702

Tunnicliffe, S., & Davis, I. (2009). Changeset. Retrieved March 24, 2018, from http://vocab.org/changeset/

Uschold, M., & Gruninger, M. (2004). Ontologies and semantics for seamless connectivity. *SIGMOD Record*, *33*(4), 58–64.

W3C. (2012). Introduction to SKOS - SKOS Simple Knowledge Organization System. Retrieved March 28, 2018, from https://www.w3.org/2004/02/skos/intro

Zeginis, D., Tzitzikas, Y., & Christophides, V. (2011). On Computing Deltas of RDF/S Knowledge Bases. *ACM Transactions on the Web*, *5*(3), 14:1–14:36.

Zeng, M. L., & Mayr, P. (2018). Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review. *International Journal on Digital Libraries*. https://doi.org/10.1007/s00799-018-0241-2

About the author

Qing Zou is a digital initiatives librarian at Lakehead University Library, Canada. His research interests include Linked Data, integrated library systems, digital libraries, and digital archives.