

Uma Balakrishnan – Verbundzentrale des GBV (VZG), Göttingen, Germany
Dagobert Soergel – Dpt. of Information Science, Univ. at Buffalo, United States
Olivia Helfer – WNYLRC, Buffalo, NY, United States

Representing Concepts through Description Logic Expressions for Knowledge Organization System (KOS) Mapping

Abstract:

Mapping among KOS can be achieved by representing the concepts in each KOS through a canonical expression:
 DDC: 386.8 Inland waterway >Ports \longleftarrow <isa> Traffic station \square <inEnvironment> Inland water transport \longrightarrow Binnenhafen
 Canonical expression GND

We explored representing concepts through description logic expressions using a sample of 150 library KOS classes and subject headings, 50 each from the Dewey Decimal Classification (DDC), the Regensburg Verbandsklassifikation (RVK), and the German Integrated Authority File (GND). Working from the ground up, we compiled a small vocabulary of relationships/roles and the beginnings of a faceted classification of elemental concepts. Large-scale application of this approach requires a large hierarchically structured vocabulary of relationships/roles and a universal faceted classification. We discuss the task of developing these tools drawing on many sources.

1.0 Introduction

A description logic expression is a combination of elemental concepts giving for each elemental concept the role it plays in the context. Adding the roles is a refinement of simple semantic factoring. A DL expression is a formal definition of a concept (Baader et al. 2017). We use a simplified version of DL expressions, similar to the Semantic Code (Perry and Kent 1958). On the other hand, we found, it necessary to allow for nesting in DL expressions, indicated by []. Some examples:

RVK: ZO3700 Technology > Traffic > Traffic safety = <isa> safety \square <inEnvironment> Transportation
GND: Verkehrssicherheit (Traffic safety) = <isa> safety \square <inEnvironment> Transportation
RVK: PN808 Law > ... > Hazardous material = <isa> law (topical area) \square <appliesTo> [<isa> goods \square <hasCharacteristic>hazardous]
DDC: 343.0938 Law > ... > traffic safety = <isa> law (topical area) \square <appliesTo> [<isa> safety \square <inEnvironment> Transportation]
DDC: 343.09322 Law > Transportation > Transportation of goods > Hazardous material = <isa> law (topical area) \square <appliesTo> [<isa> transportation system \square <objectTransported> [<isa> goods \square <hasCharacteristic>hazardous]]
GND: Gefahrgutbeförderungsrecht (Law on transportation of hazardous goods) = <isa> law (topical area) \square <appliesTo> [<isa> transportation system \square <objectTransported> [<isa> goods \square <hasCharacteristic>hazardous]]

Classes or subject headings from two KOS that have the same DL expression refer to the same concept and can be mapped with skos:directMatch. A reasoner working on

a database of DL expressions can infer other relationships between concepts. For example, there is an associative relationship between RVK: PN808 and DDC: 343.0938, since **Safety** and **hazardous** are related; see Section 2.2.

For the context of this work see Balakrishnan et al. 2018, 2019. For the theoretical basis see Soergel 1972, 2011, 2017.

2.0 Creating DL expressions and analyzing pair-wise mappings

2.1 Creating DL expressions for classes / subject headings - The challenges

Requires a good understanding of KOS structure and considerable domain knowledge.

Understanding KOS structure. Consider a class in its hierarchical context

RVK: CM 5000 Information theory, cybernetics is actually
 RVK: CM 5000 Psychology > General, history and methods > Information theory, cybernetics
 = <isa>Information theory, cybernetics \sqcap <usedAsMethodIn> psychology

Domain Knowledge is essential; often need to look up a definition.
 A DL expression is crystallized domain knowledge

DDC: 150.1 Psychology > philosophy, theory, systems, schools
 = <isa> philosophy, theory, systems, schools \sqcap <studies> psychology
 GND: Philosophical Psychology
 = <isa> psychology \sqcap <studies> topics in philosophical psychology \sqcap
 <usesMethod> philosophical method
 GND: Historical psychology
 = <isa>psychology \sqcap <studies> [<isa>topics in psych. \sqcap <inTimePeriod> past]

Ambiguity: Caption is ambiguous, perhaps used both ways. Two DL expressions

DDC: 610. 82 Women in Medicine#1 = <isa> Person \sqcap <sex> female \sqcap <agentIn> Medicine
 DDC: 610. 82 Women in Medicine#2 = <isa> Person \sqcap <sex> female \sqcap <patientIn> Medicine
 GND: Ethnomathematics#1 = <isa> math. \sqcap <practicedAmong> identifiable cultural group
 GND: Ethnomathematics#2 = <isa> curriculum subject \sqcap <studies> [<rel> culture, <rel> math.]

2.2 Analysis of mapping pairs based on DL Expressions

Given the DL-Expressions for each class or subject heading in a pair taken from KOS A and KOS B, a system can infer the type of mapping, as shown in the examples.

skos:exactMatch

RVK: CP 3200 General psychology > Feelings, emotion = <isa> emotion
 GND: Feeling, emotion = <isa> emotion
 DDC: 150.9#1 Psychology > History, [biographic treatment, biography] >
 = <isa> history \sqcap <studies> psychology
 RVK: CM 2000 Psychology > History of psychology = <isa> history \sqcap <studies> psychology
 The history part of the DDC class (without biography) is an exact match for the RVK class.

skos:narrowMatch
<p>RVK: ZO 9300 Traffic, transport > Transportation system > Transportation of goods = <isa> transportation system \sqcap <objectTransported> goods</p> <p>GND: Law of transportation of goods = <isa> law (topical area) \sqcap <appliesTo> [<isa> transportation system \sqcap <objectTransported> goods]</p>
skos:broadMatch
<p>DDC: 386.6 Inland waterway and ferry transportation > Ferry transportation = <isa> transportation system \sqcap <inEnvironment> inland water transport \sqcap <hasSchedule> regular schedule</p> <p>RVK: ZO 6080 Inland water transport, canals = <isa> transportation system \sqcap <inEnvironment> inland water transport</p>
skos:relatedMatch
<p>DDC: 150.1 Psychology > Philosophy and theory, systems and schools = <isa> philosophy or theory or discussion of systems or schools \sqcap <studies> psychology</p> <p>RVK: CM 5000 Psych. > General, history and methods > Information theory, cybernetics = <isa> Information theory, cybernetic \sqcap <usedAsMethodIn> psychology</p> <p>Both DL expressions include psychology; method relates somewhat to systems or schools</p>
No match – Mapping Error. Lack of domain knowledge of mapping editor
<p>DDC: 150.9#1 Psychology > History, biographic treatment, biography > = <isa> history or biography \sqcap <studies> psychology</p> <p>GND: Historical psychology = <isa>psych. \sqcap <studies> [<isa> topics in psych. \sqcap <inTime> past]</p>

3.0 Toward a system relationships/roles and a universal faceted classification

To apply the approach illustrated at large scale requires a large list of relationships/roles and a large universal faceted classification. The following sections give a flavor of what needs to be done, but the task is monumental.

3.1 Relationship types / roles

For this exploration we introduced relationship types/roles as needed. Some are quite obvious, such as <isPartOf>, <produces>, or <hasCharacteristic>, but others are not so common or quite specialized, such as the following: <usesMethod>, <usedAsMethodIn>, <studies>, <appliesTo>, <inEnvironment>, <agentIn>, <patientIn>, <objectTransported>, <atEducationLevel>.

Our ultimate goal is either to locate each relationship/role in a standard or widely used ontology or to contribute to some inventory of relationship types / roles.

3.2 A universal faceted classification of elemental concept

At this stage of the project, we introduced elemental concepts as needed for the DL expressions, standardizing terminology in this small set of 150 classes and subject headings. The next step would be to develop a universal faceted classification, a monumental task using many sources, including standard classifications, see Figure 1.

The nature of the classes in DDC and RVK often requires elemental concepts that are defined using *or* such as

Comparison or harmonization (with NT *Comparison* and NT *Harmonization*)

Cognition or intelligence

Philosophy or theory or discussion of viewpoints or schools (frequent subdivision in DDC)

<p>We need <i>Level of education</i>, as seen from some concepts we used as components. The International Standard Classification of Education (ISCED) maintained by UNESCO includes a classification for <i>Level of education</i>. For purpose of illustration we built a hierarchy consisting mainly of ISCED concepts (ISCED does not present these as a hierarchy). We added a number of concepts, some to group ISCED concepts, some extensions further down (Kindergarten), because we know or expect that these are needed in building DL expressions for DDC and RVK.</p>	<ul style="list-style-type: none"> ISCED level 0 or 1 Early childhood education to Primary ed. . ISCED: ISCED level 0 – Early childhood education . . ISCED: Early childhood educational development . . ISCED: Preprimary education . . . Kindergarten . ISCED: ISCED level 1 – Primary education Secondary education . ISCED: ISCED level 2 – Lower secondary education . ISCED: ISCED level 3 – Upper secondary education ISCED: ISCED level 4 – Post-secondary non-tertiary ed. ISCED: Tertiary education . ISCED: ISCED level 5 – Short-cycle tertiary education . ISCED: ISCED level 6 – Bachelor’s or equivalent level . ISCED: ISCED level 7 – Master’s or equivalent level . ISCED: ISCED level 8 – Doctoral or equivalent level K-12 education . NT Kindergarten . NT ISCED level 1 – Primary education . NT Secondary education Youth and adult education level . Youth education level . NT Secondary education . Adult education level . NT ISCED: ISCED level 4 – Post-sec. non-tertiary ed. . NT ISCED: Tertiary education
--	--

Figure 1. Hierarchy of Level of education

3.3 Concept definitions

Elemental concepts must be defined, particularly concepts that have been introduced specifically for DL expressions. Consider our pattern for DL expressions for subject disciplines:

environmental science = <isa> subject discipline Π <studies> environment

But what to do with *psychology*? We could come up with

psychology = <isa> subject discipline Π <studies> mental states and processes and behaviors

But that is a cryptic characterization of what psychologists study. Better:

psychology = <isa> subject discipline Π <studies> topics in psychology

Now, we can create a full definition of *topics in psychology* and put many narrower terms under it, for example from (www.apa.org).

For another example, consider several related meanings of *law*:

law (subject discipline), law (topic area), and Body of law, statutes.

4.0 Conclusions and future work

In our exploration representing KOS concepts through simple DL expressions worked generally well. Comparing KOS mapping using DL expressions with mappings found in our database convinced us that it is worthwhile to test the idea further in a larger pilot that would use proper software to support developing and maintaining the systems discussed in Section 3 and to partially automate the creation of DL expression using linguistic analysis of captions.¹

References

- Baader, Franz, Ian Horrocks, Carsten Lutz, and Uli Sattler. 2017. *An Introduction to Description Logic*. 1 edition. Cambridge: Cambridge University Press.
- Balakrishnan, Umamaheswari, Jakob Voß, and Dagobert Soergel. 2018. "Towards Integrated Systems for KOS Management, Mapping, and Access. Coli-Conc and its Collaborative Computer-Assisted KOS Mapping Tool Cocoda." In *Challenges and Opportunities for Knowledge Organization in the Digital Age: Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal*, edited by Fernanda Ribeiro and Maria Elisa Cerveira. Advances in knowledge organization 16. Baden-Baden: Ergon, 128-136.
- Balakrishnan, Umamaheswari and Dagobert Soergel. 2019. "Concept Mapping Through a Hub: Coli-Conc Pilot Study." Presented at *ISKO-LC 2019 Conference*. <https://doi.org/10.5281/zenodo.3257136>
- Perry, James W. and Allen Kent. 1958. *Tools for Machine Literature Searching: Semantic Code Dictionary, Equipment, Procedures*. New York: Interscience Publishers.
- Soergel, Dagobert. 1972. "A General Model for Indexing Languages: The Basis for Compatibility and Integration." In *Subject Retrieval in the Seventies*, edited by Hans Hanan Wellisch and Thomas Daniel Wilson. New York: Greenwood; College Park, Md.: University of Maryland, School of Library and Information Services, 36-61.
- Soergel, Dagobert. 2011. "Conceptual Foundations for Semantic Mapping and Semantic Search." In *Concepts in Context: Proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization July 19th - 20th, 2010*, edited by Felix Boteram, Winfried Gödert, and Jessica Hubrich. Würzburg: Ergon, 13-35.
- Soergel, Dagobert. 2017. "The Principle of Compositionality and Entity-Relationship Modelling: Faceted Classification in a Broader Context." In *Faceted Classification Today: Theory, Technology and End Users: Proceedings Of The International UDC Seminar 2017, London (UK), 14-15 September*, edited by Aida Slavic and Claudio Gnoli. Würzburg: Ergon Verlag, 43-60.

¹ The data sets created in this exploration (list of relationships/roles, partially structured list of elemental concepts, definitions, and DL expressions) are available from the authors.