Proposal for NKOS Consolidated Workshop 2021
Theme: Terminology/Vocabulary Development and KOS Mapping

Transforming the Integrative Levels Classification to SKOS: representation of numbers, dates and people via parallel facets

Claudio Gnoli*, Douglas Tudhope**
* Science and Technology Library, University of Pavia, Italy
** Hypermedia Research Group, University of South Wales, UK

The Integrative Levels Classification (ILC) is a freely-faceted knowledge organization system (KOS) under development since 2004 by an international team of researchers (Gnoli 2016; Park *et al.* 2020). The ILC second edition has recently been converted into SKOS RDF for the purposes of use as linked open data, e.g. in the BARTOC repository (BARTOC). This presentation (a) reflects on the transformation to SKOS RDF representation and (b) considers the appropriate representation of numbers, dates and people both in ILC notation and SKOS.

Transformation to SKOS

The ILC is a complex classification scheme with various features that are challenging to represent in SKOS. Previous work expressing other complex classification schemes in SKOS is briefly reviewed. Significant aspects of the ILC transformation using the STELETO transformation tool are described. Various issues arising are discussed, including the challenges posed by compound faceted descriptors with coordination of concepts, which can also be relevant for such other schemes as Dewey (see e.g. Kyrios et al. 2021), and a limited extension of SKOS to incorporate RDF properties in order to model the ILC's facet indicators together with specialized sub-properties of *skos:related*. Some possible routes towards addressing the complexity of complex classification schemes in indexing and search applications are suggested (Binding et al. 2021).

The second part of the presentation considers the potential of faceted classification to permit the notation for certain phenomena to be derived from that of other phenomena via parallel facets. We will illustrate this through the examples of numbers, dates and persons, all connected via parallel facets. Numbers, dates and persons are connected because the notation for dates (periods) is derived from that for numbers and persons from times.

Numbers belong to main class *a* of forms and logico-mathematical structures, and its subclass *an* of quantities. The next letter specifies order of magnitude: *anb* negative quantities, *ann* units, *ano* tens, ... *ant* millions, etc. Further letters are used like digits, both negative and positive, to specify the exact quantities: e.g. *q* is positive two and *t* is positive five, so that *annq* means the number 2, *annqt* means 2.5, *antt* means five millions, *anttq* means 5,200,000 and so on. Although representing decimal digits by letters may appear a surprising solution, it allows the automatic sorting of negative and positive digits in an appropriate sequence: for example, the number -7 is *anbng* and -4 is *anbnj*.

Similar issues arise with **dates** expressed in years of the Common Era, as these exist in both negative (BCE) and positive (CE) form. In ILC2, dates are listed under class *r* because they are a cultural creation connected to calendar traditions (though referring to *b* spacetime in absolute terms). Class *rab* of historical periods can be divided into millennia, centuries, decades and years by the same "digits" illustrated above. This allows the expression of dates

from 9999 BCE, roughly corresponding to the beginning of the Holocene (rise of agriculture), to 9999 CE. For example, *rabg* means the 8th millennium BCE and *rabpwox* means the year 1809 CE. Periods can further be specified by additional digits for months, days, hours etc., e.g. *pwoxqm* is 12 February 1809.

Another related array is that of *px* **persons**. Persons can be specified by various facets, including occupation, place of birth and date of birth. However, the most effective way to classify persons has been found to be their birth time, which produces an automatic chronological ordering. Basically the same idea is implemented in Ranganathan's Colon Classification. Facet *p91 [rab]* allows to specify persons by birth time, with time taken from *rab* historical periods: thus *px91g* means "persons born in 8th millennium BCE" and *px91pwox* means "persons born in 1809 CE". In SKOS data, this is expressed by the following triples:

<http://www.iskoi.org/ilc/2/class/p91>
  <http://www.w3.org/2000/01/rdf-schema#domain>
  <http://www.iskoi.org/ilc/2/class/p> .
<http://www.iskoi.org/ilc/2/class/p91>
  <http://www.w3.org/2000/01/rdf-schema#range>
  <http://www.iskoi.org/ilc/2/class/rab> .

On the other hand, faceted compounds, such as *px91g,* are not usually included in the schedules (and SKOS transformation) and need to be generated in each case at the application stage when indexing or searching.

To produce a first set of examples in the schedules, some tens of persons have been selected from lists of the most famous persons found on the Internet, including prominent leaders, thinkers and artists. Among them, three are all born in 1809 -- Charles Darwin, Abraham Lincoln and Edgar Allan Poe. Adding digits for month and date allows to specify Poe as *px91pwoxpt* "persons born on 19 January 1809"; however, by a strange coincidence both Darwin and Lincoln were born on 12 February 1809, so that even the birth hour is needed to distinguish them in a general list of people: *px91pwoxqme* (born on 12 February at 3 AM) works for Darwin and *px91pwoxqmh* (born at 6 AM) works for Lincoln! In practice, such detailed distinctions will not be needed often, as persons in turn occur in faceted combinations connected with specific topics, so that *t93pwox* "polities, ruled by persons born in 1809" will most probably refer to Lincoln.

Conclusion

Complex classification schemes, such as the ILC, are not straight forward to represent in SKOS. While most of the elements of the ILC can be modelled within SKOS, the ILC's facet relationships and its ability to express coordinated compound descriptions extend beyond the SKOS standard. The presentation discusses the use of *rdf:Property* sub-hierarchies to represent ILC facet indicators. Extensions of the associative relationship (sub-properties of *skos:related)* are used to model the ILC's fundamental facet relationships. This approach permits a method for the representation of faceted compound descriptors in the ILC and potentially other faceted classification schemes.

Our examples also show how a classification based on phenomena can define every concept uniquely in a specific place and connect it to other classes through parallel facets, which can be represented in RDF by specification of property range. While parallel facets can be found here and there in other systems, the structure of ILC and its SKOS representation allow their definition in a formal way, that is suitable for systematic exploitation in both classifying and browsing interfaces.

Acknowledgements
Ceri Binding designed and implemented the SKOS transformation of the ILC and created the RDF ILC representation. Andreas Ledl and Jakob Voss have made the ILC and its RDF representation available on BARTOC.

References

BARTOC. ILC (edition 2) in BARTOC registry. https://bartoc-skosmos.unibas.ch/ilc2/

Binding, C, Gnoli, C, Tudhope. D. 2021, 'Migrating a complex classification scheme to the Semantic Web: expressing the Integrative Levels Classification using SKOS RDF', *Journal of Documentation.* https://pure.southwales.ac.uk/en/publications/migrating-a-complex-classification-scheme-to-the-semantic-web(b8c20e03-9f0b-4151-9eab-e564b9e27d20).html

Gnoli, C. 2016. “Classifying phenomena, part 1: Dimensions”, *Knowledge Organization*, 43(6), pp 403-415.

Integrative Levels Classification. http://www.iskoi.org/ilc/

Kyrios, A., Wehrling P. et al. 2021. 'Updates from the Dewey as Linked Data Working Group', European Dewey Users Group 2021 Meeting. Available at: https://edug.pansoft.de/tiki-index.php?page=2021+meeting

Panzer, M, Zeng M. 2009. ‘Modeling Classification Systems in SKOS: Some Challenges and Best-Practice Recommendations’, *Proc. International Conference on Dublin Core and Metadata Applications,* pp. 3-14. Available at: https://dcpapers.dublincore.org/pubs/article/view/974

Park, Z, Gnoli, C, Morelli , D. 2020. ‘The second edition of the Integrative Levels Classification’, *Journal of Data and Information Science,* 5(1), 39-50

Zeng, M, Panzer, M, Salaba, A. (2010), “Expressing classification schemes with OWL2”, in Gnoli, C. & Mazzocchi, F. *Paradigms and conceptual systems for knowledge organization: proceedings of the Eleventh International ISKO Conference,* Ergon, Würzburg. pp. 356-362.