Commerce, see also Rhetoric: cross-discipline relationships as authority data for enhanced retrieval

Claudio Gnoli
University of Pavia Science and Technology Library, Pavia, Italy

Rodrigo De Santis
Paraná Federal Institute of Education Science & Technol., Irati, Brazil

Laura Pusterla
University of Pavia Science and Technology Library, Pavia, Italy

Abstract: Subjects in a classification scheme are often related to other subjects in different hierarchies. This problem was identified long ago by Hugh of Saint Victor (1096?-1141). Still with present-day bibliographic classifications, a user browsing the class of architecture under the hierarchy of arts may miss relevant items classified in building or in civil engineering under the hierarchy of applied sciences. To overcome these limitations we have developed SciGator, a browsable interface to explore the collections of all scientific libraries at the University of Pavia. Besides showing subclasses of a given class, the interface points users to related classes in the Dewey Decimal Classification, or in other local schemes, and allows for expanded queries that include them. This is made possible by using a special field for related classes in the database structure which models classification authority data. Ontologically, many relationships between classes in different hierarchies are cases of existential dependence. Dependence can occur between disciplines in such disciplinary classifications as Dewey (e.g. architecture existentially depends on building), or between phenomena in such phenomenon-based classifications as the Integrative Levels Classification (e.g. fishing as a human activity existentially depends on fish as a class of organisms). We provide an example of its representation in OWL and discuss some of the details.

Keywords: relationships; existential dependence; query expansion; disciplines; phenomenon-based classification; Dewey Decimal Classification; Integrative Levels Classification

1. Introduction

The most typical structure of all classifications (be they synthetic or not, discipline-based or not, aimed at classifying books or other objects) is a hierarchical tree of classes. Every concept is subsumed under a more general one, and so on, thus forming one big pyramidal graph. This structure offers many benefits for both navigation between concepts and search of classified items: for example, by a single search for a truncated notation a user can retrieve all items classified in the selected class or in any more specific class.

On the other hand, the real world is obviously more complex than a single big pyramid. Concepts are related in many ways, not just by vertical dominance. While the hierarchical structure is an effective facility for displaying some relationships, it also hides other relationships that may be relevant too.

This problem was identified long ago by Hugh of Saint Victor (1096?-1141), the Saxon theologian and philosopher whose classification of knowledge had a profound impact on Western knowledge organization in the subsequent centuries. As recently reported by Olson (2010: 134), while discussing his system Hugh observed that e.g. a class such as commerce could be classified not only among the practical arts, but in principle also under rhetoric, as "a peculiar sort of rhetoric [...] for eloquence is in the highest degree necessary to it". In other words, despite any classification of knowledge into separate disciplines, or more generally into separate main classes, a subject listed in any given class is often related to different classes as well.

Beyond Hugh's particular examples, which are obviously related to the culture of his time, this logical problem still affects any classification of the present day, and the retrieval of information resources through it. Indeed, due to the hierarchical structure of classifications, each subject has to be listed under one branch or another, no matter how many others have some semantic relationship with it. As users will browse bibliographic items classified under a certain branch, they will be missing potentially relevant ones classified under related branches.

This paper explores strategies to cope with this problem in both conceptual classification structures and their implementation as authority data in online catalogues or digital libraries. We will start by discussing the classical application of a disciplinary classification in library catalogues (Section 2), and present a web-based library tool for exploring classes and relationships between them in science and technology domains (Section 3). Then we will transcend the special case of discipline-based library classifications, by considering non-hierarchical relationships in the more general perspective of ontology (Section 4) and their representation in the Web Ontology Language (OWL) (Section 5).

2. Cross-discipline relationships in libraries

In traditional library classifications, relationships other than hierarchical are usually expressed in the schedules as "see also" links, pointing users from a given place in the scheme to a different one belonging to another hierarchy. As most library classifications are based on disciplines, such relationships can be described as cross-discipline ones: an example in the Dewey Decimal Classification (DDC) is the relationship between class `378` for university, belonging to the discipline of education, and class `727.3` for university buildings, belonging to the discipline of architecture. "See also" references are also common in verbal subject headings lists, such as the Library of Congress Subject Headings, and are basically equivalent to "related terms" (RT) recorded in thesauri.

In terms of authority control, cross-discipline relationships can be recorded in special fields of the classification reference database accounting for relationships of a class with others in a different hierarchy. The UDC Master Reference File includes field 125 for references to the notation of other classes, which "represent[s] 'see also' reference, i.e. it points to related classes. When there is more than one reference examples the group is ordered according to the UDC filing rules" (UDC, 2003; 2012).

In reviewing the requirements of interfaces to classifications, Slavic (2006) lists "hyperlinks on class numbers, see also reference" among the possible functions of browsing and classified display; she emphasizes that "a systematic (classification) display can be adapted for search expansion, enhancing both recall and precision, through exploiting hierarchical and associative relationships, provided that semantic linking in the vocabulary is fully supported by vocabulary data".

Despite the opportunity to leverage this kind of authority data for retrieval, most library catalogues hardly provide any kind of link in their browsable lists of classes, as a wide survey of Italian online catalogues has shown (Casson, Fabbrizzi & Slavic, 2011). Display of related concepts to be used for query expansion has been demonstrated with a faceted thesaurus (Binding & Tudhope, 2004; Tudhope et al., 2006).

Similar cross-disciplinary references would also be very useful on library shelves arranged by some classification scheme. Unfortunately, applications of this kind are not common. A recent suggestion was to provide shelves containing a given DDC class with QR codes pointing to related classes in the online catalogue (Green, 2013), so that users can leverage their smartphones to navigate across the classification tree, and integrate the material space of the library with the digital space of its catalogue. An Italian institution making use of QR codes to connect signage at open shelves with the catalogue, and paying attention to illustration of the structure of DDC to users, is the Library of Social Sciences at the University of Florence (Fabbrizzi 2014). In this implementation, display of cross-discipline relationships is delegated only to the catalogue:

> "each classified subject has relationships with other subjects, especially those which share one or more concepts within different disciplines [...]. If these relationships are recorded in the catalogue, it is possible to give the users the opportunity of a network of search paths that otherwise would not be evident due to the organization by disciplines" (Fabbrizzi 2014: 115).

3. SciGator: an interface for exploring cross-discipline relationships

The Science and Technology Library of the University of Pavia was founded in 2009 by merging pre-existing smaller libraries, as part of a reorganization process. It gathers information resources mainly concerning engineering, architecture, natural science and mathematics. More resources in the related disciples of physics, chemistry, pharmacology, and medicine are gathered in the separated Science Library and Medical Library. As all these libraries are increasingly adopting DDC for shelving, providing classification data in the catalogue is an important way to overcome the scattering of information resources across several institutions and buildings.

Besides searching the university catalogue by authors and titles, users can access information on the collections through the university libraries website: http://biblioteche.unipv.it/BST. Since 2015, this includes a browsable interface called SciGator, which allows navigation across DDC classes used in all three scientific libraries mentioned above. Only classes assigned to a significant number of owned documents (i.e. some 5 or more shelfmarks) are displayed. Data from local classifications previously used for shelving older resources in the same libraries are also included, and mapped to approximately equivalent DDC classes.

The relevant classification data are stored in a MySQL database that can be queried by a dynamic web page written in PHP. These include:

- DDC class notation;
- caption in Italian;
- caption in English;
- scope notes;
- related DDC classes from different hierarchies;
- equivalent classes from local schemes.

A book to be shelved often deals at one time with subjects listed in several places in DDC schedules, e.g. with `624` civil engineering, `690` building and `720` architecture. Although a single class has to be chosen for the shelfmark, based on the disciplinary perspective prevailing in the book, links to others will be suggested in the SciGator interface. Table 1 lists cross-discipline relationships that have been found to be relevant to the particular collections of these libraries (irrespective of whether they are recorded as "see also" relationships in the published editions of DDC or not).

Table 1: *Cross-discipline relationships in the scientific libraries in Pavia*

004 computer science	←	519.4 numerical analysis
004 computer science	←	621.39 computer engineering
005.1 programming	←	511.3 mathematical logic
152.1 sensoryperception	←	612.8 physiology of nervous system
304.5 genetic factors in social processes	←	591.5 ethology
338 microeconomy	←	650 management
344 social and environmental laws	←	333.7 natural resources
344 social and environmental laws	←	363 security and social problems
344 social and environmental laws	←	618 public health
346 private and civil law	←	710 land planning
363.7 pollution and environmental problems	←	577 ecology
519.4 numerical analysis	←	515 calculus
569.9 palaeontology of hominids	←	599.9 physical anthropology
581.3 plant genetics	←	576.5 general genetics
599.935 human genetics	←	576.5 general genetics
610.72 medical statistics	←	519 statistics
611 human anatomy	←	571.3 comparative anatomy
612.015 medical biochemistry	←	572 biochemistry
612.64 human embriology	←	571.8 embriology
616.052 medical genetics	←	599.935 human genetics
616.89 psychiatry	←	153 mental processes
620.2 applied acoustics	←	534 physics of sound and vibrations
621.3 electrical engineering	←	537 electromagnetism
621.31 electric energy	←	333.7 natural resources
621.36 optical engineering	←	535 optics
621.4 engines	←	536 thermology
622 mining engineering	←	551 geology
624.176 earthquake engineering	←	551.22 seismology
627 hydraulic engineering	←	532 fluid dynamics
627 hydraulic engineering	←	551.48 hydrogeology
628.5 pollution control technologies	←	363.7 pollution and environmental problems
630 agriculture	←	580 botany
631.53 plant reproduction	←	581.3 plant genetics
632 agrarian pathology	←	581.2 plant pathology
660.6 biotechnology	←	572 biochemistry
660.6 biotechnology	←	610.28 biomedical instrumentation
663.2 oenology	←	634.8 viticulture
664 food technology	←	641 food science
681 precision tools and circuits	←	621.3815 components
690 building	←	624 civil engineering
696 interior plants	←	628 sanitary engineering
711.4 city planning	←	307 communities
720 architecture	←	690 building
725.5 health buildings	←	610.28 biomedical instrumentation
727.3 university buildings	←	378 university education

728 houses	←	307 communities
912 cartography	←	551 geology

Once a user has identified a relevant class, she can start three different kinds of search, each marked by a different icon, for corresponding items in the online catalogue:

- search all records having this class, or a subclass of it, in the Shelfmark field of location data, that is, all documents shelved in this class in any of the library sections that are already organized by DDC;
- search the records above, plus records having this class, or a subclass of it, in the Classification field of bibliographic description data (often assigned by other libraries) irrespective of their shelfmarks;
- search the records above, plus records having one of the related classes, or a subclass of them, in either the Shelfmark or the Classification field.

In all three cases, results are sorted in reverse order of publication data, assuming that most recent documents on the given subject can be more relevant for a majority of users. Also, results can be further refined by the standard search options available in the catalogue (only documents in one selected library, only documents of a selected type, only documents published within a selected range of years, only documents within the resulting set having a specific author or word in the title, etc.). However, experience suggests that few users pay attention to such refinement options.

Clearly, the second and third search options will yield an increasing number of results as compared to the first one. Indeed, more inclusive options produce a higher recall, especially in the case of the expanded query that includes several related classes. This will avoid missing relevant documents that have been assigned related classes, but at the same time will reduce precision and increase information overload: for example, some documents on building techniques may be irrelevant to a user only looking for information on the artistic aspects of architecture.

To face this situation, we have identified two strategies, that have been embodied in the search interface:

- making expanded query only an option, with clear instructions concerning its specific purpose, rather than making it the default search mode;
- limiting expansion to a single step in the graph of relationships; e.g., selecting civil engineering will also include building (as it depends on civil engineering), but not architecture (as it depends on building).

The latter strategies build on reported experience with testing query expansions by thesauri (Tudhope, Alani & Jones, 2001). While all concepts in a knowledge organization system can be virtually connected in a single graph, as one moves from one node to the next, the relationship becomes less and less strict. As a result, concepts connected through more than one or two steps, although having some semantic relationship with the concept searched, may be scarcely relevant to most users.

For this purpose, it is useful to have a closer look on the nature of the relationships between classes in a classification system. This is discussed in the next section.

4. Existential dependence between classes

Relationships other than hierarchical, such as "see also" relationships in classifications or RTs in thesauri, are usually described as "associative". Association may be based on several categories of meaning, for example a shared place or a shared time (like with the relationship between night and stars). However, these kinds of loose relationships seem to be of limited use in information retrieval. On the other hand, many relationships across a classification scheme that appear to be more significant for retrieval, such as the relationship between building and architecture, can be described as cases of existential dependence.

Existential dependence is a kind of ontological dependence (Lowe, 2015) in which the extension of a class depends on another for its own existence. Architecture cannot exist if there are no buildings where it can be applied. Buildings in turn cannot exist if there are no engineering principles to create them. Classes related by existential dependence often belong to different hierarchies in a classification scheme. To take another example, while knowledge on crops may belong to the hierarchy of agriculture, it also depends on botanical knowledge on cereals listed under biology, and on meteorological knowledge on climates listed under Earth sciences. This often makes it useful to provide users with links to such related classes, in order to search information resources, or to navigate between them, in more efficient ways.

As most bibliographic classifications, such as the UDC or the DDC, are organized according to disciplines, existential dependence in them takes the form of a link between disciplines: agriculture depends on botany and on meteorology.

Ultimately, dependence between disciplines can be reduced to dependence between phenomena: the existence of crops depends on the existence of some plants and of some climates. In phenomenon-based classifications, such as the

Integrative Levels Classification (ILC)[1], this relationship can thus take the form of a link between classes of phenomena. Such links can also be represented as semantic factors concurring to the meaning of the class of phenomena (Gnoli, 2013); indeed, in a vocabulary or a formalized ontology one could define crops as the result of cultivating some kinds of plants, hence the concept of plants will be a semantic factor needed in the definition of crops. The ILC database structure includes a field for semantic factors. Another project of a phenomenon-based classification, the Basic Concepts Classification, plans to make wide use of synthetic notations derived from notations of the "basic concepts" forming the more complex concepts required (Szostak, 2014). As this may eventually result in long and complex classmarks, clearly a balance has to be found between what is expressed by post-coordinate synthetic notation and what is recorded as pre-coordinate semantic factors in the database.

5. OWL representation of relationships

This kind of authority data, together with the other structural elements of the classification, can also be represented as linked data for the purposes of interoperability. SKOS may be a natural option here, although this format has only a limited capability of representing the sophisticated structure of faceted classifications (Gnoli et al., 2011). Other RDF formats may then be considered, such as the more powerful Web Ontology Language (OWL) (Zeng, Panzer & Salaba, 2010). In the following, we provide an OWL example of representation of existential dependence between classes of such a phenomenon-based classification as ILC.

In ILC, the class `v` "technologies" have subclasses including `vc` "fishing", `vd` "mining" and `vf` "forestry". Other main classes of ILC are `m` "organisms" with subclasses including `mq` "animals" and `g` "bodies" with subclasses including `gx` "crystals". Dependence relationship allows us to express that `vc` "fishing" depends on `mq` "animals"; `vd` "mining" depends on `gx` "crystals"; and `vf` "forestry" depends on `mp` "plants", as graphically represented in Figure 1.

{insert figure 19-Figure1_Gnoli.tif}

Figure 1: *Graphical representation of some ILC classes and their dependence relationships*

The OWL code for classes `v` "technologies", `vc` "fishing" and the dependence

[1] Integrative Levels Classification (ILC) is available at http://www.iskoi.org/ilc/ilc.php.

property `dependsOn` are presented below.

Class `v` "technologies":

```
  <owl:Class rdf:about="http://www.iskoi.org/ilc/ilc_ontology#v">
       <rdfs:label>v</rdfs:label>
       <rdfs:comment xml:lang="en">technologies</rdfs:comment>
       <rdfs:subClassOf rdf:resource="http://www.iskoi.org/ilc/
              ilc_ontology#Classes"/>
</owl:Class>
```

Class `vc` "fishing":

```
  <owl:Class rdf:about="http://www.iskoi.org/ilc/ilc_ontology#vc">
       <rdfs:label>vc</rdfs:label>
       <rdfs:comment xml:lang="en">fishing</rdfs:comment>
       <rdfs:subClassOf rdf:resource="http://www.iskoi.org/ilc/
              ilc_ontology#v"/>
       <owl:Restriction>
              <owl:onProperty rdf:resource="http://www.iskoi.org/
                     ilc/ilc_ontology#dependsOn" />
              <owl:allValuesFrom rdf:resource="http://www.iskoi.
                     org/ilc/ilc_ontology#mq" />
       </owl:Restriction>
</owl:Class>
```

Property `dependsOn`:

```
  <owl:TransitiveProperty rdf:about="http://www.iskoi.org/ilc/
       ilc_ontology#dependsOn">
       <rdfs:label xml:lang="en">dependsOn</rdfs:label>
       <rdfs:range rdf:resource="http://www.iskoi.org/ilc/
              ilc_ontology#Classes"/>
       <rdfs:domain rdf:resource="http://www.iskoi.org/ilc/
              ilc_ontology#Classes"/>
</owl:TransitiveProperty>
```

The type `TransitiveProperty` asserts that the property `dependsOn` is able to be inherited for the whole hierarchy. It means that a class at a greater degree of specificity, e.g. `mqvhj` "salmoniformes", will automatically keep the dependence relationship with `vc` "fishing", throughout degrees `mq` "animals", `mqv` "chordates" and `mqvh` "ray-finned fish".

However, the transitivity of `dependsOn` has a side effect: it also expresses that `v` "fishing" could also depend on `mqvtoc` "canidae", or, in other words, that we would be able to fish a dog. A solution can be to create another property in class `mq`, called `fishable`, to set its value for each subclass, defining whether an animal can be "fishable" or not, and to set a new restriction for property `dependsOn` in class `vc` "fishing":

```
  <owl:Restriction>
       <owl:onProperty rdf:resource="#isFishable" />
       <owl:hasValue rdf:resource="true" />
</owl:Restriction>
```

It is indeed possible in OWL to define properties unlimitedly and to represent different types of relationships between classes, properties and individuals (beyond the usual "is-a" and "is-part-of" relationships). On the one hand, this makes the initial work of creating a new scheme more complicated. On the other hand, it allows the model to be more flexible and powerful, allowing the representation of sophisticated schemes, which seems to be promising in the case of ILC.

6. Conclusion

Knowledge organization systems, including classifications, can imply complex and rich graphs of relationships between concepts. The classical hierarchical structure and display of classification schemes only show some of them, that is, the relationships between more general and more specific classes of the same branch. This is already a powerful tool, for both navigation and search, as it allows e.g. using truncated notation in queries to retrieve all subclasses in the branch in one search.

On the other hand, non-hierarchical relationships between classes, such as existential dependence, are lost in most interfaces. We have shown how these can be represented to enhance both exploration of the whole classification scheme and recall in search. To achieve this, we had to operate on both a conceptual plane, by evaluating which and how many relationships should be used without obtaining retrieval of scarcely relevant information, and a technical plane, by recording the relevant relationships in a suitable database structure as part of classification authority data. We have also shown how such representation can be translated into RDF structures for uses in the semantic Web.

A more precise and rich representation of relationships as authority data, than is usually implemented, has the potential of providing users with more powerful tools for knowledge discovery, and making them more aware of the complex network of connections that form the corpus of human knowledge.

Acknowledgments

The paper benefits from discussions during first author's Erasmus Staff Training in May 2013 at the University of South Wales Hypermedia Research Unit, led by Douglas Tudhope, and from practical work at the University of Pavia Science and Technology Library, directed by Anna Bendiscioli.

References

Binding, C.; Tudhope, D. (2004). KOS at your service: programmatic access to knowledge organization systems. *Journal of Digital Information*, 4 (4). Available at: https://journals.tdl.org/jodi/index.php/jodi/article/view/110/109.

Casson, E.; Fabbrizzi, A.; Slavic, A. (2011). Subject search in Italian OPACs: an opportunity in waiting? In: *Subject access: preparing for the future.* Edited by P. Landry et al. Berlin: De Gruyter, pp. 37-50.

Fabbrizzi, A. (2014). An atlas of classification: signage between open shelves, the Web and the catalogue. *JLIS.it: Italian Journal of Library, Archives and Information Science*, 5 (2), pp. 101-122. Available at: http://leo.cineca.it/index.php/jlis/article/view/10068.

Gnoli, C. (2013). Facets, levels, and semantic factoring. *SRELS Journal of Information Management*, 50 (6), pp. 751-762.

Gnoli, C. et al. (2011). Representing the structural elements of a freely faceted classification. In: *Classification & ontology: formal approaches and access to knowledge: proceedings of the International UDC Seminar 2011, 19-20 September, The Hague, Netherlands.* Edited by A. Slavic, E. Civallero. Würzburg: Ergon, pp. 193-206.

Green, R. (2013). Is this all they have? In: *025.431: the Dewey blog*. Available at: http://ddc.typepad.com/025431/2013/09/.

Lowe, (2015). Ontological dependence. Revised version. In: *Stanford encyclopedia of philosophy*. Available at: http://plato.stanford.edu/entries/dependence-ontological/.

Olson, H. (2010). Earthly order and the oneness of mysticism: Hugh of Saint Victor and Medieval classification of wisdom. *Knowledge Organization*, 37 (2), pp. 121-138.

Slavic, A. (2006). Interface to classification: some objectives and options. *Extensions & Corrections to the UDC*, 28, pp. 24-45.

Szostak, R. (2014). The Basic Concepts Classification as a bottom-up strategy for the semantic Web. *International Journal of Knowledge Content Development & Technology*, 4 (1), pp. 39-51.

Tudhope, D.; Alani, H.; Jones, C. (2001). Augmenting thesaurus relationships: possibilities for retrieval. *Journal of Digital Information*, 1 (8). Available at: https://journals.tdl.org/jodi/index.php/jodi/article/view/181/160.

Tudhope, D. et al. (2006). Query expansion via conceptual distance in thesaurus indexed collections. *Journal of Documentation, 62 (4), pp. 509-533.*

UDC (2003). *Master Reference File: manual.* The Hague: UDC Consortium. Available at: http://www.udcc.org/mrfmanual.pdf.

UDC (2012). *Data elements in UDC MRF exports*. The Hague: UDC Consortium. available at: http://www.udcc.org/files/udc_data_elements_mrf11.pdf

Zeng, M.; Panzer, M.; Salaba, A. (2010). Expressing classification schemes with OWL 2: exploring issues and opportunities based on experiments using OWL 2 for three classification schemes. In: *Paradigms and conceptual systems in knowledge organization: proceedings Eleventh International ISKO Conference, 23-26 February 2010, Rome, Italy*. Edited by C. Gnoli and F. Mazzocchi. Würzburg: Ergon, pp. 356-362.

CLAUDIO GNOLI is a librarian at the Science and Technology Library, University of Pavia. His research interest is in classification theory and he participates actively in the work of the UDC Editorial team (UDC Consortium) and the International Society for Knowledge Organization. He has published papers on this subject in several international journals and conference proceedings. He is a member of the scientific advisory boards of the Universal Decimal Classification Consortium (UDCC) and of the journal Knowledge Organization, and vice-president of the International Society for Knowledge Organization (ISKO).

LAURA PUSTERLA is a librarian at the Science and Technology Library, University of Pavia, where she works on cataloguing and systematic shelf-arrangement. She has recently graduated in Archive, Documentation and Library Science at the University of Pavia with a final paper on indexing historical documents supervised by Paul G. Weston.

RODRIGO DE SANTIS VIEIRA DA SILVA holds a master degree from the Brazilian Institute of Information in Science and Technology (IBICT). He is Professor at the Paraná Federal Institute of Education Science and Technology. His research interests are in information science and technology, knowledge organization and their application to the domain of music. Rodrigo participates in the Integrative Levels Classification project and was a founder of the Aquela Música! project for a semantic database of Brazilian popular music.