Moving towards a Topic-Based DDC

Rebecca Green
Assistant Editor, Dewey Decimal Classification,
OCLC Online Library Computer Center, Inc.
Dewey Editorial Office, LM-556, Library of Congress,
101 Independence Ave., S.E., Washington, DC, 20540 USA
greenre@oclc.org

ABSTRACT
Although oft regarded as a class-based system, many potential uses of the Dewey Decimal Classification (DDC) system are better supported by a topic-based view. This paper explores automated strategies for identifying equivalence and hierarchical relationships between Relative Index headings, one of the steps required to fully implement a topic-based data model for the DDC. Data used in detecting relationships beyond the small number of relationships already explicitly coded include qualifiers in headings, variant name notes, hyphenated words, acronyms, the structure of Relative Index headings, adjectival modification, equivalence and hierarchical relationships between Library of Congress Subject Headings, and synonymy and hypernymy relationships in WordNet.

Keywords
Dewey Decimal Classification, DDC, topics, Relative Index, Library of Congress Subject Headings, WordNet, equivalence relationships, hierarchical relationships.

INTRODUCTION

The Dewey Decimal Classification (DDC) system is often regarded as a class-based system, a perception supported by the common activity of assigning class numbers to bibliographic resources. Both physical and virtual browsing of a collection classified by the DDC is supported by this class-based view.

But other uses of the DDC (e.g., discovery) are better supported by a topic-based view of the classification. Indeed, the class-based view relies on the character of a class as a topic neighborhood (Green & Panzer 2010). Moreover, relationships between DDC classes are often mediated through topics (Green 2011, Green & Panzer 2011). Searching a collection by the components of synthesized numbers (field 085 in the MARC Bibliographic Format) also relies on a topic-based view.

Advances in Classification Research, 2014, November 1, 2014, Seattle, WA.

Analysis leading to a topic-based data model of the DDC (Green et al. 2013) revealed a close affinity between topics and the text of class descriptions, on the one hand, and between topics and terms in the DDC's Relative Index (RI), on the other hand. Indeed, we may say that topics are reflected both by the text of class descriptions and by RI headings. This does not mean, however, that the correspondence between the expression of a specific topic in the class descriptions and in the Relative Index is apparent. The language of class descriptions, while controlled to a degree, has the veneer of natural language, whereas the language of Relative Index headings is more controlled. Multiple topics may be expressed in a single syntactic unit in a class description; that is, it is difficult sometimes to isolate the expression of individual topics in class descriptions, whereas RI headings each aim to express a single topic, albeit often complex. Moreover, some RI headings correspond to mapped terminology rather than being motivated directly by the class description. This is possible because it is not possible in general to enumerate all the topics belonging to a class.

We seek automated/semi-automated strategies for moving the DDC towards a full implementation of the topic-based data model. Steps required in that implementation include the following:

- Identification of topics reflected in class descriptions
- Identification of the correspondence between topics-reflected-in-class-descriptions and Relative Index headings
- Identification of relationships between topics (relationships found in the DDC are summarized in Mitchell 2001 and Mitchell & Vizine-Goetz 2010)
- Capture of topic and topic relationships in the MARC (XML-encoded) classification and authority formats currently used for storing, displaying, and manipulating DDC data

Significant headway has already been made on automating the first and second of these steps, and informal proposals have been worked out for how topic and topic relationships could be captured in the MARC classification and authority formats (Green 2008; Green 2012; Green 2014b). The focus of this paper is thus on automated ways to establish

relationships between topics, by identifying equivalence and hierarchical relationships between the topical RI headings now assigned to DDC classes. The 000s (Computer science, information & general works), to which 2280 Relative Index terms are currently assigned, are used as a case study to determine the extent to which topical RI headings can be related to each other automatically using existing data.

RELATIONSHIPS IN THE DDC RELATIVE INDEX

Relationships between Relative Index headings are chiefly expressed in two ways. The first of these is built into the structure of RI headings, which consist of a main heading and, optionally, one or more subheadings, which restrict the scope of the heading. Thus, the topic expressed by an RI consisting of a main heading and a subheading is often narrower than the topic expressed by the main heading. But, as will be seen later, sometimes the topic expressed by such an RI is better construed as being narrower than the topic expressed by the subheading.

The second manner in which relationships are expressed between Relative Index headings is through see-also references from one heading to another. Such references are asymmetric (in that they are unidirectional). These references may express equivalence relationships (as in Computer software *see also* Computer programs) or hierarchical relationships (as in University libraries *see also* Academic libraries). But it's not the case that these see-also references exist to capture semantic relationships. After all, the semantic nature of the relationship between two headings connected through a see-also relationship is not always discernible in the internal DDC database. It would be more accurate to say that the current see-also references take advantage of relationships, whether implicit or explicit, to direct the user from one heading to a heading under which additional relevant numbers may be found. That is, current see-also references are an efficiency mechanism tied to the print index so that the subheading structure under one heading need not always be repeated under synonymous headings or under narrower headings.

The goal of identifying (and characterizing) the relationships between Relative Index headings can be seen to go well beyond the current state of Dewey data.

EQUIVALENCE RELATIONSHIPS

The task of identifying equivalence relationships in the Relative Index is made simpler by restricting the task to identifying which RI headings that index the same number are synonymous. Of the 931 classes in the 000s, less than half are indexed by two or more RI headings.

Current RI Equivalence Relationships

The lowest hanging fruit are the 49 pairs of Relative Index headings that have already been explicitly coded in the DDC database as synonyms. For example, Data banks, Databanks, and Data bases have all been coded as synonyms of Databases. Desktop computers, Digital microcomputers, Home computers, Microcomputer workstations, Microcomputers, Microprocessors, and Workstations (Microcomputers) have all been coded as synonyms of Personal computers.

Qualifiers in Headings and Variant Name Notes

The editorial rules governing the development of the DDC call for parenthetical qualifiers to be added to headings in various circumstances. Among the qualifiers called for by the rules are single synonyms and acronyms. If it is awkward to list a synonym as a qualifier, the synonym(s) may be given instead in a variant name note. The 000s include only one instance of a caption with a synonym qualifier--Unidentified flying objects (UFOs)—at 001.942. Five variant-name notes exist in the 000s. These tell us, for instance, that Flying saucers and Unidentified flying objects are equivalent terms, as are Data compaction, File compression, and Data compression.

Hyphenated words

Hyphens can be used both for separation, thereby breaking a word into parts, and for uniting, thereby making of multiple words a single word. That is, the alternative to a hyphen may be a space or nothing at all. Since little consensus exists in the use of hyphens, both of these alternatives should be considered whenever a hyphen appears in a word. A user seeking access to the DDC by a given term may represent it as a set of words, as a hyphenated word, or as a single word.

In identifying these alternative word forms, we have used words with hyphens as a pivot point. On the one hand, in Relative Index headings with hyphens, we construct an alternate form in which the hyphens are replaced by spaces; if an RI heading indexing the same class matches this alternative form, the two headings are designated to be equivalent. On the other hand, and in like manner, we construct an alternate form in which the hyphens are deleted and the parts of the hyphenated word run together without separation; this form is tested against other RI headings indexing the same class number, and matching terms are designated as being equivalent. Although 77 hyphenated RIs index the 000s, only four of them have equivalent RIs adopting either of the alternatives noted above.

By using the hyphenated form as a pivot point, we may have missed equivalent forms in which a space in one form is replaced by nothing in an equivalent form. This is the case among the first group of equivalent RI headings noted above, those which are already coded as equivalent in the DDC database. There we had "Data banks" and "Databanks" and also "Data bases" and "Databases." These are equivalent headings in which the hyphenated form which might exist does not exist. We should therefore

amplify the techniques set out above by taking multiword RIs and testing against the other RIs indexing the same class the various alternate forms in which one or more of the spaces between words is deleted.

Acronyms

Identifying the fuller forms of acronyms begins with first detecting that an acronym exists. Relative Index headings containing 2 or more contiguous upper-case letters were identified as probable acronyms. The letters in the candidate acronym became the basis of two regular expressions, against which other headings indexing the same number were searched. The first of these regular expressions used the letter of the acronym as the first letters of words; the second merely maintained the existence of the letters of the acronym, in order. This approach identified EDP and Electronic data processing, on the one hand, and PIM and Personal information management programs, as equivalent RIs. The second regular expression was used only if the first was not matched. This approach was needed to identify MARC format and Machine-readable cataloging—format as equivalent, as well as PRECIS and Preserved Context Indexing System. (Pre-identifying acronyms meant missing such acronyms as Unesco; there is no real impediment to treating every one-word heading as a possible acronym, in which case United Nations Education, Scientific, and Cultural Organization would also have been identified as being equivalent to Unesco.).

LCSH Equivalent Relationships

The steps above identified equivalent relationships among RIs indexing the same number, using data in the Dewey database. We now turn to look at the identification of additional equivalence relationships using data outside of the DDC. The first source of this additional data we used was *Library of Congress Subject Headings* (LCSH).

Some Relative Index headings have the same form as an LC subject heading. Other Relative Index headings may match lead-in vocabulary for an LCSH. In either case, the set of terms designated as lead-in vocabulary to an LCSH, plus the authorized LCSH itself, constitute a synonym set. A match between a Relative Index heading and any member of an LCSH synonym set sets in motion a search to see if any other member of that synonym set indexes the same Dewey number as the first heading. That heading and each subsequent RI heading matching a member of the LCSH synonym set constitute equivalent terms.

For example, the Relative Index heading Bliss Bibliographic Classification at 025.434 matches an authorized LCSH that has four lead-in vocabulary terms. One of these lead-in terms is Bliss Classification, which matches another RI that indexes 025.434. The equivalence of these two RIs is established through the thesaurus-like structure of LCSH. Six sets of equivalent terms in the 000s, involving 17 RI headings, are established in this manner.

WordNet Synonym Sets

Another source of equivalent terms is WordNet, a lexical database of general English. WordNet provides coverage for over 155,000 words, most of which are nouns. At the heart of WordNet are “synsets” (synonym sets)—sets of synonymous word senses. WordNet synsets are used in the same manner as LCSH equivalence sets to establish equivalence among Relative Index headings.

How can we be sure that the word senses that WordNet sees as synonymous match the word senses in the Relative Index? For example, in WordNet, “delusions” and “hallucinations” are seen as synonyms, but we don’t have those kinds of delusions in mind at 001.96 Errors, delusions, superstitions. Our safety net consists of the word sense disambiguation that is achieved through looking only at the RIs that index the same DDC class. The RI Hallucinations does not index 001.96. If it did, then, yes, we probably would have those kinds of delusions in mind.

WordNet synsets establish the equivalence of approximately 90 terms in 40 sets of equivalent RIs. However, many of these relationships duplicate equivalent relationships also discovered by the approaches mentioned above. Some of the new equivalences found through WordNet include Bigfoot and Sasquatch, Electronic computers and Computers, Systems programs and Systems software, Annuals and Yearbooks, and Reference books and Reference works.

Extending Equivalence Relationships

We take advantage of the regularity in Relative Index heading structure to extend the equivalence relationships discovered above. As a general rule, two Relative Index headings that end with the same subheading are considered equivalent if the parts of the Relative Index headings preceding the shared ending subheading are synonymous. Thus, Punch cards—nonelectronic data processing and Punched cards—nonelectronic data processing are considered equivalent because Punch cards and Punched cards are marked as equivalent in the DDC database; GUI (User interface)—systems programs and Graphical user interfaces—systems programs are considered equivalent because GUI and Graphical user interfaces were previously established as synonyms through our handling of acronyms. And Annuals (Publications)—publishing and Yearbooks—publishing are considered equivalent, because WordNet treats “annuals” and ”yearbooks” as synonyms.

HIERARCHICAL RELATIONSHIPS

In identifying equivalence relationships between Relative Index headings, we restricted our consideration to RI headings that index the same number. While some hierarchically-related Relative Index headings may index the same Dewey number, we would not expect this to be the general pattern. We would typically expect hierarchically-related RIs to index hierarchically-related classes, whether

through the DDC notational hierarchy or through its structural hierarchy (as extended by see references). But where the broader term represents a comprehensive or interdisciplinary treatment of a topic, even the restriction to hierarchically-related classes may be too narrow.

Current RI Hierarchical Relationships

As was the case with equivalence relationships, some number of hierarchical relationships are encoded directly in the DDC database. Again such relationships are governed by concerns for efficiency in the print index, with see-also references leading from narrower terms to broader terms if at least three extra numbers will be found at the broader term. However, such relationships are made only if the broader term is "not apparent." Because hierarchical relationships display only from narrower headings to broader headings, even though the relationships are coded in both directions in the database, we will discuss hierarchical relationships between RI headings from this same perspective.

In the 000s, 56 broader-term relationships have been recorded. The majority of these (46 of the 56) lead to one or the other of three broader terms: Processing modes—computer science (16), Computer communications (18), and Mobile computing devices (12). Examples of hierarchical relationships captured in the DDC database include:

- Tablet computers BT Mobile computing devices,
- Electronic mail BT Computer communications,
- Atlantis BT Legendary places, and
- Learned societies BT Nonprofit organizations.

Relative Index Subheadings

Since subheadings restrict the meaning of a Relative Index heading, it is not infrequently the case that a broader Relative Index heading can be identified simply by deleting the final subheading of an RI. (These are one type of "apparent" hierarchical relationship that would not be encoded in the database under current rules.) Representative of hierarchically-related RIs that follow this pattern are:

- Indexing—subject cataloging BT Indexing,
- Libraries—data processing—personal computer programs BT Libraries—data processing,
- Encyclopedias—French language BT Encyclopedias,
- Parliamentary rules—legislatures BT Parliamentary rules, and
- Newspapers—United States BT Newspapers.

But it is not always the case that the restrictiveness of subheadings operates in a strictly linear order. While some subheadings narrow the preceding topic within the original semantic space, others shift the topic to a different semantic space in which the main heading is the narrowing element. For example,

- Adolescents—libraries for BT Adolescents

does not seem an appropriate relationship, any more than

- Art—information systems BT Art,

or

- Rock music—songs—catalogs BT Rock music—songs.

The subheading "libraries for" changes the semantic type from the semantic type of the main heading to libraries. The subheadings "information systems" and "catalogs" do much the same thing. Thus better hierarchical relationships might be:

- Adolescents—libraries for BT Libraries;
- Art—information systems BT Information systems, or better yet,
- Art—information systems BT Humanities—information systems; and
- Rock music—songs—catalogs BT Rock music—catalogs.

A key principle in the development of Relative Index headings is that the Relative Index should not duplicate the structure of the schedules. That is, if the Relative Index heading for a class is X and it has a subclass that is more specific with respect to aspect Y, we should not expect to find that the Relative Index heading for the subclass is X—Y, which would merely duplicate the schedule. Since the purpose of the Relative Index is to show all the places that a given subject is treated in the schedules, the Relative Index heading at the subclass described above is more likely to be Y—X than X--Y. This means that Relative Index headings consisting of a main heading and a subheading are often of the semantic type of the subheading.

Over a thousand broader-term relationships were posited in the 000s on the basis of subheading structure. A more sophisticated approach is needed, however, to distinguish between subheadings that narrow the semantic space established by the preceding part of the heading and subheadings that establish a new semantic space. The issues involved are revisited in the paper's Discussion section.

LCSH Hierarchical Relationships

Just as equivalence relationships in LCSH records were used to establish equivalence relationships between matching RIs, so, too, hierarchical relationships between RIs can be established on the basis of hierarchical relationships between LCSHs, possibly supplemented by equivalence relationships (RIs may correspond either directly to authorized LCSH or to lead-in vocabulary). Consider, for example, these relationships recorded for the LCSH Science and technology libraries:

150 ## $a Science and technology libraries
450 ## $w nne $a Technical libraries
550 ## $w g $a Special libraries

On the one hand, Technical libraries was previously the authorized entry for the concept now referred to as Science and technology libraries. On the other hand, the LCSH Special libraries is broader (as evidenced by the $w g coding in the 550 field) than Science and technology libraries. On the basis of this LCSH data, we are able to establish that the RI Special libraries, which indexes 026, is broader than the RI Technical libraries, which indexes 026.6. Ninety-nine broader term relationships for RIs in the 000s were established using this data source.

WordNet Hypernyms

WordNet records a number of semantic and other linguistic relationships between its synsets, including hypernymy and hyponymy relationships. These relationships are the linguistic equivalent of BT and NT relationships, respectively. Among them, WordNet distinguishes between (standard) class-based hypernymy and instance hypernymy. WordNet also records that some hierarchical relationships are whole-part relationships (which WordNet refers to as holonymy and partonomy relationships). In establishing hierarchical relationships between RIs on the basis of hierarchical relationships in WordNet, we have made use of both class-based and instance-based hypernymy relationships, but not of holonymy relationships. We achieved word sense disambiguation by looking for hypernym/hyponym relationships in the same class or in the upward hierarchy of the class indexed by the hyponym, with upward hierarchy defined either notationally or structurally.

The following hierarchical relationships in the 000s are representative of the forty-six identified on the basis of hypernymy/hyponymy relationships in WordNet (the class number indexed by the RI term is given after the term):

- Microcomputers (004.16) BT Digital computers (004)
- Internet (004.678) BT Computer networks (004.6)
- Assembly languages (005.136) BT Programming languages (005.13)
- Card catalogs (025.313) BT Library catalogs (025.31)
- Conferences BT Meetings (both 060)

Subordination by Adjectival Modification

A common way of forming subordinate classes in English is to add an adjectival modifier to a noun, which yields the pattern Adjective Noun BT Noun. (A variant of this pattern is Noun Noun BT Noun, in which the first noun of a noun-noun phrase operates pragmatically as if it were an adjective.) This is not merely a common pattern in English, but also in the Relative Index, given that the word order within a main heading or subheading typically imitates the natural word order of English.

Since Relative Index headings are invariably built around nouns, the search for these patterns was implemented simply as a search for pairs of one-word and two-word Relative Index headings in which the two-word heading consists of the word in the one-word heading preceded by any word, and in which the two headings both index the same class or in which the one-word heading indexes a class in the upward hierarchy of the class indexed by the two-word heading.

Representative of the 107 hierarchical relationships identified in the 000s by matching these patterns are:

- Small-scale systems (003.7) BT Systems (003)
- Analytical bibliography (010.42) BT Bibliography (010)
- Public libraries (027.4) BT Libraries (027)
- Maori almanacs (039.9944202) BT Almanacs (030)
- Learned societies BT Societies (both 060)
- Pictorial journalism (070.49) BT Journalism (070.4)

DISCUSSION

Over a dozen routines have been executed to identify equivalence and hierarchical relationships among RI headings. How successful have they been as a whole in building a relational web of Relative Index headings? How might RI subheading structure, which contributes the greatest number of relational links between RI headings, but is undergirded by the weakest support, be better analyzed to produce higher-quality results?

Results

Given that semantic relationships have been captured in the DDC database primarily to support see-also references and given that these see-also references exist to keep the size of the print Relative Index from ballooning, it should not be surprising that the number of explicitly coded equivalence and hierarchical relationships in the Relative Index is reasonably small: the 49 equivalence relationships and 56 hierarchical relationships now explicitly coded in the 000s account for fewer than 5% of the 2280 Relative Index headings in 000s being related to any other heading.

The methods set out above identified 80 equivalence relationships and 1441 hierarchical relationships. These numbers do not tell us directly how many Relative Index headings would thereby be related to some other Relative Index heading: on the one hand, each relationship affects two Relative Index headings; on the other hand, some Relative Index headings are involved in multiple relationships. The 1521 identified relationships involve 1961 Relative Index headings. This results in fully two-thirds of the Relative Index headings being related to at least one other RI.

Relative Index Subheadings Revisited

As previously noted, the apparent success reported above, with two of every three RI headings being related automatically to at least one other RI heading, is muted by the realization that the discovery of the largest number of relationships is the result of implementing the technique

with the weakest support. What can be done differently to achieve better results?

We approached this question through a random sample of twenty (complex) RI headings-with-subheadings in the 000s. For each of these headings, we explored the available data to see if it could be used to identify appropriate broader term RI headings. In most cases, it could. However, the analysis calls for use of a variety of data elements and techniques. Testing over a larger number of RI headings will be required to determine how complete the approach is; prioritization of techniques and/or conflict resolution among multiple outputs may likewise be needed. For purposes of illustration, we will restrict our exploration to RI headings with a main heading and one subheading. Further work will be needed to determine if the approaches set forth here scale up to RI headings with two or more subheadings.

First, we consider circumstances in which the broader term of the complex heading is its main heading, i.e., headings in which the techniques described earlier yield appropriate results:

- *The subheading is a chronological subheading*. All examples with chronological subheadings in the 000s have two or more subheadings; however, an appropriate broader term can be identified by deleting the chronological subheading, no matter which position it occurs in. Unfortunately, the broader terms thus generated for the 000s are not in use, so we illustrate with an example from the 700s: Art metalwork—to 4000 B.C. BT Art metalwork.

- *The subheading is a geographic subheading*. An appropriate broader term can be identified by deleting the geographic subheading, e.g., Museums—Washington (D.C.) BT Museums. If there is more than one subheading and one of them is a geographic subheading, an appropriate broader term can be identified by deleting it, no matter which position it occurs in. (If both chronological and geographic subheadings occur, two appropriate broader terms can be identified by deleting each separately from the original heading.)

- *The subheading gives disciplinary or subdisciplinary context*. For example, the disciplinary subheading "computer science" occurs with RI headings that index numbers throughout the computer science development, i.e., numbers in 004 Computer science, 005 Computer programming, programs, data , and 006 Special computer methods, while the subdisciplinary subheading "library operations" in the final position occurs only with numbers in 025 Operations of libraries, archives, information centers. Broader RI headings can be identified by deleting disciplinary or subdisciplinary subheadings in the final position, e.g., Finance—library operations BT Finance. (Note that the DDC number indexed by Finance is the interdisciplinary number for the topic. The hierarchical relationship between a main heading and the main heading in a disciplinary context represents perhaps a particular kind of hierarchical relationship, just as instantiation is a particular kind of hierarchical relationship). But in some cases—when the RIs with and without the (sub)disciplinary subheadings both index the same number, e.g., General catalogs—library science and General catalogs, the two RI headings might be considered equivalent instead.

- If any of the other approaches for identifying hierarchically related headings applies to either the main heading or a subheading, an appropriate overall broader term RI heading can be identified by substituting within it the broader term RI heading for the component. For example, by using the technique associated with subordination by adjectival modification, we find Hybrid computers—graphics programming BT Hybrid computers—programming. (Hybrid computers—graphics programming BT Computers—graphics programming would work as well, but the latter heading does not exist.) By using LCSH hierarchical relationships, we identify Manuscripts—cataloging BT Nonbook materials—cataloging. (The authority record for the LCSH Manuscripts gives Nonbook materials as its broader term.)

- We extended our equivalence relationships based on a general rule that two Relative Index headings ending with the same subheading are considered equivalent if the parts of the Relative Index headings preceding the shared final subheading are synonymous. We could adopt a parallel general rule for hierarchical relationships: two Relative Index headings ending with the same subheading are hierarchically related if the parts of the Relative Index headings preceding the shared ending subheading are hierarchically related. Consider, for example, the RI Real-time processing—programming, indexed to 005.273. Real-time processing is the caption at 004.33 and Processing modes is the caption at 004.3, evidence that Real-time processing is subordinate to Processing modes. The RI corresponding to the caption at 004.3 is Processing modes—computer science and the RI corresponding to the caption at 004.33 is Real-time processing, so we hypothesize Real-time processing BT Processing modes—computer science. Applying our general rule we hypothesize also Real-time processing—programming BT Processing modes—computer science—programming. This example showcases two additional (and independent) techniques for identifying broader RI headings. The adoption of the general rule for extending hierarchical relationships is

clear and unexceptional; the identification of hierarchical relationships based on RIs corresponding to captions of hierarchically-related classes is not so straightforward as suggested above and may require considerable "fuss" to make it work properly.

Next, we consider circumstances in which the broader term of a complex heading consisting of a main heading and one subheading is its subheading, i.e., headings in which the techniques described earlier would likely *not* yield appropriate results:

- *The subheading is a form subheading*. When the subheading is a form subheading, e.g., bibliographies, dictionaries, serials, statistics, the semantic type of the term (and thus also of the broader term) is that of the form subheading. This gives us Computer graphics—dictionaries BT Dictionaries. But the best broader term one level up for headings with a form subheading is not always the form subheading. Best books—bibliographies BT Books—bibliographies is a better choice than Best books—bibliographies BT Bibliographies.
- *The subheading ends with a preposition*. We have previously indicated that Adolescents—libraries for BT Libraries is a better hierarchical relationship than Adolescents—libraries for BT Adolescents, since the main heading becomes the object of the preposition in the subheading.
- *The subheading implicitly ends with a preposition.* It is relatively uncommon for Relative Index subheadings to end with prepositions. At the same time, it is not so very uncommon for subheadings to stand in relation to a main heading as if the subheading ended in a preposition. Consider the following examples:
  - Academic libraries—collection development [in]
  - Best books—bibliographies [of]
  - Library materials—preservation [of]
  - Union catalogs—library cooperation [through]

  How might we discover in such cases that the semantic type of the broader term is that of the subheading? In the first case above, we can discover that Academic libraries—collection development BT Collection development on the basis of the LCSH record for Academic libraries—collection development, which gives as a broader term Collection development (Libraries). (For the record, the RI Libraries—collection development does not exist.) The second case is taken care of by the technique for handling headings ending with form subheadings, as already discussed. The third case is again solved with reference to the corresponding LCSH, which is Library materials—conservation and restoration. The record for this LCSH gives both Library materials—Preservation (which allows us to make the match in the first place) and Preservation of library materials (which gives us the subheading followed by a preposition) as lead-in vocabulary. Thus we can conclude Library materials—preservation BT Preservation. (For the record, the RI Preservation—bibliographic materials, which is essentially synonymous with Library materials—preservation, also exists.) The fourth case depends on the caption at 021.642 Cooperation through union catalogs, which is the number indexed by Union catalogs—library cooperation. On the basis of this data, and in the absence of Catalogs—library cooperation, we get Union catalogs—library cooperation BT Library cooperation.

This revisiting and further exploration of the structure of RI headings suggests that appropriate broader term RIs for complex RIs may be identifiable automatically in a large number of cases. The multiplicity of approaches and the need at times to interweave those approaches leads one to wonder, however, if the processing needed to identify appropriate broader term RIs might not also posit broader term RI relationships that are not so appropriate. This conundrum can be resolved only through testing.

CONCLUSION

Our investigations-to-date suggest that a significant part of the effort needed to convert the DDC into a topic-based system could be handled automatically. However, additional work will need to be undertaken to see if that potential can be realized. Nowhere is this more true than with respect to the identification of hierarchical relationships involving complex RIs. If this piece of the puzzle can be brought largely under automatic control, the likelihood of success in the project as a whole will be greatly increased.

REFERENCES

Green, Rebecca. (2008). Making visible hidden relationships in the Dewey Decimal Classification: How Relative Index terms relate to DDC classes. In: *Culture and identity in knowledge organization:* p*roceedings of the Tenth International ISKO Conference, 5–8 August 2008,Montréal, Canada*, edited by Clément Arsenault & Joseph T. Tennis, 8–14. Würzburg: Ergon Verlag.

Green, Rebecca. (2011). See-also relationships in the Dewey Decimal Classification. In: *Proceedings from North American Symposium on Knowledge Organization 2011, June 16–17, Ryerson University, Toronto, Canada*, vol. 3, edited by Richard P. Smiraglia, 41–51. Toronto, Canada.

Green, Rebecca. (2012). Modeling topics in the DDC. Presentation for the June 2012 meeting of the Dewey Decimal Classification Editorial Policy Committee (EPC). EPC Exhibit 135-23.3.

Green, Rebecca. (2014a). Relative Index: Managing synonyms. Paper for the June 2014 meeting of the Dewey Decimal Classification Editorial Policy Committee (EPC). EPC Exhibit 137-28.1.

Green, Rebecca. (2014b). Topic-enhanced classes. Paper for the June 2014 meeting of the Dewey Decimal Classification Editorial Policy Committee (EPC). EPC Exhibit 137-28.3.

Green, Rebecca & Michael Panzer. (2010). The ontological character of classes in the DDC. In: *Paradigms and conceptual systems in knowledge organization: proceedings of the Eleventh International ISKO Conference, 23–26 February 2010, Rome, Italy*, edited by Claudio Gnoli & Fulvio Mazzochi , 171–179. Würzburg: Ergon Verlag.

Green, Rebecca & Michael Panzer. (2011). Relationships in the notational hierarchy of the Dewey Decimal Classification. In: *Classification & ontology: formal approaches and access to knowledge: proceedings of the International UDC Seminar, 19–20 September 2011, The Hague, The Netherlands*, edited by Aida Slavic & Edgardo Civallero, 161–176. Würzburg: Ergon Verlag.

Green, Rebecca; Diane Vizine-Goetz; Marcia Lei Zeng; & Maja Žumer. (2013). From modelling to visualization of topic relationships in classification schemes. In: *Classification & visualization: interfaces to knowledge: proceedings of the International UDC Seminar 24–25 October 2013, The Hague, The Netherlands*, edited by Aida Slavic, Almila Akdag Salah, Sylvie Davies, 177–190. Wurzburg: Ergon Verlag.

Library of Congress. (199x?–). *MARC standards*. [Washington, D.C.]: Library of Congress, Network Development and MARC Standards Office. http://purl.fdlp.gov/GPO/gpo13828.

Mitchell, Joan S. (2001). Relationships in the Dewey Decimal Classification system. In: *Relationships in the organization of knowledge*, edited by Carol. A. Bean & Rebecca Green, 211–226. Dordrecht: Kluwer Academic.

Mitchell, Joan S. & Vizine-Goetz, Diane. (2010). Dewey Decimal Classification (DDC), *Encyclopedia of Library and Information Sciences, Third Edition, 1:1*, 1507–1517.

Princeton University. (2010). WordNet: A lexical database for English. WordNet. Princeton University. Retrieved 28 October 2014 from <http://wordnet.princeton.edu>.