

Making Visible Hidden Relationships in the Dewey Decimal Classification: How Relative Index Terms Relate to DDC Classes

Rebecca Green

The schedules of the Dewey Decimal Classification (notation 001-999) include over 25,000 subject classes, while the accompanying tables contribute more than 8500 additional classes. Of a necessity, the scheme offers several types of tools for locating relevant classes, including hierarchical organization of the classes, class-elsewhere notes, see and see-also references, and the Relative Index. In addition to acting as a locating device, the Relative Index indicates which of multiple class numbers for a topic to use for interdisciplinary works and, to a limited extent, comprehensive works. What it does not generally do at present is indicate the relationship between the topic named by the index term and the DDC class that it indexes.

DDC classes are often gathering points for a set of topics. Some of these topics correspond to the caption for the class. Others of the topics are found in class-here notes and are said to “approximate the whole” of the class. These topics may be more-or-less coextensive with the meaning of the class; in any case, they cover at least half the content of the class. Yet other topics occur in including notes and are said to be in “standing room” at the class. Classes are subject to further expansion, in which present standing room topics may become future subdivisions, as justified by literary warrant. For example, Figure 1 gives the caption, including note, and class-here note for class 021.7 Promotion of libraries, archives, information centers; “public relations” is considered a topic broad enough to approximate the whole of the promotion of libraries, archives, information centers, while “friends of the library organizations” is a more specific topic, not covering the majority of the promotion concept.

021.7	Promotion of libraries, archives, information centers
	Including friends of the library organizations
	Class here public relations

Figure 1. DDC class example

The prior statement that the Relative Index does not indicate the relationship between the topic named by an index term and the DDC class that it indexes can now be understood to mean that no record has been kept of which Relative Index terms correspond to the caption, which to class-here concepts, and which to including concepts; that is, which correspond to the overall class, which approximate the whole of the class, and which are only in standing room. The study reported on here represents an attempt to make such designations automatically.

Why would it be useful to ascertain the nature of the relationship between an index term and a class that it indexes? (a) For one thing, standard subdivisions (commonly

recurring forms and topic aspects—for example, serials, geographic treatment, management—that apply broadly across the scheme) can be added for topics that approximate the whole, but not for those in standing room. Explicit indication whether Relative Index terms correspond to topics that approximate the whole or that are in standing room would help human classifiers use the scheme more faithfully and consistently; as the scheme is increasingly used in automated contexts, such a characterization could also be used in verifying the correctness of built numbers or in computer-aided generation of built numbers. (b) The process of determining whether or not an index term approximates the whole of the class it indexes begins with determining which part of a schedule or table entry the index term corresponds to. A side benefit of this determination is identifying which parts of the entry lack a corresponding index term, a situation that might warrant remedy. (c) Addition of DDC numbers to subject authority (e.g., LCSH) records depends on the conceptual relationship between the subject LCSH (an index term) and the DDC class(es) with which it commonly co-occurs (Mitchell 2006). (d) It is hoped that rigorous analysis of the classification as a knowledge representation scheme will eventually lead to automated support of reasoning through DDC classes. It will be crucial in such reasoning to know the scope of topics associated with the class relative to the class itself; do they correspond to or approximate the whole of the class or are they more specific concepts?

Why is the determination whether Relative Index terms correspond to the caption, to class-here concepts, or to including concepts anything other than trivial? For example, once the possible difference in case is accounted for, where is the challenge in determining that the index term “Friends of the library organizations” corresponds to the including concept in 021.7? The first difficulty lies in the fact that individual caption, class-here, and including concepts are not well delineated in the scheme to start with. For instance, the including note for 133.3 Divinatory arts reads “Including fortune-telling by bones, dice, pendulum.” Here we find three concepts—fortune-telling by bones, fortune-telling by dice, and fortune-telling by pendulum—only the first of which occurs fully spelled out in the schedule. A second difficulty lies in the attempt to adopt a form for index terms that is consistent across the scheme, and if possible, also consistent with the form of entry for Library of Congress subject headings. The Relative Index terms that correspond to the fortune-telling concepts at 133.3 include Bones—divination, Dice—divination, and Pendulums—divination. Here the difficulty of detecting the appropriate relationships resides in a difference in number between “pendulum” and “pendulums;” the substitution of a synonym, divination (a morphological variant of a word in the caption for 133.3), for fortune-telling; and a reordering of the core elements of process and instrument. A third type of difficulty goes beyond the typical morphological, syntactic, and lexico-semantic bases of paraphrase and extends into less-regular and more pragmatically oriented relationships. Consider these other Relative Index terms for 133.3, all of which correspond to the phrase “occult methods of foretelling the future” in the class-here note: Forecasting—occultism, Predictions—occultism, and Prophecies—occultism. The relationship of forecasting, predicting, and prophesying to foretelling the future is clear to any human who understands the meaning of the terms, but they might not be listed in lexical resources as synonyms.

Work to date has been limited to the 19,944 index terms (out of 67,611 index term assignments total) without subheadings or qualifiers (both of which often name concepts outside the scope of the local class, giving what Miksa [2006] refers to as “conceptual context”). Three types of matches have been generated: In an EXACT match, the index entry matches either the caption or the subject portion of the including note or class-here note exactly and fully. (The subject portion of an including or class-here note is identified by deleting the initial “Including” or “Class here”, as well as the phrases “comprehensive works on” and “interdisciplinary works on.”) In an IN match, the entire index entry occurs as a phrase within the caption or subject portion of the including or class-here note. In a KEYWORD match, all of the words of the index entry occur within the caption or within the subject portion of the including or class-here note, but order is not necessarily maintained. Differences of case and number are ignored throughout, as are commas and semicolons.

Table 1 gives the results of the current level of processing. Almost two-thirds (65.33%) of the index entries find a match among the caption, including note, or class-here note; a small percentage (1.75%) match more than one of the three.

	EXACT	IN	KEYWORD
Caption	3869	2376	472
Including note	439	3028	373
Class-here note	770	1688	246

Table 1. Preliminary matching results

With the exception of IN matches for including notes, the caption generates more matches than do the including and class-here notes. This is due to the fact that all classes have captions, but many lack including and/or class-here notes; indeed, of 24,754 unbracketed classes, only 4,461 have including notes and 4,860 have class-here notes. This might lead one to assume that only the classes with including or class-here notes need to be considered (or maybe even only those with including notes, since concepts that correspond to class-here notes and approximate the whole of the class are similar to those that match the caption); however, classes may have index terms in conceptual standing room without having including notes.

Future work to be carried out as part of this study includes:

- Matching on morphological variants
- Matching on synonyms
- Partial matching (some but not all of the words in the index term match)
- Developing heuristics for dealing with subheadings
- Developing heuristics for dealing with “missing” data (absent including and/or class-here note)
- Developing a voting scheme for situations where an index term matches more than one of the caption, an including note, and a class-here note
- Evaluating a random sample of the automatic matches

After the current study, the feasibility and value of identifying finer relationships (e.g., hierarchical relationships, cardinality of relationships) between index terms and DDC classes will also be investigated.

References

- Miksa, Francis. 2006. The DDC Relative Index. In Mitchell, Joan S. and Vizine-Goetz, Diane, eds., *Moving beyond the presentation layer: Content and context in the Dewey Decimal Classification (DDC) system*, 65-95.
- Mitchell, Joan S. 2006. Dewey numbers in authority files (discussion paper [for DDC Editorial Policy Committee]). Available http://www.oclc.org/dewey/discussion/papers/epc_126-35.doc.