

# Computer-assisted abridgment of a classification scheme

Rebecca Green and Joan S. Mitchell  
OCLC Online Library Computer Center, Inc., USA

## Abstract

Work is under way to define and test a model for semi-automatic abridgment of the Dewey Decimal Classification (DDC) system. This abridgment is guided by data in the current full and abridged editions and by principles and practices that inform development and maintenance of the classification. The model is immediately applicable to the development of the next abridged edition of the DDC; other applications include derivation of different views of the DDC and management of multilingual content at various levels of development.

## Introduction

The editorial team of the Dewey Decimal Classification (DDC) system is involved in a long-term effort to derive the abridged edition of the classification semi-automatically from the full edition (Green & Mitchell, 2009). Publication of the next abridged edition of the DDC is expected in early 2012, and the edition will be based on the recently published full edition of the DDC (Dewey, 2011). The effort to derive the abridged edition semi-automatically is guided by several motivations, both practical and intellectual:

- A similar process might inform derivation of different views of the DDC, for example derived versions of the classification with contractions, expansions, and/or flexible structures.
- The DDC mixed translation model (Mitchell, Rype, & Svanberg, 2008, 2009) features classes in the vernacular at an abridged level of notation combined with full-edition English-language classes at subordinate levels.
- Translations of the abridged edition are nearly always expanded in targeted areas (Beall, 2003).
- The current process of intellectual logical abridgment is costly (for the Dewey editorial team and for translators).
- The abridgment process has the potential of serving as a quality control measure for review of the full edition.

In short, we anticipate an increasing need for a programmatic approach to delivering flexible editions—abridged with contractions, abridged with expansions, full with contractions, full with expansions. In order to do so, we need a set of rules that will work in most of these circumstances with minimal intellectual effort required for post-editing. The most prominent need is for a set of rules to produce a logical abridgment from the full edition.

## Background

Derivation of a logical abridgment of the DDC depends on understanding the overall structure of the system. The general model is based on the interplay of classes, topics, relationships, and access points. Classes correspond both to a category description and to a class notation. Topics are primarily gathered into a class through its category description, especially through its caption and specific kinds of notes. Of special importance are *including* and *class-here* notes. Access points (including both Relative Index terms and mapped headings from external subject heading schemes / thesauri) lead to DDC classes on the basis of topics in the

classes.<sup>1</sup> In addition to the relationships among classes, topics, and access points just noted, DDC classes are structured into a polyhierarchy that is expressed notationally or, when there is more than one parent for a class, using *see* references.

Abridging such a classification scheme can be approached by addressing four questions: 1) Which classes should be retained? 2) Which topics should be associated with those classes? 3) Which relationships should be retained? 4) Which access points should be retained?

## Classes

The convergence of questions 1) and 2) suggests that in an abridgment of the DDC, classes that are retained are 'the same, only different'. They are the same as the original classes in that they correspond to the same notation; they may be different from the original classes in that their category descriptions may associate a different—an expanded—set of topics with them.

On the one hand, different degrees of abridging the classification are possible. On the other hand, a stable philosophy has evolved of developing abridged editions of the DDC (of which there have been fourteen) for maximum relevance in classifying collections of approximately 20,000 titles. We can therefore leverage past efforts by using the classes found in the current abridged edition as the starting point for the basic notational framework; classes can then be added (or deleted) based on class usage in WorldCat.

## Topics

Abridging a classification scheme does not mean that topics in classes of the full edition that are not retained in the abridgment are no longer relevant. Such topics, if only implicitly, become members of the lowest class in the notational hierarchy above them that is retained. The question is how many of those topics to make explicit.

Answering that question depends in part on a distinction made in the DDC with respect to how topics and classes are related. Some topics are said to approximate the whole of the class, while others are said to be in standing room in the class. That distinction reflects the fact that even the full edition is not as specific as it could be, since a certain level of literary warrant is required before developing a specific class. We might then think of topics that approximate the whole of a class as those that would belong in that class no matter how specifically the system is developed, while topics in standing room in a class would belong to a more specific class if it existed. Topics in *class-here* notes approximate the whole of the class, while topics in *including* notes are in standing room in the class.

## Relationships

A variety of relationships between classes or between topics structure the system (see Mitchell, 2001). The notational hierarchy is the primary relational axis along which the abridgment takes place. *See* references express hierarchical relationships that are not expressed notationally. Other kinds of class-to-class or class-to-class-for-topic relationships are expressed through *class-elsewhere* notes, *do-not-use* notes, *relocation* notes, *discontinuation* notes, *option* notes, and *see-also* references. The relationships to be retained (question 3) are discussed under 'Additions to Other Notes'.

## Abridgment guidelines

We begin with two basic assumptions:

- The abridged edition is a logical abridgment of the full edition.

---

<sup>1</sup> The Relative Index relates topics to disciplines: the topics are arranged alphabetically with terms identifying the disciplines subarranged alphabetically under each topic.

- The abridged edition is derived from the full edition based on a set of rules with minimal intellectual intervention.

As noted previously, we can use the current classes found in Abridged Edition 14 as the starting point for the basic notational framework. These classes are then initially populated by following the same basic guidelines that have been used for creating abridged classes from full-edition classes in manual generation of abridged editions:

- All captions will be the same as those found in the full edition.
- All *including* and *class-here* notes in the full edition will be listed in records in the abridged edition at the same level of notation.
- All *class-elsewhere/see/see-also* references in the full edition will be listed in records in the abridged edition at the same level of notation.

These guidelines provide the classes to be retained with the same basic content they have in the full edition. Our attention now turns to which topics and relationships to add to those classes from full-edition classes not retained in the abridgment.

#### Additions to *Including* Notes

Topics in *including* notes are standing room topics. In the full edition these topics are often candidates for expansion; that is, standing room topics may be moved to a newly developed subordinate class. It stands to reason then that when the scheme is being abridged rather than expanded, topics from full edition subordinate classes should be added to *including* notes. At the same time, it would be counterproductive to add all topics from all subordinate classes to classes retained in the abridgment. First, topics are only taken from classes that will not appear in the abridgment. Further, after some experimentation, we have adopted the rule of thumb to take topics from classes that appear in the full edition one level below the class whose category description is being modified and to limit the topics taken to those appearing in captions and *class-here* notes. If a caption is ascertained to be a 'node label',<sup>2</sup> captions from its subclasses are given as parenthetical examples; standard subdivision<sup>3</sup> captions are excluded. Topics from captions are also skipped if they already appear in the caption of the class being modified.

We will illustrate using one example throughout (see Figure 1), so first, a little background: the class 153.9 appears in both Abridged Edition 14 (Dewey, 2004) and the current full edition, DDC 23 (Dewey, 2011). In the full edition, 153.9 has six subdivisions, two of which relate to standard subdivisions; of the other four, three are one level down and the fourth is two levels down. None of the subdivisions appears in the abridged edition.

In the full edition, 153.9 does not contain an *including* note; the *including* note shown here has been automatically derived from the captions of subordinate classes one level down in the hierarchy plus topics in *class-here* notes associated with those classes. In order to support comparison with the full edition, topics in *including* notes are sequenced according to the order of the subordinate numbers from which the topics are

---

<sup>2</sup> A 'node label' is a caption for a class that operates as a node for gathering together substantive subclasses, but where the class itself lacks substantive meaning; the caption at 321.804, Systems defined by method of selecting chief executives, is of this sort. We treat a caption as a node label if no Relative Index terms are assigned to the class, but Relative Index terms are assigned to two or more of its subclasses.

<sup>3</sup> Standard subdivisions (located in Table 1) represent frequently recurring physical forms or approaches applicable to any subject. They can be added to any schedule or table number for topics that approximate the whole of that number, unless instructed otherwise.

taken. This is different from the alphabetical order generally specified in the DDC editorial rules.

Figure 1. Computer-assisted abridgment example.

153.9 Intelligence and aptitudes

Standard subdivisions are added for either or both topics in heading

Including intelligence tests; comprehensive works on testing and measurement of cognition, of conscious mental processes, of intelligence and personality; aptitude tests; vocational interest tests; superior intelligence

Class here intellect, intelligence levels, multiple intelligences

Class emotional intelligence in 152.4; class factors in differential and developmental psychology that affect intelligence and aptitudes in 155; class use of aptitude and vocational interest tests for academic prognosis and placement in 371.26; class comprehensive works on vocational interests in 158.6

*For personality tests, see 155.2; for educational tests and measurements, see 371.26; for neuropsychological tests, see 616.804/75*

Additions to *Other Notes*

*Class-here notes*

Given that class-here notes contain only topics that approximate the whole of the class, no additions are made to those notes from subordinate classes. The *class-here* note shown in Figure 1 duplicates the *class-here* note that appears in 153.9 in the full edition.

*Class-elsewhere notes*; see references

*Class-elsewhere* notes and *see* references both refer to classes outside the class where the reference is made. Candidate topics for inclusion in these abridged notes generally come from corresponding notes in subordinate classes in the full edition that do not have their own entry in the abridgment. Notes of either type are brought over without limitation on how many levels down they are. A key aspect of processing *class-elsewhere* notes and *see* references is converting the class numbers they refer to, to their abridged equivalents. Some class numbers do not require conversion per se, as the number in the full-edition note also belongs in the abridgment. For other class numbers several strategies are adopted: 1) If a record for the number occurs in the full edition and includes a segmentation note (in which a slash indicates the end of the corresponding abridged-edition number), the number as segmented there is used as the abridged equivalent; 2) If the number is a standard subdivision number, it is separated into its base number and standard subdivision parts. If the base number occurs in the abridgment, then the standard subdivision part is analyzed to determine its abridged equivalent (using these strategies recursively), taking into account if the abridgment occurs in Table 1 (Standard Subdivisions) or in Table 2 (Geographic Areas, Historical Periods, Biography); 3) The rightmost digit of the class number is repeatedly deleted until the number is found in the abridged edition and the result is presented as a segmented number, with the segmentation mark inserted at the end of the abridged number.

A variety of circumstances account for not being able to identify an abridged equivalent number using these strategies. For example, synthesized numbers built on the basis of add instructions outside of Table 1 cannot be

abridged using this set of strategies. All such numbers are reported for further review.

Sometimes when class numbers are converted to their abridged equivalents, they end up referring to the same class from which the reference is made; only if the abridged reference leads outside that class is the corresponding note added to its *class-elsewhere* note or *see* reference. *See* references that end up referring to the same hierarchy are reported for subsequent evaluation and post-processing.

In the full edition, *class-elsewhere* notes at 153.9 refer to 152.4 and to 155, while *class-elsewhere* notes at 153.94 refer to 371.264 and 158.6. Of the four classes referred to, three appear in the abridged; the numbers for these classes are retained as is. The fourth—371.264—does not appear in the abridged. In the automatically derived *class-elsewhere* note shown in figure 1, this class number is shown as its abridged equivalent, 371.26. The *class-elsewhere* notes are sequenced according to DDC editorial rules: references are arranged in class number order, except that comprehensive and interdisciplinary works numbers are given last.

The full-edition class for 153.9 does not contain any *see* references, but the class at 153.93 includes four *see* references. The first of these refers to 153.94; this *see* reference is not brought up to the abridged entry for 153.9, because its abridged equivalent is 153.9. The numbers referred to are represented by their abridged equivalents: 155.28 becomes 155.2, while 616.80475 is analyzed as an abridged number followed by a standard subdivision. As the base number 616.8 exists in the abridged edition and the abridged equivalent for 0475 in Table 1 is 04, an automatically derived segmentation, 616.804/75 is supplied), a reminder of the need to verify the number's accuracy. (In this particular case, the number needs to be compared against instructions for modified standard subdivisions under 616.1-616.8.)

#### Additional note types

The example in Figure 1 illustrates treatment of the most frequently occurring data types in the abridged edition. However, the underlying database includes nearly fifty types of data elements. We prepared a table outlining action on each. For example, we determined that simple add instructions, already customized from the abridged perspective, could be carried over from the Abridged Edition 14 database. Most data elements, however, come directly from full edition records at the same notational level.

*Relocation* notes are handled in the same way as class-elsewhere notes and see references.

*Discontinuation* notes, which report on the shift of a topic or the contents of an entire class to a more general number in the hierarchy, are already a form of abridgment. We include *discontinuation* notes up to one level below the class whose category description is being modified. Because of the discursive nature of Manual notes,<sup>4</sup> we designated them for hand editing. The inclusion of Manual notes in previous abridged editions had been in part at editorial discretion. We took the opportunity to expand inclusion of notes to any DDC 23 Manual note that addressed topics within the notational framework of Abridged Edition 15; this resulted in the addition of more than thirty Manual notes heretofore not included.

#### **Tables**

The full edition includes Tables 1-6, with Table 3 divided into three subsections (A-C). The current abridged edition contains abridged versions of Tables 1-4 (with Tables 3A-C merged into a single table). We were able to process Tables 1, 2, and 4 automatically, but created Table 3 by manually editing Abridged Edition 14 data to reflect the appropriate DDC 23 data resident in the subsections of Table 3.

---

<sup>4</sup> Manual notes provide extended advice on classifying in difficult areas and choosing between related numbers. The notes appear in separate records from the class records to which they apply, and are linked to the class records through see-Manual references.

## Access Points

Our approach to which access points should be retained (question 4) is fairly straightforward. All Relative Index terms at the same notational level plus Relative Index terms that match including or class-here topics exactly will be carried over from the DDC 23 database to the Abridged Edition 15 database. Once the automatic production is completed, we will produce a report for editorial inspection with differences between the indexes associated with Abridged Edition 14 and Abridged Edition 15.

All mapped terminology at the same notational level in the DDC 23 database will also be carried over and associated with classes in Abridged Edition 15 without any editorial intervention. Likewise, we will add the mapped terminology from *Sears List of Subject Headings*, which already contains mappings to abridged edition numbers produced by H.W. Wilson, directly to the database without further editorial intervention.

## Conclusion and future work

The content of Abridged Edition 15 can, to a large extent, be generated automatically because of several important features of the DDC. The first of these is its expressive notation, which permits automatic determination of the correspondence between DDC 23 and Abridged Edition 15 classes. This comes into play both in identifying which DDC 23 classes are in the same hierarchy as an Abridged Edition 15 class and in identifying abridged-equivalent notation for classes mentioned in DDC 23 records.

The second feature is the distinction made in the DDC between topics that approximate the whole of a class and topics that are in standing room in a class. The generation of including notes, which are at the heart of the abridgment process, is totally dependent on this distinction.

A third feature is the well-defined inventory of notes used in the DDC, along with a well-defined order in which the notes should appear in the category description of a class. Differentiating among types of notes allows us to define which aspects of the content of an abridged class should come from the corresponding class of the previous abridged edition and which should come from corresponding classes of the full edition. The well-defined order of these notes permits the correct interleaving of notes from different sources.

This work will do more than inform the development of the next abridged edition of the DDC; it also has implications for the development and maintenance of the DDC and derivations of different views of the DDC. We illustrate this latter point with a single application. The full English-language print edition of the DDC is a subset of a larger data file that includes continuously updated content plus additional synthesized numbers and access points. Some classes in the data file are logical abridgments of expansions that appear in various corresponding editions of the DDC in other languages. These logical abridgments are currently maintained intellectually by the editors. We now raise the question whether our work on semi-automatic derivation of the abridged edition can suggest new approaches to derivations of views of the DDC and maintenance of the underlying data file in a multilingual context. We believe the features of the DDC that support automatic abridgment can also be leveraged in deriving useful views in that larger context.

## Acknowledgments

The authors are grateful for the advice and comments of the Dewey editorial team. DDC, Dewey, Dewey Decimal Classification, and WorldCat are registered trademarks of OCLC Online Computer Library Center, Inc.

## References

Beall, J. (2003). Approaches to expansions: Case studies from the German and Vietnamese translations. Paper presented at the World Library and Information Congress (69<sup>th</sup> IFLA General Conference and Council), 1-9 August

2003, Berlin. Retrieved from <http://www.ifla.org/IV/ifla69/papers/123e-Beall.pdf>.

Dewey, M. (2004). *Abridged Dewey Decimal Classification and Relative Index*. 14th ed. J.S. Mitchell, J. Beall, G. Martin, & W.E. Matthews, Jr. (eds.). Dublin, OH: OCLC.

Dewey, M. (2011). *Dewey Decimal Classification and Relative Index* (23rd ed., Vols. 1-4; J.S. Mitchell, J. Beall, R. Green, G. Martin, & M. Panzer, Eds.). Dublin, OH: OCLC.

Green, R., & Mitchell, J.S. (2009). Rethinking the abridged edition. EPC Exhibit 132-35.1. Paper presented at Meeting 132 of the Decimal Classification Editorial Policy Committee (EPC), Washington, DC, 16-17 November 2009. Earlier reports issued as EPC exhibits 131-36.1 (2009), 129-35.1 (2008), and 128-381 (2007, J.S. Mitchell).

Mitchell, J.S. (2001). Relationships in the Dewey Decimal Classification system. In C.A. Bean & R. Green (eds.) *Relationships in the organization of knowledge*, pp. 211-226. Dordrecht/Boston/London: Kluwer Academic Publishers.

Mitchell, J.S., Rype, I., & Svanberg, M. (2008). Mixed translation models for the Dewey Decimal Classification (DDC) system. In C. Arsenault & J.T. Tennis (eds.), *Culture and Identity in Knowledge Organization: Proceedings of the Tenth International ISKO Conference, 5-8 August 2008, Montréal, Canada*, pp. 98-104. Würzburg: Ergon.

Mitchell, J.S., Rype, I., & Svanberg, M. (2009). Mixed translations of the DDC: design, usability, and implications for knowledge organization in multilingual environments. Paper presented at Looking at the Past and Preparing for the Future, Florence, 20-21 August 2009. Retrieved from <http://www.ifla2009satelliteflorence.it/meeting2/program/assets/MitchellRypeSvanberg.pdf>

#### Corresponding authors:

Rebecca Green can be contacted at [greenre@oclc.org](mailto:greenre@oclc.org)

Joan Mitchell can be contacted at [mitchelj@oclc.org](mailto:mitchelj@oclc.org)