WHY BUILD DEWEY NUMBERS? THE REMEDIATION OF THE DEWEY DECIMAL CLASSIFICATION SYSTEM

Tore Brattli

1. Introduction

Library classification systems use codes to represent subjects of library material. The most visible expression of classification codes is the call number label on the spine of most library books. Together with subject headings, the codes are also used in library catalogs to support subject search. The main advantages of (hierarchical) classification codes compared to subject headings, are the ability to sort documents by subject (e.g. on the shelves), perform hierarchical searches, and an independence of language, e.g. a book on *Computer graphics* has the same Dewey code (006.6) in all languages.

The *Dewey Decimal Classification* (DDC)[1] system is used in more than 200,000 libraries and more than 135 countries worldwide, making it the most widely used system for classification of library materials in the world (OCLC 2012). The system was originally developed by the American librarian Melvil Dewey in 1876. It has since gone through 22 major revisions, with the present 23rd edition (Dewey & Mitchell 2011a) published in the summer of 2011. With the 21st edition in 2000, DDC also became available online[2] as WebDewey (Trotter 1995; Mitchell & Vizine-Goetz 2009, 1510).

1.1. Adaption to print

Dewey is a comprehensive system that through more than 100 years seems to have adapted to the printed book medium in a near optimal manner. One of the challenges has been to keep an increasingly growing system into a "manageable" space (i.e. for printing and usage).

In large hierarchical classification systems like the DDC, many classes are facets or have facet-like properties that are repeated (and are redundant) throughout the system. Instead of repeating properties like serial form, humorous treatment or geographical affiliation thousands of times, these classes can be separated from the schedules[3] and represented only once in separate tables (e.g. table 1-6). Another solution is to represent the more infrequent facets once in the schedules (or as add tables in the schedules) and refer to it with number building instructions (found in notes[4]) from other schedule classes with similar subdivisions. This is the most efficient compression technique used and it saves space in polynomial order. A sub-branch with n classes (e.g. table 1) that applies to most m classes in the schedules, would take $m \times n$ classes if included as an integrated part of the schedules. Number

[1] Copyright by the Online Computer Library Center, Inc.

[2] Conversion to electronic form began in 1979. Electronic Dewey was published in 1993 on CD-ROM.

[3] The classes 000-999 are called the schedules.

[4] Notes are classification rules and guidelines that lead the user to a correct Dewey notation.

building can be traced all the way back to the 1st edition, but according to Miksa (1998, 23) it took off from edition 17, published in 1965.

A second technique, which is less space saving, is used for notes that apply to "many" classes. These notes are normally attached to one or a very few of the relevant classes. Notes with hierarchical force (i.e. that apply to all classes within the same branch) are collected in the root of the subtree. The cost is that the user must look for relevant notes in every class (and number span) on the path from the root to the class in question, both when browsing and searching. More scattered notes are referred to from relevant classes with footnote signs like * or †. For example 577.3 †Forest ecology, which refers to the footnote †*Add as instructed under 577.3-577.6*, which in turn refers to the note for number building under 577.3-577.6.

Using these two compression techniques, the DDC is able to represent a large number of notes and (buildable) classes in three volumes[5] (≈ 3000 pages). The result is very well constructed and many classification experts perceive the DDC as equivalent with number building. At the same time, this compression dramatically decreases the user-friendliness because of the extensive instructions for "decompression", which require at least library school to comprehend. There are many books on number building but probably few librarians that are able to build (or decipher) DDC-numbers, not to mention users. Finding and interpreting notes and number building are complex operations that easily bring forth associations with mind games like Sudoku or crosswords.

On the other hand, without number building the DDC would be literally impossible to print on paper. If we estimate the number of classes in the schedules to approximately 27,000 and in the tables to 9,000 (Mitchell & Vizine-Goetz 2009, 1510) (≈ 10 classes per page), the total number of buildable DDC classes are at least $10^4 \times 10^4 = 10^8$. Because of the system's dynamical nature, it is difficult to say exactly how many classes there are. For example, there seems to be no theoretical limit to how many times you can add from the schedules 001-999). A book covering *Video productions on buildings for Aerospace Museums* gives 778.598 + 7276 + 6291. In practice, the number of DDC classes is more limited. Miksa (1998, 24) estimates 10^9 (i.e. 1,000,000,000) classes, which seems reasonable. A full print version of the DDC without number building would require 100,000 volumes (each 1000 pages), or the equivalent of a relatively large library. In this article, we will use the *class* concept also for built "classes", even if it is unclear from the literature if they actually are regarded as classes.

1.2. Transition to WebDewey

The transition of the DDC from print to WebDewey has made the system more accessible and the awareness and use have increased in many countries worldwide. In Scandinavia, many Swedish libraries are now converting to DDC (from SAB[6] classification) and both Swedish and Norwegian WebDewey translations are now readily available.

[5] The fourth volume is the DDC-index.
[6] Sveriges Allmänna Biblioteksförening (SAB)

Using the terminology of Bolter & Grusin (1999), we can call this transition a remediation, where the old (printed[7]) medium is represented in a new (electronic) medium. The remediation is relatively moderate and closer to a pure digitization than a total absorption. The content is largely the same except for the addition of subject headings (e.g. MeSH) mapped to some classes, while the form is more hypermediated with clickable links and multiple windows. The main presentation has been extended from numerically sorted (or linear) to include hierarchical browsing, and the search is of course improved. Many of these features are however also "available" in the printed version because of DDC's explicit and unambiguous structure. For example, hierarchical browsing can be done in combination with the DDC summaries as a manual binary search. The level of immediacy seems nevertheless higher in WebDewey. Clickable links and a more explicit hierarchy make browsing faster and it is easier to forget the medium, at least until we reach number building or referred notes.

At the same time, it is interesting to note that the system has kept its underlying structure, which largely is tailor made for the printed book medium. The 23rd edition of DDC is recently published on paper and WebDewey is essentially a digital copy of its traditional relative. Today however, there is no need for saving space for neither classes nor notes. DDC classes can be structured as entities in a database with Dewey class numbers, headings (or captions) and notes as attributes. Different DDC tables may correspond to database tables linked together and all the $\approx 10^9$ classes could be made searchable (compared to the current ≈50,000 or 0.005%). The underlying structure can be different or separated from the presentation of the system, and the presentation could be made more user friendly.

It is difficult to find information structures comparable to the DDC still published on paper. In the 1990s, bibliographical reference works like Chemical Abstracts filled up the shelves in chemistry libraries with indexes and references. The users had to manually look up the indexes, then find the reference, then look up in the OPAC and then find the journal on the shelf. Today users simply search, click the reference and click the full text button. Services like Google and Facebook are new products of the new computer medium that would be impossible on paper.

2. Research question

Correct and precise DDC classification is difficult, time-consuming and has a high threshold of learning. This is mainly because of number building combined with locating and interpreting notes (Taylor 2006, 101). The Bachelor program in *Library and Information Science* at the Oslo and Akershus University College teach approximately60 hours of classification and indexing the first year, where DDC constitutes a major part. The author has had extensive e-mail discussions with the DDC experts at the Norwegian national library regarding the classification of seemingly simple subjects. The issue of explicit number building in a computer environment is therefore interesting to examine further.

Is number building necessary for representing today's DDC structure or is it just an "obsolete" technique with the purpose of saving printing space? Relational

[7] We here focus on the printed version, even if it today probably originates from the same (electronic) source as WebDewey.

databases do the same joining of entities from different tables, but well hidden from users. An FRBR[8] based OPAC would hardly be popular if users had to build the bibliographic records themselves. How may DDC be structured behind the scenes and presented with all relevant notes attached to each class and without number building?

Finally yet importantly, does this make the system more user-friendly? The focus in this article is primarily on subject librarians classifying library materials, but the DDC may also be used as an interface for end users searching the library catalog. In fact, the main reason for classifying library material is to improve retrieval for end users. With increasing amounts of electronic library material, most users search from everywhere else but the library, so classification search must be simple.

The method in which to answer these questions will be a theoretical analysis of the DDC structure, and the principles that define it. We will also discuss different underlying representations of the structure that may eliminate number building and propose ideas for a new presentation (or user interface).

Compared to the large distribution and importance of the DDC, remarkably few publications question or justify (user performed) number building, including comprehensive textbooks like Scott (2005) and Chan, Mitchell, & Kungl. biblioteket (2010). The Deutsche Nationalbibliothek is manually prebuilding DDC-classes in collaboration with the users of their WebDewey-system MelvilClass (Mengel, Alex, & Junger 2011). So far, the organization is linear, and a hierarchical sorting is missing. In Sweden, Svanberg (2011) calls for a better presentation of notes where all relevant notes are presented for each class. Svanberg also wants to "[…] allow the difference between the data display in WebDewey and the print DDC to be bigger, in order to take full advantage of the possibilities of the web. The underlying data should of course continue to be exactly the same". OCLC has announced that *number building assistance* is forthcoming in 2012 (Crawford 2012), but no details are available.

3. Analysis of the DDC structure

The DDC is normally perceived as a system where the classes are arranged in a hierarchical tree structure (ordered directed tree) (Mitchell 2001). Each class (or node) has from zero up to 10 subclasses, hence the term decimal. The traditional principle of hierarchy and numbering in DDC is common knowledge and well described in the DDC intro. However, this is a simplified view that does not consider number building as a structure-defining factor.

3.1. Facet representation and number building

Even though the DDC is a hierarchical system, it can represent facets (and compound subjects) quite flexibly as a part of the class number. The subject (from the schedules) is normally expressed first, followed by facets like place and time. Facets are often taken from table 1-2, but may also be taken from table 3-6, the internal tables or from the schedules. The technique is to "jump" across the hierarchies, between suitable classes in the schedules and tables. For each jump, both whole class numbers and parts of class numbers may be collected (i.e. number building) to

[8] Functional Requirements for Bibliographic Records

represent the facet added. For example, Journal of parallel processing computers = 004.618505. We start from **004.618** *Computers distinguished by processing modes, then jump to 004.3**5** *Multiprocessing, collecting the digit 5 and then jump to table 1 for standard subdivisions T1--**05** Serial publications, collecting the last 05.

To ensure that number building produces unique numbers and does not conflict with the representation of other subjects, the DDC-structure has free space "built in". We may perceive this space as the result of the compression. Identifying a facet from the class number only is sometimes difficult, since a facet does not have a set location in different numbers. Often facets begin with the facet indicator 0 (and a few times 1) (Satija 2007, 73; Michell 1997, 86), but not always (e.g. 641.596781 Zanzibari cooking). Number building (and which digits to collect) is always directed by instructions in notes or the *Introduction to the DDC*.

Number building has important implications for how we perceive the DDC structure. The jumping may violate the requirements of a directed tree and turn the hierarchical schedules and tables into a directed graph. A decisive factor here is how we perceive classes like 004.35*Multiprocessing[9]. Is 004.35 only one class or actually two (or more) classes compressed into one? With the latter alternative, we may still perceive DDC as a tree. In both cases, number building can be seen as an extension of the schedules or as the decompression to retrieve the actual structure.

3.2. Guidelines for number building and classification

A hierarchical classification system like the DDC projects documents in a multi-dimensional[10] subject space down to a single dimension. A DDC-number represents a point on the one-dimensional "bookshelf line" from 000 to 999. Sometimes several classes (more or less) match the subject of a document and sometimes it is possible to combine classes for a better match. In order to use the system as intended (e.g. number building) and classify consistently there are a number of rules and guidelines for the classifier (or catalog searcher) to follow. Some are general DDC rules that apply to the whole system and are located in the DDC introduction. Other rules only apply to one or a few classes and are located in notes or the manual. We can divide the rules into two groups: rules that contribute to defining the DDC structure and rules that specify how to use the structure.

3.2.1. Rules that define structure

Add to base number and other add notes can be compared to directed edges (or links) from one class to another across the hierarchies. Adding from table T1 is also such a rule, but since most classes can add from T1, it is more efficient to specify when not to add[11]. We may perceive these rules as the decompressing manual to retrieve the actual DDC-structure.

A question is whether or not citation order defines DDC structure through number building. It probably does, but not directly for the user to utilize. In terms of

[9] The class is also used in number building from 004.618 **Computers distinguished by processing modes*.

[10] i.e. typically more than 100,000 in the vector space model

[11] T1 can be added to all schedule classes once unless it is specified in a note or the subject is redundant.

structure, citation order is predefined through schedules, tables, notes, manual and the general rules. Options also define a more tailor made structure (but this is beyond the scope of this article).

3.2.2. Rules that specify use

Not all facets of a subject can necessarily be combined into a class number. There is no way to represent e.g. *Journal of library research*. Two of the three facets can be represented, but not all three. In some cases, a single (compound or multifaceted) subject may be represented by more than one class, e.g. Adult education 374 and 305.240715. The other group of rules specifies how to use the DDC-structure correctly and consistently. Its main task is to give directions when choosing between two or more equally relevant classes (or facets). Here we have general principles like the author's intension with the work, *First-of-two rule*, *Rule of three*, *Rule of application*, *Rule of zeros* and the *Table of last resort* (i.e. citation order). We also have more local rules like *Notes that Describe What Is Found in a Class*, preference tables and the "opposite" *First-of-two rule*.

Even if adding from T1 define structure, it is not always allowed to use this structure, e.g. Standard subdivisions cannot be added for topics in standing room, nor are other number-building techniques allowed (Dewey & Mitchell 2011b).

3.3. Hierarchy

If we divide a Dewey number into the significant digits to the left and the trailing zeros[12] to the right, the DDC class number hierarchy can be described in recursive terms as follows. A number N is superior (or higher up in the hierarchy) to another number N_{sub} if and only if N_{sub} begins with all of N's significant digits (or if N_{sub} only differs from N in at least one of the trailing zero positions). An interesting consequence of this definition is that the class 000 *Computer science, information & general works* must be regarded as the root of DDC, with 24 subclasses 001-006, 010-090 and 100-900.

A DDC class may have up to $9(n+1)$ subclasses[13], dependent of the number n of consecutive zeros used for standard subdivision. For example, the class 271.90001 is a direct subclass of 271.9 *Congregations and orders of women*. Maximum three zeros are used in the 23rd edition and the class 271.9 has in fact $7+9+9+9=34$ subclasses. Direct subclasses (or child nodes) normally have one or two digits extra, but may (as we have seen) have from zero up to four digits extra (e.g. …0001) with the rightmost digit being nonzero.

If we perceive the DDC as a directed tree, every class (node) can be reached by following exactly one unique path from the root. Since each subclass has a unique number, it follows that every class has a unique number. If we perceive the DDC as a directed graph, it follows that every path represents a unique "class number".

In WebDewey 2.0, the parent class is not always obvious, at least to the inexperienced user. If we take 025.431 *Dewey Decimal Classification* as an example, the actual parent classes are 025.43 → 025.4 → 025 → 020 → 000. It is easy to get

[12] The zeros added to make at least 3 digit numbers, e.g. 020 has two significant digits and one trailing zero.

[13] Optional numbers could add to this figure, but this is not investigated further.

the impression that 02 is the parent class of 025 (from both numbers and layout). A similar confusion arises from the variation in the headings (or captions):

000 Computer science, information & general works
000 Computer science, knowledge & systems
000 Computer science, information, general works

In both cases, the problem is that WebDewey mixes the DDC summaries with the schedules in a confusing manner. All one and two digit "classes" are merely for browsing purposes. In the printed version, this is no problem.

3.4. Optional numbers

There is no rule without exceptions. Some classes have the option to use letters from the Latin alphabet to make deviations from the standard decimal division. Local emphasis, shorter notations and improved division by other characteristics are some of the reasons mentioned in the DDC notes. Some deviations are predefined, others are up to the local library to define (Dewey & Mitchell 2011b, § 12).

Optional numbers do not alter the main DDC-structure and principles, and will not be investigated further in this article.

3.5. Number spans

In addition to classes, DDC have number spans (or centered entries), which may be seen as non-class nodes in the tree/graph structure. Number spans define logical collections within the same parent class. The official reason is to compensate for limitations in the decimal division (Dewey & Mitchell 2011b, § 4.20; Chan et al. 2010, 37). The overall aim is fill up parent classes with 10 subclasses in order to reduce notation length and to keep subjects at "equal level" of specificity at the same DDC level. If there is a choice between two or more otherwise apparently equal numbers, one should choose the number with most nonzero digits from the left. This is called the *Rule of zeros* and it favors the most specific classes from the schedules. Number spans make it possible to add a concept and a level in the hierarchy without adding any digits to the notation.

On the other hand, number spans make browsing tedious and sometimes unnecessarily complex, and some subjects have a large number of them. Taking 200 *Religion* as an example, it takes eight 'clicks' in WebDewey to reach 283 *Anglican churches*, instead of the minimal three. In addition, some of the number spans even overlap at the same level, e.g. 230-270 *Specific elements of Christianity* and 250-280 *Christian church*.

Browsing for 283 in WebDewey (🖱 = mouse clicks):
200 Religion 🖱
230-280 Christianity 🖱
250-280 Christian church 🖱
270-280 History, geographic treatment, biography of Christianity; 🖱
Church history; Christian denominations and sects
280 Christian denominations 🖱
281-289 Specific denomination or sect 🖱
283-289 Protestant and other denominations 🖱

283 Anglican churches

Instead of:

200 Religion
- 210 Philosophy and theory of religion
- 220 Bible
- 230 Christianity
- 240 Christian moral and devotional theology
- 250 Local Christian church and Christian religious orders
- 260 Christian social and ecclesiastical theology
- 270 History, geographic treatment, biography of Christianity
- **280 Denominations and sects of Christian church**
  - 281 Early church and Eastern churches
  - 282 Roman Catholic Church
  - **283 Anglican churches**
  - 284 Protestant denominations of Continental origin and related bodies
  - 285 Presbyterian churches, Reformed churches centered in America, Congregational churches, Puritanism
  - 286 Baptist, Restoration movement, Adventist churches
  - 287 Methodist churches; churches related to Methodism
  - 289 Other denominations & sects
- 290 Other religions

The notation for number spans is worth noticing. When searching for 283 *Anglican churches*, you have to choose the span 270-280. A more user-friendly notation would be 270-289, more like adding form 001-999, not 001-900.

The other reason for number spans is to make a location for hierarchical notes in order to reduce the quantity and to compress printing space. All spans contain notes, some of them numerous and comprehensive (e.g. under 930-990 *History of specific continents, countries, localities; extraterrestrial worlds*). Some number span headings are far from being necessary and user friendly as concepts for browsing, e.g. 270-280 *History, geographic treatment, biography of Christianity; Church history; Christian denominations and sects*. Others are more understandable, such as 230-280 *Christianity*, but probably not indispensable.

The relative importance of structural improvements versus space compressing as a reason for number spans will not be examined further in this article.

4. Representing the DDC structure

As we have seen, the DDC has a structure defined by the schedules, tables and structure defining notes/rules. The main reason for this complex structure definition is to reduce the overall space needed for printing. In order to present this structure in a more user-friendly way, it can be represented in different ways on a computer. We propose and analyze four different representations. Some are exact while others require minor modifications of the DDC structure. However, it is not unlikely that the structure to some degree is dependent on the medium from where it originated. It

is therefore reasonable to presume that a transition to the computer medium could have implications for the DDC structure itself.

The representations are by no means fully analyzed, but are meant to suggest new ideas and opportunities. The objective is to eliminate number building (performed by users), improve search as an alternative to browsing and make the system more user friendly.

4.1. Today's structure

The most obvious representation is of course to retain today's DDC structure with the schedules, the local add tables in the schedules and tables T1-T6. What is new is that the approximately 36,000 classes must be linked together to a much larger extent than in today's WebDewey. Every DDC class must link to its traditional hierarchical subclasses listed in the schedules and to classes referred to in *add to base number* notes if there are any. In addition, a class must link to T1 if there are no reasons not to.

When browsing this representation the users just pick a starting point and click on the most relevant continuations (i.e. links). Both T1 and add to base number classes may be displayed along with all relevant notes. Because of linking, the system software must differentiate between and keep track of the number of the class we have arrived in (e.g. 72**6.64** *Roman Catholic cathedrals*) and the actual path used to get there (e.g. 690.**664** *Construction of Roman Catholic cathedrals*).

This does not necessarily mean that all possible continuations are permitted. The DDC rules must still be followed, e.g. preference order and topics in a standing room. However, this structure is relatively easy to implement. Some rules may be automatically checked and for users it should be easier to exclude disallowed continuations than to find permitted ones.

Improved linking also gives new search options. By searching not only the classes but also the link structure, all relevant built numbers can also be found.
For example, searching *Catholic cathedrals* should find both:

- 726.64 Roman Catholic cathedrals

and built numbers like

- 726.6409421 Roman Catholic cathedrals in London
- 690.664 Construction of Roman Catholic cathedrals
- 246.964 Cathedrals as Christian art

in some relevance order.

For each class, all relevant notes may be displayed independent of structure and representation. There should be no need to collect notes e.g. from parent classes as done today. For built numbers, notes are naturally dependent on their origin, but this is no different than today. For instance, the *Class here* note under 61**2** Human physiology is obviously more adapted to human medicine than as a part of the built class 636.089**2** Physiology--domestic animals, . . .

4.2. Directed graph

So far, we have perceived DDC as a directed tree, but when we include number building we soon realize that the DDC structure is more a directed graph (Bang-Jensen & Gutin 2009; Green & Panzer 2010). The *Add to base number* notes can be

perceived as directed links from one branch of the tree to another branch and even to the root (add from 001-999). This does not mean that the DDC could not easily be transformed into a tree (see representation 3), or perceived as a tree, but the way it appears is as a graph. The classes are the nodes (also called vertices) and from each node there are directed edges (or links) to all successor nodes. Nodes with no inlinks are called sources and nodes with no outlinks are called sinks. The only source in DDC is the class 000. There are probably not many sinks, like 105 *Serial publications of philosophy*. DDC have 10 starting nodes, the main classes 000, 100… 900. Every node (except from the source 000) may have from one up to ≈ 30,000 inlinks[14] and from zero up to 36 outlinks. In addition to schedule and tables nodes, there are nodes from the internal tables, often with modified standard subdivision and facet indicators[15].

One way to number classes in a directed graph like the DDC is to let each node have its significant digit(s), normally from 1-9, but in some cases 01-09, 001-009 or even 0001-0009 for standard subdivisions (e.g., the significant digit for 51**2** *Algebra* is 2). The class number for node N is then the concatenation of all significant digits on the path from the starting node (most significant digit) following the directional links to and including N. The class number for a node is consequently dependent of the starting node and path. The heading of a node is more fixed, like 512 *Algebra*, but the subject it represents is also dependent on the path. The node 796.**33** *Football* may for instance represent itself, a library devoted to football 026.796**33**, *injuries in football* 617.1027**33** or 794.833 *electronic (video) football games*. A longer path with more digits, headings and facets always limits the subject. We can imagine each heading combined with the AND operator, e.g. 512.05 = Science AND Mathematics AND Algebra AND Serial publication. The notation is unique for each path. No pair of different paths may produce the same notation.

A search for a class heading will require more computer work. There are no prebuilt classes. A search for *injuries in football* may find both the football classes at 79**6.33** and injuries classes at 155.916, 344.0218, 571.975, 617 and 63, but no football injuries classes. The DDC software then has to search for connections (or paths) between these classes. The only connecting path is 617.1027**633**, which naturally will be ranked first.

4.3. One prebuilt hierarchical schedule

The simplest and perhaps most comprehensive representation of the DDC is as one single hierarchical schedule containing all classes, including ALL possible buildable classes. Since the DDC structure actually is a digraph, class numbers may be very long and in theory have no limit. In order to represent the system as a finite tree, we therefore have to define some cut-off point (e.g. no more than 30-digit numbers or 3 times adding from 001-999). In practice, such limitations will cause few problems and it should be possible to prebuild "every required" class.

[14] E.g. T1--05 Serial publications that may be added to most other classes in the schedule

[15] E.g. the local table under 592-599 *Specific taxonomic groups of animals*: 1 *General topics of natural history of animals*. This is a node with inlinks from more than 600 directly preceding nodes and at least 5 links out.

DDC already has a few prebuilt classes. There are classes in the schedules that overlap T1 or overlap the result of *Add to base number* notes. For the latter the built classes are merely listed, not hierarchically organized.

The DDC as a single schedule is a very large structure with $\approx 10^9$ classes. If we build 1000 classes a day it will take nearly 3000 years to generate the complete schedule, which makes a manual approach out of the question. However, an automatic generation of classes and headings should be possible, with manual adjustments for the most used classes.

In the current edition of WebDewey most classes neither have a heading nor a class number, e.g. 636.7089699471072 *Research in bone cancer in dogs*[16]. In addition, even if we have a ready-built number (e.g. from the OPAC), there is normally no better heading available than the sum of headings along the path, from 600 *Technology* to T1--072 *Research*. If we keep the most specific heading from each number building step, we end up with *Dogs*, *Veterinary medicine*, *Cancers*, *Bones* and *Research*.

In a prebuilt system, all captions (or headings) can be tailor made to fit the class content, e.g. 636.7089699471072 *Research in bone cancer in dogs*. There is no number building so browsing is easy. For each class, all legal continuations (subclasses), including *Add to base number* continuations and the tables, may be listed, together with relevant notes. Notes can be included as part of each class or be separated and linked to relevant classes. For example, the preference table under 641.5 *Cooking* may be a separate entity linked to/from 641.5 and all subordinated classes.

The most interesting opportunity with this approach, however, is search. With all classes prebuilt, it is perhaps easier to search for a class than to browse. Searching for topics like *bone cancer in dogs* or *construction of cathedrals* will give several hits that may be sorted by relevance before being presented to the user.

4.4. The subtree approach

The DDC structure can also be regarded as built up from of a number of smaller disjoint subtrees. A DDC subtree can be defined as a tree starting from T1 or any other node that is referred to in *Add to base number* notes in connection with number building. A subtree is a part of the DDC that is used at least twice, but represented only once in order to compress a large system to the printed book medium.

In this more theoretical approach, we are interested in the atomic subtrees, which are the minimal building blocks of the DDC structure. One motivation is an improved view of the structure; which subtrees are wise to keep and which are best merged with one or more of the superordinated subtrees. The subtree approach can also be used as a DDC representation when the subtrees are linked together.

4.4.1. The number of subtrees

With more than 2200 *Add to base number* notes there are likely to be many subtrees, but some refer to the same subtree and some refer to number spans with more than one subtree. For example, the 8 classes 005.434, 005.435, 005.436, 005.437, 005.44, 005.713, 006.68 and 006.78 all refer to the 7 subtrees 005.31-005.39. However, the

[16] Inspired by (Satija 2007, 135)

parent class 005.3 is not the root of a subtree, since there are no references to it. The number of subtrees is probably around 2000, but this is not investigated further.

4.4.2. Numbering of subtree classes

The classes can be numbered and named in different ways. In the current edition some subtrees belong more to one context than to others, e.g. 612 *Human physiology* that obviously has a focus on humans. At the same time, it is also a subtree connected to 636.089 *Veterinary medicine*. By keeping the current notation and heading, we emphasize the relation to medicine, but it is less suitable as a continuation for veterinary medicine. Some subtrees, like 371.1-371.8 *Schools and their activities*, seem to belong equally to all superordinated classes (or subtrees). In this case, we could use only the significant digit as class number. I will leave the question of whether it is necessary to have context-dependent subtrees in a computer environment open.

4.4.3. Number building

Classification (or number building) is the process of traversing the subtrees, starting from the root(s) of what is left of the original schedules and moving down[17] the hierarchy. From some subtree classes (often leafs), it is possible to jump (using a link) to the root of other *Add to base number* subtrees and continue down the directed paths. This continues until we hopefully reach the most relevant class, a sink or we are stopped by rules that do not allow us to go any further.

4.4.4. Examples of atomic subtrees

Each of the classes:

004.3**3***Real-time processing
004.3**5***Multiprocessing
004.3**6***Distributed processing

are one-class subtrees, only referred to from 004.618 **Computers distinguished by processing modes*.

005.36**8** Programs for specific operating systems and for specific user interfaces
005.36**82** Programs for specific operating systems
005.36**84** Programs for specific user interfaces

This is a small subtree referred to from the classes 005.34, 005.32 and 005.376.
A larger subtree is 616.1 **Diseases of cardiovascular system* with ≈32 classes. The subtree is referred to from the classes 614.59, 618.92, 618.976

4.5. Summary

The DDC has a structure that can be represented by a computer (and presented) in different ways. Every class has its set address, uniquely expressed by the class numbers and often by the headings, even if it sometimes is "utilized" in several contexts. Is the division of human medicine satisfactory for animals? If not, the alternative is to make *Human medicine* more general or to let *Veterinary medicine*

[17] A computer tree-structure is normally drawn upside down, see http://en.wikipedia.org/wiki/Tree_(data_structure)

have its own divisions. The DDC seems to have an unfounded fear of redundancy and a continuing desire to compress. For example, is the reduction of the three classes 004.33-004.36 worth the extra challenge of number building? Perhaps more of the DDC should be prebuilt when space now is unlimited.

An efficient representation is however not better than its source. Except for the third one, the suggested representations are probably isomorphic, i.e. structurally identical to each other (and to the DDC). The differences are more on a conceptual level. The class numbers and/or the headings can be more or less generalized (i.e. not exact representations), depending on how well we want them to fit the different contexts. For instance, by using the significant digit as class number, we remove the belonging to a certain context (or subtree) represented by the number. To recommend a representation will be qualified guessing. Maybe a combination of the first and third representations would be a good option.

Search is likely to have extensive opportunities, which in most cases may perform better than browsing alone. With 10^9 DDC classes, a representation must support relevance-sorted search, both for single classes and classes related through links. A combination is also interesting to investigate further since the search results in some cases would be a graph.

5. The DDC without number building

The four DDC representations we have analyzed could all be used to present (and search) the DDC, without number building and with all relevant notes attached to each class. It is up to the presentation software to find all relevant classes and notes and present them in a user-friendly way.

We will here propose some ideas on how classes can be presented on a computer without number building. Our example subject is the complex subject *Research in bone cancer in dogs*.

5.1. Number building in the current edition of WebDewey:

Build notation	**Class visited**	**Class heading**	**Relevant notes (or action)**
600	600	Technology	Starting from the Schedule
630	630	Agriculture	
636	636	Animal husbandry	
	636.1-636.8	Specific kinds of domestic animals	
636.7	636.7	Dogs	
	636.701-636.708	Specific topics in husbandry of dogs	Add to base number 636.70 the numbers following 636.0 in 636.01-636.08 **(i.e. jump back to 636)**

	636	Animal husbandry	
636.70**8**	636.0**8**	Specific topics in animal husbandry	
636.70**89**	636.0**89**	Veterinary medicine	Add to base number 636.089 the numbers following 61 in 610-618 **(i.e. jump to 610)**
	610	Medicine & health	
636.7089**6**	61**6**	Diseases	
	616.1-616.9	Specific diseases	
636.708**969**	61**6.9**	Other diseases	
636.7089**699**	61**6.99**	Tumors and miscellaneous communicable diseases	
636.7089**6994**	61**6.994**	†Cancers	
	616.9942-616.9949	Cancers of other organs and of regions	Add to base number 616.994 the numbers following 611 in 611.2-611.9 **(i.e. jump to 611)**
	611	Human anatomy, cytology, histology	
	611.1-611.9	Gross anatomy	
636.708969947	611.**7**	Musculoskeletal system, integument	
636.7089699**471**	611.**71**	Bones	**(Jump to T1--0 standard subdivisions)**
636.708969947**107**	T1--**07**	Education, research, related topics	
636.7089699471**072**	T1--**072**	Research	

5.2. Without number building

Without number building, each class is presented with all possible (and hopefully permitted) subclasses and all relevant notes. We only show how it may appear prebuilt for the classes that involve number building (in blue). Number spans are also eliminated.

636.7 Dogs (1st number building)

636.7001 Philosophy and theory
636.7002 Miscellany
: :
636.7009 Historical, geographic, persons treatment

636.701 Ranches and farms
636.707 Puppies (Young of animals)
636.708 Specific topics in animal husbandry (🖱 our choice)
636.71 Breeds of dogs
636.72 Nonsporting dogs
636.73 Working and herding dogs
636.75 Sporting dogs, hounds, terriers
636.76 Toy dogs

636.7089 Dogs -- Veterinary medicine (2nd number building)
…01 Philosophy and theory
…02 Miscellany
: :
…09 Historical, geographic, persons treatment
…1 Human (or actually dog) anatomy, cytology, histology
…2 Human (dog) physiology
…3 Personal (dog) health and safety
…4 Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine
…5 Pharmacology and therapeutics
…6 Diseases (🖱 our choice)
…7 Surgery, regional medicine, dentistry, ophthalmology, otology, audiology
…8 Gynecology, obstetrics, pediatrics, geriatrics

For Veterinary medicine, the source standard subdivision it is not obvious. We can choose between T1 and 610.

The next one is not easy for the average user to comprehend. Here we have a † reference to more than two pages of add notes under 618.1-618.8 (for adding 001-009 and 01-09). In addition, we have add notes under 616.994**2**-616.994**9** (for adding 2-9).

636.708966994 Dogs -- Veterinary medicine -- †Cancers (3rd number building)
…001 Philosophy and theory
…002 Miscellany
: :
…009 History, geographic treatment, biography
…01 Medical microbiology
…02 Special topics
…03 Rehabilitation
…04 Special classes of diseases
…05 Preventive measures and surgery
…06 Therapy
…07 Pathology
…08 Psychosomatic medicine
…09 Case histories
…1 †Cancers of cardiovascular organs and blood
…2 Respiratory tract diseases

…3 Digestive system diseases
…4 Endocrine diseases
…6 Genital diseases
…7 Musculoskeletal diseases (🖱 our choice)
…8 Nervous system diseases
…9 Regional and topographical anatomy

To add from T1 the user must be made aware of the preference table under T1, topics in a standing room, redundancy, etc. This however is not relevant in our example, and we can safely add from T1.

636.70896699471 Dogs -- Veterinary medicine -- Bone cancer
(4th number building)
…01 Philosophy and theory
…02 Miscellany
…03 Dictionaries, encyclopedias, concordances
…04 Special topics
…05 Serial publications
…06 Organizations and management
…07 Education, research, related topics (🖱 our choice)
…08 History and description with respect to kinds of persons
…09 Historical, geographic, persons treatment

We have now reached our DDC number without any *Add to base number* notes and number building.

6. Conclusion and future work

Even after nearly 20 years in electronic form, WebDewey is largely a copy of the printed DDC. WebDewey is certainly hypermediated with multiple windows, clickable links and hierarchical browsing, but the underlying structure with number building and sparse notes is a construction tailor made for the printed book medium. Reduced redundancy is of course still an important principle (e.g. in database theory), but the structure is ready for revision and at least end users should be spared the decompression. A number building free DDC may increase immediacy and make it more user-friendly. Classification is in any case difficult for most people.

In this article, we have examined the Dewey structure and four different representations for the computer medium. We have also suggested ideas for a new interface without number building. The research is only in its beginning and more work needs to be done in this field. The actual redundancy of DDC subtrees and the underlying representation of the DDC itself are interesting topics to examine further. The same is an actual test of a number building free user interface that includes search for all classes, not only the 50,000 from the printed edition, but also the 1,000,000,000 buildable classes.

7. References

Bang-Jensen, J., & Gutin, G. Z., *Digraphs Theory, Algorithms and Applications*. London: Springer-Verlag 2009

Bolter, J. D., & Grusin, R., *Remediation: Understanding New Media*. Cambridge, Mass.: MIT Press 1999

Chan, L. M., Mitchell, J. S., & Kungl. biblioteket, *Dewey decimalklassifikation/principer och tillämpning*. Stockholm: National Library of Sweden 2010

Crawford, L., "Dewey Products & Services" Retrieved 15. april, 2012, from http://www.oclc.org/multimedia/2012/files/pla-dewey-products-services.pdf

Dewey, M., & Mitchell, J. S., *Dewey Decimal Classification and Relative Index* (23rd ed.). Dublin, Ohio: OCLC Online Computer Library Center 2011a

Dewey, M., & Mitchell, J. S. (2011b), "Introduction to Dewey Decimal Classification" in: ed. J. S. Mitchell, *Dewey Decimal Classification and Relative Index*. Dublin, Ohio: OCLC Online Computer Library Center 2011b, xliii-lxxi. Retrieved from http://www.oclc.org/dewey/versions/print/intro.pdf

Green, R., & Panzer, M., "The ontological character of classes in the Dewey Decimal Classification" paper presented at the *Eleventh International ISKO Conference*, Rome, 23-26 February 2010

Mengel, T., Alex, H., & Junger, U., "Future Requirements in Respect of WebDewey Functionality" Deutsche Nationalbibliothek 2011. Retrieved from www.slainte.org.uk/edugit/docs/**EPC134-23.4.1**.pdf

Michell, J. S., "Challenges Facing Classification Systems: A Dewey Case Study" Paper presented at the *Knowledge Organisation for Information Retrieval Sixth International Study Conference on Classification Research*, University College London 1997

Miksa, F. L., *The DDC, the Universe of Knowledge, and the Post-Modern Library*. Albany, N.Y.: Forest Press 1998

Mitchell, J. S., "Relationships in the Dewey Decimal Classification System" in: ed. C.A.B.R. Green, *Relationships in the Organization of Knowledge*. Dordrecht, NL: Kluwer 2001, 211-226

Mitchell, J. S., & Vizine-Goetz, D., "Dewey Decimal Classification (DDC)" in: eds. M. J. Bates & M. N. Maack, *Encyclopedia of Library and Information Science*. Boca Raton, Fla.: CRC Press 2009

OCLC, "DDC Translations Retrieved" March 22, 2012. Retrieved from http://www.oclc. org/dewey/about/translations/default.htm

Satija, M. P., *The Theory and Practice of the Dewey Decimal Classification System*. Oxford: Chandos Publishing 2007

Scott, M. L., *Dewey Decimal Classification: A Study Manual and Number Building Guide*, Westport, Conn.: Libraries Unlimited 2005

Svanberg, M. (2011), "EPC 134-23.4.2 WebDewey 2.0 Swedish experiences" Paper presented at the *European DDC Users Group Technical Issues Working Group*, Stockholm, Sweden. Retrieved from http://www.slainte.org.uk/edugit/docs/EPC-134-23.4.2.pdf

Taylor, A. G., "Teaching the Dewey Decimal Classification System" *Cataloging & Classification Quarterly*, 42(3-4), 2006, 97-117

Trotter, R., "Electronic Dewey: The CD-ROM Version of the Dewey Decimal Classification" *Cataloging and Classification Quarterly* 19(3/4), 1995, 213-34

Biographical note

Tore Brattli is Associate Professor of Documentation studies at the University of Tromsø, Norway. His research interests include databases, information retrieval and digital representation. E-mail: tore.brattli@uit.no

Summary

Correct Dewey classification is demanding and time consuming. Many of the challenges with the Dewey system are related to locating and interpreting notes (i.e. classification guidelines), and number building.

Today's Dewey structure is a result of more than 100 years of optimizing a comprehensive classification system to the printed book medium. In order to limit the system into a "manageable" size, facets and facet-like subjects are represented only once and instead referred to from relevant classes for number building. A similar technique is used to reduce the number of notes.

With the remediation of Dewey from printed to computer media, space is not limited and there is no need to compress the classification system. Number building can be eliminated, and all relevant notes attached to each class. Despite the fact that the system now has been available in electronic form for almost 20 years, it is still largely a copy of the printed version.

This article first investigates how the Dewey system may be presented for users without number building, in order to make it more immediate and user-friendly. We first analyze the Dewey structure, and then look at different representations of the structure suited for computer media. Finally, some ideas for a new presentation without number building are proposed.

Key words

Dewey Decimal Classification, WebDewey, Number building, Remediation, Data compression, Directed graph, Facet representation

