A Steven Pollitt
View-Based Systems Limited, Armitage Bridge, Huddersfield, UK
Amanda J Tinker
Centre for Database Access Research, The University of Huddersfield, UK

Enhanced View-Based Searching through the Decomposition of Dewey Decimal Classification Codes

Abstract: The scatter of items dealing with similar concepts through the physical library is a consequence of a classification process that produces a single notation to enable relative location. Compromises must be made to place an item where it is most appropriate for a given user community. No such compromise is needed with a digital library where the item can be considered to occupy a very large number of relative locations, as befits the needs of the user. Interfaces to these digital libraries can reuse the knowledge structures of their physical counterparts yet still address the problem of scatter. View-based searching is an approach that takes advantage of the knowledge structures but addresses the problem of scatter by applying a facetted approach to information retrieval. This paper describes the most recent developments in the implementation of a view-based searching system for a University Library OPAC. The user interface exploits the knowledge structures in the Dewey Decimal Classification Scheme (DDC) in navigable views with implicit Boolean searching. DDC classifies multifaceted items by building a single relative code from components. These codes may already have been combined in the schedules or be built according to well-documented instructions. Rules can be applied to decode these numbers to provide codes for each additional facet. To enhance the retrieval power of the view-based searching system, multiple facet codes are being extracted through decomposition from single Dewey Class Codes. This paper presents the results of applying automatic decomposition in respect of Geographic Area and the creation of a view (by Geographic Area) for the full collection of over 250,000 library items. This is the first step in demonstrating how the problem of scatter of subject matter across the disciplines of the Dewey Decimal Classification and the physical library collection can be addressed through the use of facets and view-based searching.

1. Introduction

Conceived pragmatically by Melvil Dewey in 1876 as a means of arranging physical items on library shelves, the Dewey Decimal Classification Scheme is now an available tool for electronic knowledge organisation and information retrieval. Work to extend the scope of Dewey to incorporate both physical and electronic domains began when it was produced in machine-readable form, a century after the scheme's inception. Ten years later it was utilised for the first implementation of a classified approach to subject searching in a library OPAC (Markey & Demeyer, 1986). The emergence of the Internet probably provided the greatest impetus for Dewey to extend its remit to electronic contexts. The last few years have seen how Dewey, and the notion of classification in general, is being rediscovered as a viable alternative to keyword search engines for a more selective and quality-controlled approach to Internet resource discovery.

Internet Subject directories utilising the DDC range from those which transplant the Dewey hierarchies and notation without adaptation in their paper-based format (e.g. BUBL Link/5.15), those incorporating slight adaptations (e.g. Canadian Information by Subject), to OCLCs NetFirst database which has incorporated multiple Dewey codes for items and recast the Dewey summaries to make them more appropriate for end-user searching (Vizine-Goetz, 1998).

In an earlier paper (Tinker et al, 1999) it was asserted that for optimal subject retrieval the DDC should forgo its physical, shelf-motivated aspects, such as a single relative location, and be applied in a way that takes full advantage of the electronic domain. It was suggested that multiple Dewey codes and, in turn, additional facets could be derived by decomposition

to identify frequently used notations, applicable across subject disciplines. This will allow greater specificity in subject searching and collocate concepts that might otherwise be distributed across the Dewey Schedules.

This paper describes the decomposition process as part of the development of a view-based searching OPAC at the University of Huddersfield. Details of 255,212 titles held in the library were downloaded into a Microsoft Access relational database. Elements of the Dewey Decimal Classification Schedules and Tables were stored in the same database and programs and queries developed to progress the creation of facets and views.

2. Decomposition of Synthesised Dewey Notations

The DDC is essentially an enumerative classification but, particularly since publication of the 18th edition with its extension to seven auxiliary Tables, has made increasing use of facets and synthesis as advocated by Ranganathan (1965). The tools for the classifier comprise the Schedules; Auxiliary Tables for number building, denoting recurrent bibliographic forms and common subjects (Table 1, Standard Subdivisions) and subject facets such as geographic location, historical period and language (Tables 2-7); the relative index, with synonyms for collocating topics scattered across the Schedules; and lastly, a Manual for explication and clarification.

A Dewey classification code can be assigned to an item using a ready-made notation in the schedules. This notation may already signify more than one facet (e.g. 032 General Encylopedic Works in English, 061.1 General Organisations in Canada). Additional facets can be added by re-using schedule code components or notations from the seven auxiliary tables. This notational synthesis is applied either under instruction in the schedules or without instruction using the first auxiliary table, the Standard Subdivisions. The ability to identify individual facets from the synthesised codes opens up significant possibilities for enhanced information retrieval (Liu 1996, Miksa 1998, Mitchell 1997).

The task of decomposition of Dewey Decimal Class Codes is not trivial. Wajenberg (1983) recommended enhancing MARC coding with the constituent components of the Dewey Code for an item "that will consistently encode the hierarchy of the numbers, and consistently identify and analyze synthesised numbers". This approach would preserve the original components before synthesis and obviate the need for decomposition. In the absence of such an enhancement there is no option other than to apply decomposition techniques.

Liu (1996) describes a rule-based approach for the 700 class and the Tables. The report is most encouraging, producing 100% success (excluding incorrect Dewey Codes) in a test with 6,000 codes. The system built in the study, DND - Dewey Number Decomposer, does not appear to have been extended. Decomposition efforts are not restricted to DDC; Gerhard Riesthuis (1997) has done work on decomposition of complex UDC notations.

Given the need to cover the complete Dewey schedules, and mindful of the complexity of the instructions on synthesis, we focussed on the decomposition notations that used the auxiliary tables, namely Table 1 (Standard Subdivisions) and Table 2 (Historical, geographic, persons treatment). The following sections review the analysis to produce the Geographic Area Facet.

3. Extracting the Geographic Area Facet

The three ways in which Geographic Area is signified in Dewey:

- When the geographic area is the subject of study and is therefore allocated a caption and notation in the Schedules without needing to consult the Area Table (Table 2); for example, in a few limited cases such as the geographic treatment of philosophy (Chan et al, 1996, p.111).
- From Table 2, *with instruction* in the Schedules to 'add to base number'.

- From Table 2, *without instruction* in the Schedules via Table 1 (Standard Subdivisions) -09 notation (Historical, geographic, persons treatment).

T2--0	Table 2. Geographic Areas, Historical Periods, Persons
T2--1	Areas, regions, places in general
T2--2	Persons
T2--3-T2--9	Specific continents, countries, localities; extraterrestrial worlds
T2--3	The ancient world
T2--4-T2--9	The modern world; extraterrestrial worlds
T2--4	Europe Western Europe
T2--5	Asia Orient Far East
T2--6	Africa
T2--7	North America
T2--8	South America
T2--9	Other parts of world and extraterrestrial worlds Pacific Ocean islands

We focussed our decomposition efforts on notations T2--4 to T2--9 (The modern world, extraterrestrial world). The Geography facet currently extends only to the first 3 levels with the exception of British Isles which has the full hierarchy. Examples of the loss of detail are:

Included:		Excluded:	
T2--71	Canada	T2--7135	Lake Ontario region
T2--713	Ontario	T2--71354	Metropolitan Toronto and Regional Municipality of York

This detail will be introduced in the near future given automated procedures for generating views from the Tables.

3.1. Geography already in the Schedule Summaries

Geography appears in the schedule summaries under the following Dewey Classes: 060 General organizations & museology; 070 News media, journalism, publishing; 190 Modern western philosophy; 270 History of Christianity & Christian Church; 310 Collections of general statistics; 550 Earth Sciences; 910 Geography and travel; 930-990 General history. The decomposition of codes is described for the three sets: -

- 191-9 Philosophy of Countries
- 914-919 Geography of Places
- 94-99 General History of Places

3.1.1 Philosophy of Countries.

190 Modern western philosophy
191 Philosophy of United States & Canada
192 Philosophy of British Isles
193 Philosophy of Germany & Austria etc..

Some 732 items were identified in 191-198; all had valid Table 2 codes found by removing the 19 in the notation and replacing it with 4 for 192-198 and with 7 for 191. The following are examples of titles retrieved and the associated codes:

Dewey Code	**Title**	**Table2 Code**	**Caption**
191	The philosophy of John Dewey	T2--**7**	North America
196.1	The tragic sense of life in men and nations	T2--**461**	Northwestern Spain
198.9	Kierkegaard	T2--**489**	Denmark and Finland

3.1.2 General History of Places

Geography appears in 940-990 and the coding can be matched with that in Table 2 by removing the 9 and matching the remaining number.

940 General history of Europe
950 General history of Asia Far East
960 General history of Africa
970 General history of North America
980 General history of South America
990 General history of other areas

A simple search (Dewey Code like 94* or 95*.. or 99*) retrieved 10,053 items all of which provided valid numbers from Table 2. The items include:

Dewey Code	Title	Table2 Code	Caption
956.704	International perspectives on the Gulf conflict	T--567	Iraq
966.7030924	The autobiography of Kwame Nkrumah	T2--667	Ghana
971.71042	The Great depression	T2--7171	Prince County (PEI)

3.1.3 Geography of Places

Geography of places appears in 914-919 section, the coding can be matched with Table 2 by removing the 91 and matching the remaining number.

913 Geography of & travel in ancient world
914 Geography of & travel in Europe
915 Geography of & travel in Asia
916 Geography of & travel in Africa
917 Geography of & travel in North America
918 Geography of & travel in South America
919 Geography of & travel in other areas

A simple search (Dewey Code like 914* or 915*.. or 919*) retrieved 877 items all of which provided valid numbers from Table 2. The items include:

Dewey Code	Title	Table2 Code	Caption
914.2195	The gardens and parks at Hampton Court	T2--42195	Richmond upon Thames London
914.9503	The Byzantines and their world	T2--495	Greece
916.76	East Africa : a travel survival kit	T2--676	Uganda and Kenya

3.2 Adding from Standard Subdivisions

Geographic notations added to Schedule numbers without instruction are the easiest to identify given the Standard Subdivisions facet indicator 0 followed by 9. The relevant area code is then added to this --09 notation from Table 2:

A search of the library collection to find titles with the geographic notation either directly or indirectly following the decimal point 094-099 (including variations of the standard subdivisions 009, 0009 etc.) resulted in the retrieval of 11,324 titles. There were only 37 titles where the code did not map to a Table 2 notation leaving 11,287 titles including:

Dewey Code	Title	Table2 Code	Caption
069.0941	Museums and the shaping of knowledge	T2--41	British Isles
133.43097445	A delusion of Satan : the full story of the Salem witch trials	T2--7445	Essex County (Massachusetts)

3.3 Add to Base Number from Instructions in Schedules

The remaining geographic notations are those formed by an 'add to base number from Table 2' instruction in the Schedules. Such instances were located using a search on Dewey

for Windows to find mention of 'Table 2' which led to the creation of a file of numbers for which these add instructions apply. In the future a more precise set of codes should be established using a relational Dewey database. Nevertheless, this process demonstrates the principle. This file of base numbers was subsequently matched against Dewey notations in the library database.

This resulted in 409 Dewey Codes used to search the 255,212 library items. Two hundred and thirty one of the 409 codes found 28,189 items. Twenty-one of these 231 codes did not have any added numbers. Searching using the remaining 210 codes found 28,136 items, 2,624 did not have added numbers. Matching examples from Dewey 320.4 Structure and functions of government include:

Dewey Code	Title	Table2 Code	Caption
320.4**45**	Costituzione italiana	T2--**45**	Italian Peninsula
320.4**667**	Ghana in transition	T2--**667**	Ghana
320.4**71**	Government in Canada	T2—**71**	Canada

3.4 Viewing the Geography Facet

Items with a geography facet are scattered across the classification as displayed in Table 1. These items can now be viewed via the "by Geographic Area" view as shown in Figs 1 and 2.

Dewey Class	Embedded in the Schedules	From standard subdivisions	From Table 2 with instruction	Total
000	18	153	59	**212**
100	732	42	23	**797**
200		47	113	**160**
300		7,823	7,437	**15,260**
400		7	39	**46**
500		129	60	**189**
600		695	309	**1,004**
700		2,207	1,178	**3,385**
800		81		**81**
900	10,930	103	1346	**12,379**
Total	**11,680**	**11,287**	**10,564**	**33,531**

Table 1. Sources of items for the Geographic Area Facet

4 Conclusion

Work is continuing to construct views of facets to retrieve library items via Type of Material (Books, Videos, Sound Recording etc.), Types of Book (Dictionaries & Encyclopaedia, Bibliographies, Biographies and Directories), Language, Historical Periods, Persons, Management, Research, Year of Publication and others. The object-oriented approach will also lead to views identifying Authors, Publishers and Borrowers. An evaluation is planned for later this year.

Acknowledgements

We are pleased to acknowledge the assistance and encouragement of Joan Mitchell, Editor of the Dewey Decimal Classification and President of OCLC Forest Press.

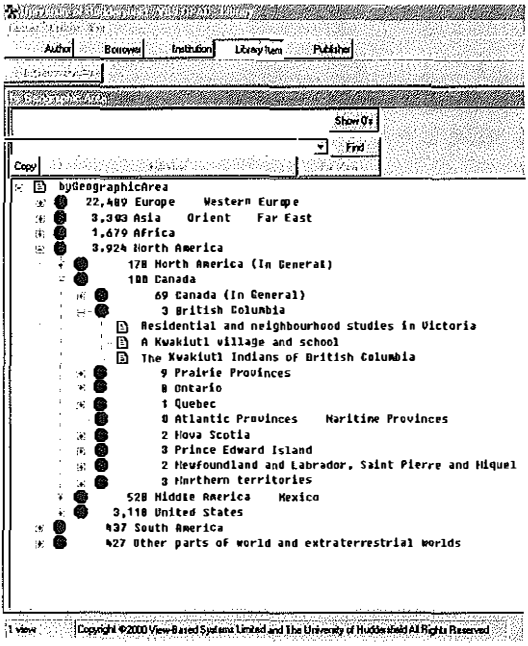

Figure 1. The Top Level "by Geographic Area" View of Library Items with Expansions

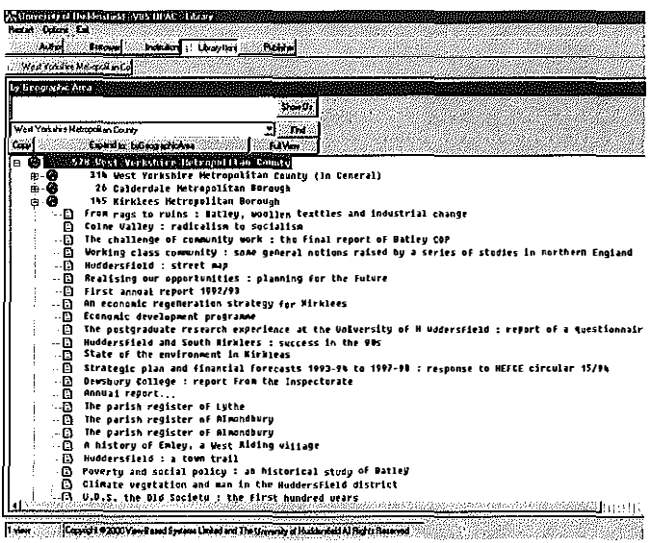

Figure 2. "by Geographic Area" view of Library Items after a direct search for West Yorkshire

References

BUBL Link/5:15 Catalogue of Internet Resources [http://www.bubl.ac.uk/link/ddc.html]. Accessed 13 March 2000.

Chan, L.M., Comaromi, J.P., Mitchell, J.S. & Satija, M.P. (1996) *Dewey Decimal Classification: a practical guide.* Albany, NY: Forest Press.

Liu S (1996) Decomposing DDC Synthesised *Numbers 62nd IFLA General Conference Bejing , China* (http://ifla.org/IV/ifla62/62-sonl.htm)

Markey, K & Demeyer, A.N. (1986) *Dewey Decimal Classification online project: evaluation of a library schedule and index integrated into the subject capabilities of an online catalog. Final report to the Council of Library Resources, report number OCLC/OPR/RR-86/1.* Dublin, OH: OCLC.

Miksa, F.L. (1998) *The DDC, the universe of knowledge and the post-modern library.* Albany, NY: OCLC Forest Press.

Mitchell, J.S. (1997) Challenges facing classification system: a Dewey case study, In *Knowledge Organisation for Information Retrieval, proceedings of the sixth international study conference on classification research*, University College London 16-18 June 1997, The Hague: FID, 85-89.

National Library of Canada. *Canadian Information by Subject.* [http://www.oclc.org/oclc/netfirst/]. Accessed 13 March 2000.

OCLC *Dewey Decimal Classification Research Agenda* [http://www.oclc.org/oclc/fp/research/agenda.htm]. Accessed March 2000

OCLC *NetFirst* [http://www.oclc.org/oclc/netfirst/netfirst.htm]. Accessed 13 March 2000.

Ranganathan, S.R.(1965) *A descriptive account of the Colon Classification.* Bangalore, Sarada Ranganathan Endowment for Library Science (reprinted 1990).

Tinker, A.J., Pollitt, A.S., O'Brien, A. & Braekevelt, P.A. (1999) The Dewey Decimal Classification and the transition from physical to electronic knowledge organisation. *Knowledge Organization*, 26(2): 80-96.

Vizine-Goetz, D. (1998) Dewey as an Internet subject guide. In: W Mustafa el Hadi, J Maniez & A S Pollitt (eds.), *Stuctures and relations in knowledge organisation. Proceedings of the fifth international ISKO conference 25-29 August 1998, Lille, France.* Wüzburg: Ergon Verlag, 191-197.

Dewey decimal classification. Centennial 1876-1976. Facsimile reprinted by Forest Press Division Lake Placid Educational Foundation.Â The plan of the following Classification and Index was developed early in 1873. It was the result of several months' study of library economy as found in some hundreds of books and pamphlets, and in over fifty personal visits to various American libraries. In this study, the author became convinced that the usefulness of these libraries might be greatly increased without additional expenditure.