The Dewey Decimal Classification and the Transition from Physical to Electronic Knowledge Organisation*

Amanda J. Tinker[α], A. Steven Pollitt[αχ], Ann O'Brien[β] and Patrick A. Braekevelt[χ]

[α]Centre for Database Access Research, School of Computing and Mathematics, The University of Huddersfield, UK

[β]Department of Information Science, Loughborough University, UK

[χ]View-Based Systems Limited, Yorkshire Technology Park, Armitage Bridge, Huddersfield, UK.

Amanda Tinker is working towards a PhD as a Research Assistant in the Centre for Database Access Research (CeDAR) at the University of Huddersfield. She graduated from the University of York in 1995 with a BA in English Language and Linguistics. She worked as a Graduate Trainee at the Bodleian Library, Oxford, before studying for an MSc. in Information and Library Studies at Loughborough University.

Steven Pollitt is Reader in Information Science in the School of Computing and Mathematics at the University of Huddersfield and Technical Director of View-Based Systems Limited. He has a first degree in Computer Science and was awarded a PhD for research into An Expert Systems Approach to Document Retrieval. He has directed a number of research projects in the Centre for Database Access Research at Huddersfield University, two of which were funded by the British Library: Matching OPAC User Interfaces to User Needs and HIBROWSE for Bibliographic Databases, leading to the current developments in view-based searching.

Ann O'Brien is a lecturer in the Department of Information Science, Loughborough University. Her teaching and research interests are information organisation and retrieval, more particularly, subject analysis, indexing and classification.

Patrick Braekevelt is a Senior Programmer and Systems Administrator at View-Based Systems Ltd. and has a first degree in Chemistry. He has City and Guilds qualifications as a programmer and has been working on the software interface development of several research projects in the Centre for Database Access Research at Huddersfield University.

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. 31 refs.

ABSTRACT: The physical organisation of items on library shelves using any classification scheme is inevitably a compromise. The best efforts to achieve an arrangement that is helpful to users will be thwarted by the multifaceted nature of these items and the specific needs of the user and the library. Items on a particular subject will be scattered throughout the library building(s) across disciplines, by physical form, by frequency of use and whether and for how long they may be borrowed. Even though the rich information content of multifaceted items may be represented in the notation, the items required by a user will be scattered across library shelves when the item is placed in a single relative location. This paper explores these issues using examples from a University Library classified using the Dewey Decimal Classification Scheme (DDC). The electronic context of the library OPAC can transcend the constraints imposed by the predominantly physical nature of library collections, yet the current use of classification schemes in on-line systems retains many of these limitations. Examples of such systems applying DDC on the World Wide Web are discussed and compared with a system

that seeks to use DDC in what is called view-based searching. The interface and the resulting browsing and searching capability of a View-Based Online Public Access Catalogue (OPAC) are described. Ways in which subject access to library collections can be improved and disciplinary scatter resolved by assigning multiple class numbers to items and exploiting the rich Dewey structure in a faceted form are discussed. It is suggested that the informative power of visual classificatory structures at the search interface will be beneficial to the broader learning experience of the user. The paper concludes that the application of classification schemes in electronic interfaces should not be bound by the physical constraints that no longer apply in an electronic context but be exploited to provide a complete, flexible and individual interface as determined by the needs of each user.

1. Introduction

> "The objective of maximum use of [library] stock ... can be assisted in two ways:
> (a) by personal assistance to readers
> (b) by so arranging and displaying the collection that its resources are made as far as possible self-evident to the user. This will include the provision of catalogues and indexes... "
> Jack Mills (1960), The problem of arrangement in a library,
> A modern outline of library classification, Chapter 1, p.1.

The physical arrangement of library collections is still today inherently problematic. These problems stem from the limitation that permits only one 'best sequence' to accommodate an ever-increasing growth and diversification of knowledge and, in turn, meet users' multifarious needs.

Prior to 1876, books were allocated a subject location by way of labels on shelves and then arranged in acquisition order. It was Melvil Dewey who freed libraries and their users from the shackles of shelf labelling by assigning a relative location to an *item*, using a decimal notation, the Dewey Decimal Classification (DDC), that allowed the item to move around the library together with its location code. This revolutionary idea of subject coding individual items according to ten hierarchical categories, rather than locating the item on a designated shelf location, also provided for greater specificity in subject description.

The revolution of information technology enables system developers to free the library user from the physical constraints of the library shelves via the Online Public Access Catalogue (OPAC). Classification schemes have a significant role to play in these electronic systems. Such initiatives involving DDC have been the subject of research and development for some years by researchers such as Karen Markey (1989). With the advent of the Internet, classification schemes are now being exploited in new ways for online searching and browsing.

This paper describes the problems of physical arrangement when using the DDC in a university library. Examples of using DDC on the Web are presented to show how these problems are being revisited when DDC is applied to electronic information collections. A discussion of the yet unrealised/untapped value of classificatory structures in current commercial OPAC interfaces precedes a description of a view-based searching OPAC, seeking to improve the quality of subject access, freed from the constraints of physical arrangement.

2. DDC and the physical library context

First published in 1876 and now in its 21st edition, the Dewey Decimal Classification Scheme is perhaps the best known and most widely used classification in existence.

Already a de facto international standard through its adoption across the world, this status has been officially recognised since its inclusion in MARC records by the Library of Congress and other national bibliographies in January 1980 and the BNB a year later (Foskett, 1982, p.339). Although the DDC is traditionally associated with public libraries it has a growing presence in academic and national libraries, notably the open shelf collections of the Bibliothèque Nationale in Paris and the arts and humanities sections of the new British Library at St. Pancras.

The scheme organises universal knowledge systematically and hierarchically into ten main categories, each followed by ten divisions and then ten sections. 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.

Examples showing how DDC is applied in the Library at Huddersfield University are presented to highlight several potential difficulties experienced by users when subject searching or browsing at the physical 'coalface' of the library shelves. These examples also demonstrate the richness inherent in Dewey

knowledge structures that is compromised by the necessity of using a single relative location, and show instances of information loss that result from DDC rules necessitating choice between subject facets. The examples are taken from the main central library and relate to the subject of textiles, taught in the university and once the core industry of the region.

2.1 Disciplinary basis and resultant scatter

Like the majority of classification schemes, the DDC is structured by discipline, the consequence being "that there is likely to be no single place for a given subject. A subject may appear in any discipline" (Forest Press, 1998). This notational scatter results in items relating to the same subject being distributed throughout the library building. The Relative Index is Dewey's means of finding the location of subjects within the Schedules. For example, the following illustrates a search of this index on Dewey for Windows (the electronic 21st edition) for the subject of 'Textiles':

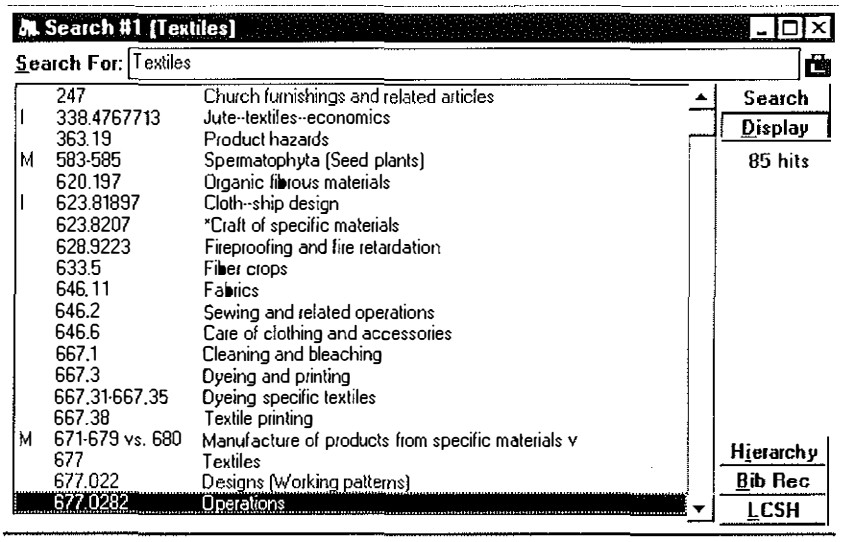

Figure 1: Search for 'Textiles' on Dewey for Windows

'Textiles' is principally located within the discipline of Technology (677, Manufacturing - Textiles). However, as illustrated by the 85 hits in Figure 1, the subject of textiles is also to be found within other disciplines, and hence Dewey Main Classes, from Religion (247, Church Furnishings) to Fine and Decorative Arts (746, Textile Arts). Such scatter in knowledge organisation and decimal notation is then mirrored in the physical library arrangement.

The searchable OPAC can also be used to study this scattering. A subject keyword search was performed on the University of Huddersfield library OPAC to locate 'textile' items. Figure 2 reveals how textile-related items can be found outside the main location for textiles (677) and the extent to which they are physically disparate within the library itself.

The major 'textile' category, situated on Level 3, constitutes the best shelf-browsing territory, and some users may not look beyond this. However, users will also find possibly relevant items on this same floor in clothing, textile arts, architecture (e.g. textile mills), chemical engineering (e.g. dyeing, bleaching, printing of textiles), mathematics (statistics for those in the textile industry), and even sanitary engineering (e.g. pollution control in textile finishing industries). Further textile items relating to the economics and management of the textile industry are available after climbing six flights of stairs to Level 6.

Dewey Code	Caption	Library Floor Location
330	Economics	Level 6
380	Commerce	Level 6
519	Mathematics (Statistics)	Level 3
628	Sanitary engineering	Level 3
658	Management	Level 6
660	Chemical engineering	Level 3
677	**Textiles**	**Level 3**
687	Clothing	Level 3
725	Architecture	Level 3
746	Textile arts	Level 3

Figure 2: Shelf locations for 'Textiles' in the University of Huddersfield Library

Scatter is not generated solely by discipline, but compounded by the extra variables of the item type (e.g. printed-book, audio-visual) and lending status (e.g. quick reference, short-loan) where typically such material is housed separate to the main collection, in

our case on Level 4 - the library entrance hall. Foskett (1982) stresses the importance of the library online public access catalogue (OPAC) for enabling collocation in the face of such scatter.

> In addition to physical form, other factors may influence the place where we choose to keep any given item: we may decide that it may not be removed from the building, so it is placed in the reference section; or it may be suitable for children rather than adults, so it is placed in the children's section; or it may be a rarely used work which is placed in the stack rather than in the section open to the public. All of these factors emphasize the importance of the records, as opposed to the items themselves, for we can gather together in one place the records of items which themselves must perforce be scattered. Foskett, p.4.

Like many subjects, 'Textiles' is victim to disciplinary, form-based and lending status scatter. This may have little adverse effect in the case of known-item searching, disregarding the inconvenience of moving from floor to floor, but may result in information loss when browsing the library shelves if all relevant physical locations are not identified and explored. Added to which, the activity of borrowing means that the entire library collection will never be on the shelves at a single time (Pollitt, 1982).

2.2 Facets, synthesis and citation order

The recognition of subject facets and their application in notational synthesis within the DDC is often characterised as a recent development. However, Miksa (1998) notes that the roots of the synthesis now forming an integral part of the scheme can be traced back to as early as the first edition in 1876:

> ... even here Dewey had already taken important steps towards the systematization of form, genre, language, and geographical categories which were essential to the classification of books as books, but non-essential or 'accidental' to the classification of pure knowledge categories... In addition, he also established a single place in the schedules - 550s Geology - the principle of using a geographic or place facet to be synthesized with topical headings. He did this under the guise of a 'Divide like' instruction, however, rather than in terms of an additive device. In creating all of these provisions, one sees the beginnings of what would eventually become the faceted and synthetic side of the system. (p.7)

More explicit application of synthesis appeared in the 17th edition (1965), with the first mention of the word 'facet' and also 'additive' (i.e. 'add to base number... ') rather than 'division' to describe number building, and later in the 18th edition (1972) when all seven auxiliary tables were created (Miksa, 1998, p.29). Despite introducing further complexity into the system, the trend towards synthetic devices is successively increasing as new editions are published.

The gains achieved in moving towards a faceted approach, however, far outweigh the commensurate complexity, by providing a rich classification grounded in the economies of re-use. The use of synthesis in Dewey enables composite information-bearing items to have their several aspects expressed within the classificatory notation.

Full details from Library Catalogue

Your Search: **Title Browse = Textiles and training in Portugal**
Displaying Record: **1 of 1**

Author	Serrão Andrez , Jaime
Title	Textiles and training in Portugal , Jaime Serrão Andrez, Mário Caldeira Dias
Publisher	Berlin: CEDEFOP , 1988
Contributors	Dias , Mário Caldeira European Centre for the Development of Vocational Training
Subjects	Textiles --Education And Training --Portugal

Location	Collection	Classno	Copy
Holly Bank Library	Loan Collection	677.0071 09469 SER	
		Circulation Data	
	Available		

Figure 3: Bibliographic record from 'Textiles'

The bibliographic record shown in figure 3 illustrates a book title with three facets (Textiles, Training, Portugal). These are expressed explicitly within the Library of Congress Subject Heading and are also encoded in the Dewey notation:

600 Technology
670 Manufacturing
677 Textiles
T1--(0)07 Education, research
T1--(0)071 Education
T1—09 Geographic treatment
T2—4 Europe
T2—46 Iberian Peninsula & adjacent islands
T2—469 Portugal

The synthesis, or number building, in the above example is accomplished by adding from Table 1 (Standard Subdivisions) and Table 2 (Geographic Area), thus combining the facets of *context* and *place* with the enumerated entry for 'Textiles'. Therefore, although shelved in the main 'textile' section in the library, facets from other disciplines are encoded into the notation. As discussed in the previous section, the reverse is also true since other disciplines outside the main 'Textiles' location may include textile-related items:

Full details from Library Catalogue

Your Search: **Title Browse = waxing and waning of the textile industry in West Yorkshire**
Displaying Record 1 of 1

Title	The waxing and waning of the textile industry in West Yorkshire
Publisher	London : Channel Four , 1997
Description	Videocassette c.15 min col VHS
Series	Geographical eye
Notes	Extract for 11-14-year-olds, from 'Geographical Eye' Off-air recording
Subjects	Textile industry --England --Yorkshire --History

Location	Collection	Class no	Copy
Holly Bank Library	Audiovisual Collection	338476770942 81 WAX	Videocassette
	Circulation Data		
	Available		

Figure 4: Textile-related bibliographic record from the discipline of Economics

The record in figure 4 belongs to a videocassette on the textile industry, located in Dewey in the discipline of economics. As in the previous example, several aspects can be derived from the title and LCSH (textiles, economics, geographic location, historical perspective). With the exception of the historical perspective, these facets are again encoded within the rich structure of the Dewey notation:

300 Social sciences
330 Economics
338 Production
338.4 Secondary industries and services
338.47 Goods and services
338.47677 Textiles
T1--09 Geographic treatment
T2--4 Europe
T2--42 England & Wales
T2--428 Northeastern England
T2--4281 West Yorkshire

It can be seen from this example that although located within the discipline of economics, the Dewey 'textile' notation 677 is buried yet still available within the classification number, having been re-used upon instruction in the schedules. Another example in Dewey of this re-use of uniform notation is the heading *Equipment for textile and clothing technologies* (681.7677) within the subdivision of 'manufacturing for specific uses'. The 677 notation is encoded in this ready-built number to denote 'Textiles'. Joan Mitchell, the editor-in-chief of the DDC, explains that a move towards a standard notation is desirable not only for ease of use but also for providing "a meaningful notation for subject retrieval" with a long-term research goal being the "automatic decomposition of Dewey numbers into facets for verification of synthesized numbers and for improved retrieval" (Mitchell 1997, p.86). A standard notation and the links between disciplines it provides may in-

form the application of Dewey in a faceted form that can be manipulated at the electronic interface.

When building any number, a citation order comes into play, reflecting the physical basis of a scheme primarily intended for shelf organisation. Ranganathan realised the constraints the physical world would place on his faceted classification when moving from the Idea Plane to the Notational Plane:

> It is the duty of documentalists to spread the multi-dimensional universe of knowledge along one line. We must make a linear spectrum of it... we have to map an n-dimensional informational space on a one dimensional space. This is the problem in the organisation of information for knowledge and use. S.R. Ranganathan, (1965, p.198).

Ranganathan devised the Colon Classification which mapped the multiple dimensions of the universe of knowledge into a linear Personality: Matter: Energy: Space: Time (PMEST) citation order. In the DDC, citation order for multiple subject facets is governed by particular instructions in the Schedules. Interestingly, the Universal Decimal Classification (UDC) which was developed from the DDC by Paul Otlet and Henri LaFontaine in 1895 has no such designated citation order (Marsh, 1999). Unlike Dewey, UDC was conceived as a bibliographic index rather than for the physical arrangement of library shelves (although it is now used for such purposes, particularly by special libraries) whereby the facet order is determined by each individual institution. Miksa (1998) has argued that future developments should focus on "making the DDC into a more malleable system " with "a vast array of moveable or interchangeable facets of categories, a system that is perhaps best called an object relational database management system of categories". This seems to imply a move away from a prescribed citation order, which although perhaps currently not a realistic prospect in a physical domain marries well, as will be illustrated, with the electronic context of relational databases (Pollitt, 1997, pp.51-52).

This section has described cases in Dewey where synthesis is provided for, albeit within the constraints of a given citation order, motivated by Dewey's application in a physical context of single relative location. In certain cases, however, such facet combining is prohibited in number-building and a choice must be made.

2.3 The need for choice

Despite growing provision for number building in Dewey, some composite and interdisciplinary items are unable to have their various dimensions represented within the classificatory notation and are instead accommodated, rather awkwardly, by a single subject facet.

> In the case where a co-extensive number is not found in the schedules or cannot be built, the classifier must choose one subject or one facet/aspect and ignore the other(s). The question is then which to ignore and which to consider. Chan et al (1996, p.54)

For example, when an item involves two subjects within the same discipline, the subject given the fuller treatment is selected and in cases of equal treatment the 'First-of-two-Rule' entails selection, unless instructed otherwise, of that number which occurs first in the Schedules. Similarly, with items having two aspects of the same subject not provided for by number building, a choice is made according to a Preference Order or explicit instruction in the Schedules whilst taking account of the author's viewpoint. Another slightly different case is the 'Rule of Three' for which a multifaceted item is classified, with a consequent loss of specificity, using an inclusive broader number.

Knowledge organisation schemes are conceived according to choice, implicit in any arrangement, and in the case of the DDC, Melvil Dewey chose to arrange by discipline. The disciplinary basis of Dewey and the ensuing scatter produced by the physical need of single relative location was discussed earlier in this paper. The scheme is therefore founded on this choice to organise by discipline which is still the basic principle of the scheme today. Consequently, the growth in interdisciplinary and multidisciplinary knowledge has caused some difficulty. Beghtol claims that these changes in the nature of knowledge creation have not been reflected in the bibliographic classification schemes which "purport to organize the world of knowledge":

> The major systems that have predominated in the twentieth century were originally predicated on the academic disciplines. This structural principle is no longer adequate because multidisciplinary knowledge production has overtaken more traditional disciplinary perspectives and produced communities of cooperation whose documents cannot be accommodated in a disciplinary structure. Beghtol, (1998, 1).

The DDC has some interdisciplinary numbers, first introduced in the 18th edition (Beghtol, 1998, 5). In the absence of such numbers, the classifier is again faced with a choice, with precedence given to the discipline having the fuller treatment or where this does not apply then the less specific class of Generalities

(000) is a suggested option. These rule-governed choices are neither applicable nor required when Dewey is applied in an electronic context.

The previous sections have identified several examples of information loss, stemming from Dewey's intended use as a scheme for arranging physical items on shelves, namely:

- A single relative location which generates disciplinary scatter both throughout the notation and the library itself
- A citation order when synthesising subject facets
- Choice in the face of certain composite or interdisciplinary and multidisciplinary items not accounted for by notational synthesis

The rich, but under-utilised, structure and notational re-use in Dewey was highlighted by decomposition of synthesised codes.

Ways in which this rich structure can be applied in an electronic context are being explored in a project investigating the application of the Dewey Decimal Classification Scheme and Library of Congress Subject Headings in a view-based searching OPAC for subject searching and browsing library collections (Pollitt, 1998; Pollitt, Tinker & Braekevelt, 1998). Following a brief introduction to the application of the DDC in electronic domains, the remainder of this paper will describe the initial work of this project.

3. DDC and the electronic context

3.1 Dewey on the World Wide Web

Work has already been undertaken to prepare the Dewey Decimal Classification Scheme for use in an online environment, beginning with the publication of the DDC in machine-readable format in 1979. This was followed by Karen Markey's influential study in 1986, which saw the first implementation in an online catalogue of a library classification scheme for subject searching and browsing (Markey, 1989; Vizine-Goetz, 1996). More recently the hypertext Internet environment has offered opportunities for using subject heading arrangements in online searching, evidenced by the popularity of 'home-grown' subject schemes such as *Yahoo* (http://www.yahoo.com/).

Others have turned to traditional library classification schemes to provide Internet subject access. Examples can be found of DDC summaries being transplanted, essentially in their paper-based form (although some use slightly modified captions) together with the decimal notation, in subject gateways such as:

- *BUBL Link/5.15*
  <http://bubl.ac.uk/link/ddc.html>
- *Canadian Information by Subject* at the National Library of Canada
  <http://www.nlc-bnc.ca/caninfo/ecaninfo.htm>
- *Dewpoint* (formerly *CyberDewey*) by David Mundie
  <http://ivory.lm.com/~mundie/CyberDewey/CyberDewey.html>
- *Net Sites by the Numbers* at Tempe Public Library, Arizona
  <http://www.tempe.gov/library/netsites/>
- *The New Athenaeum: Internet Resource Guides Developed by Libraries* by Robert J Tiess
  <http://members.spree.com/athenaeum/mguide1.htm>
- *Webray* at Morton Grove Public Library, Illinois
  <http://www.webrary.org/ref/weblinksmenu.html>
- *XmlTree - Directory of Content* by Wavefront Ltd., London
  <http://www.xmltree.com/>

In these gateways, Dewey is utilised in a menu-driven interface, with the user selecting a single Dewey class and browsing down/up each hierarchy via a series of screens until the desired Internet resources have been located. Non-Dewey caption direct searches are also usually available via a form-based interface.

BUBL Link/5.15 is perhaps the best known and most widely used of the above subject gateways in the UK. A new interface with more search options was introduced in February 1999. Users are provided with several 'views' onto the data, namely:

- Subject menus
- A-Z of subject index terms
- Dewey
- Country of origin
- Type of material
- Updates (bimonthly additions)
- Random links

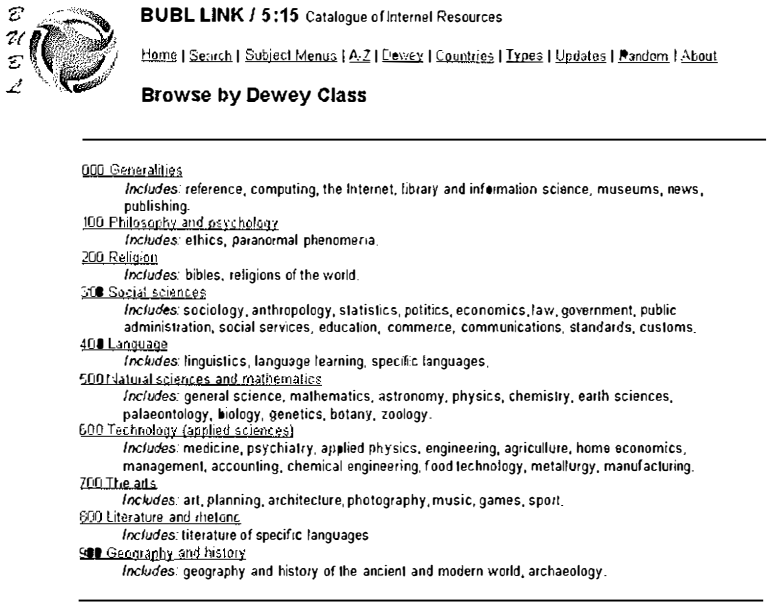

Figure 5: Browsing by Dewey on BUBL

These 'views', however, can only be accessed in a linear manner, one at a time. As will be demonstrated in the following section, a view-based searching interface would enable a user to access these views simultaneously and combine their contents in a multidimensional Boolean search with a classified results display. However, multidimensional access could also be assisted at the current BUBL Dewey interface if we remember that the DDC is now operating within an electronic not physical domain, as the following example will demonstrate:

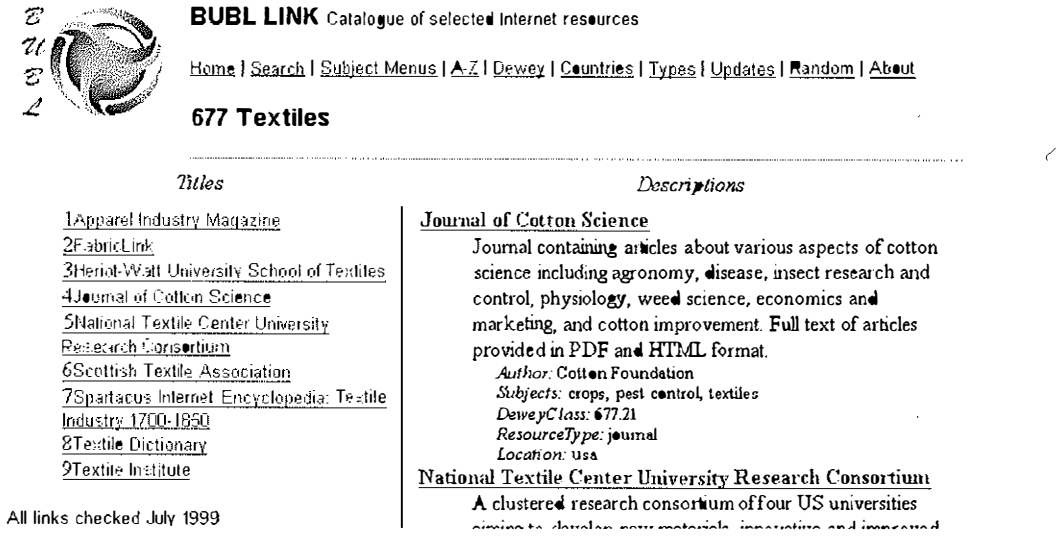

Figure 6: Browsing 'Textiles' on BUBL

Figure 6 shows a listing for the online *Journal of Cotton Science*, classified to the Dewey 677 'Textiles' category. However, as can be seen from the journal's abstract and subject index terms, the multi-faceted subject of Cotton Science includes not only the product (cotton fabric) but also the cultivation (cotton crop). Consequently due to the single Dewey code, as previously demonstrated by examples from the University of Huddersfield OPAC, when browsing the BUBL Dewey view there is a scatter across disciplines and information loss. As this journal is located uniquely in 677 'Textiles', it cannot be accessed at the Dewey interface from a cotton cultivation perspective; this aspect is only available upon consultation of

the Subject Menu or subject term A-Z views. Since a single relative location is not applicable here, further Dewey codes could be included to enable access from a variety of starting points such as: 633.51 Cotton (Agriculture), 632.5 Weeds, 632.7 Insect Pests, 338.17351 Cotton (Agricultural Economics) and 380.1 Commerce (Trade). A decision or policy would need to be established to determine how exhaustive this 'classification as indexing' would be, or some weighting scheme applied to indicate the relative importance (extent of coverage) of each descriptor.

Thus, in an electronic context where there is no requirement to locate an item in one single place, multiple Dewey notations and their accompanying captions can be assigned to enable a user to access a subject from a variety of angles. As Searing states:

> The classification structure is a theoretical map of human knowledge, but in practical terms, it is a map of the stacks, guiding readers to places where works on similar topics can be found in physical proximity. In the electronic environment, texts are freed from the limitations of physicality, so in theory any number of classes can be assigned . Searing (1996, 320)

This then applies as equally to the library OPAC, without necessitating any disruption in a collection's actual physical arrangement. In counteracting this information loss, search quality may be improved whilst a single location for shelving purposes is maintained by identifying the principal Dewey code for the item. Davies (1989) claims that the ability to search across disciplines, or 'horizontally', may also stimulate creativity and quicken the resolution of research problems by connecting researchers in different disciplines working on analogous research questions, previously unrealised due to differences in terminology.

Beghtol concurs with this practice of classification as pre-coordinate indexing:

> Bibliographic classification systems express multidisciplinary topics easily if one assumes a classified catalogue (manual or electronic) with multiple notation access points for each document. Beghtol (1998, 4)

For example, *Iter: the Bibliography of Renaissance Europe* (http://iter.library.utotonto.ca/iter), cited by Beghtol, assigns multiple Dewey codes to online journal articles and reviews to improve access to interdisciplinary materials; future plans are to produce a browse-based Dewey interface.

OCLC, the owners of Forest Press and the publishers of the DDC, has also been applying Dewey in the Internet environment of its *NetFirst* database, but has appropriately recast the traditional DDC summaries into more meaningful end-user terminology for this purpose. For example, 'Generalities' is replaced by the more explicit 'Books, Computers, Internet', bringing to the surface particular subjects buried deeper in the hierarchies without requiring changes to Dewey's logical structure. In addition to terms from Dewey's relative index, a valuable resource in this caption revision is enhancement of the Dewey knowledge base using associated LC subject headings (an ongoing service by OCLC), other subject-oriented knowledge bases/thesauri (ExTended Concept Trees project) and current free-text vocabulary (Wordsmith project) (Vizine-Goetz, 1997). Beyond the three-level summaries, frequently used class codes in *NetFirst* will be the focus for revision.

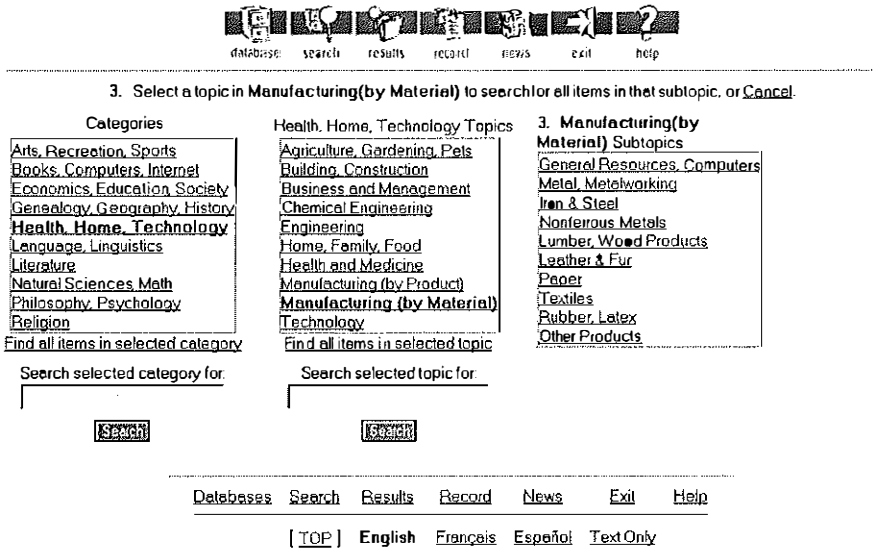

Figure 7: NetFirst browse for 'Textiles'

Figure 7 shows the revised three-level summaries in a browse for the subject of textiles. Significantly, *NetFirst* has relinquished use of the famous decimal notation. Currently users can only browse the database using these first three levels. Vizine-Goetz says this is a "good first step" but believes that there is a need to look beyond this "since many of the DDC numbers assigned to NetFirst records extend more than four or more digits past the decimal point" (Vizine-Goetz, 1996). This need for specificity will also apply in the use of Dewey in library OPACs.

Upon selection of 'Textiles' in the third level summary (Figure 7), the user is presented with a result set of 33 records, as shown in Figure 8.

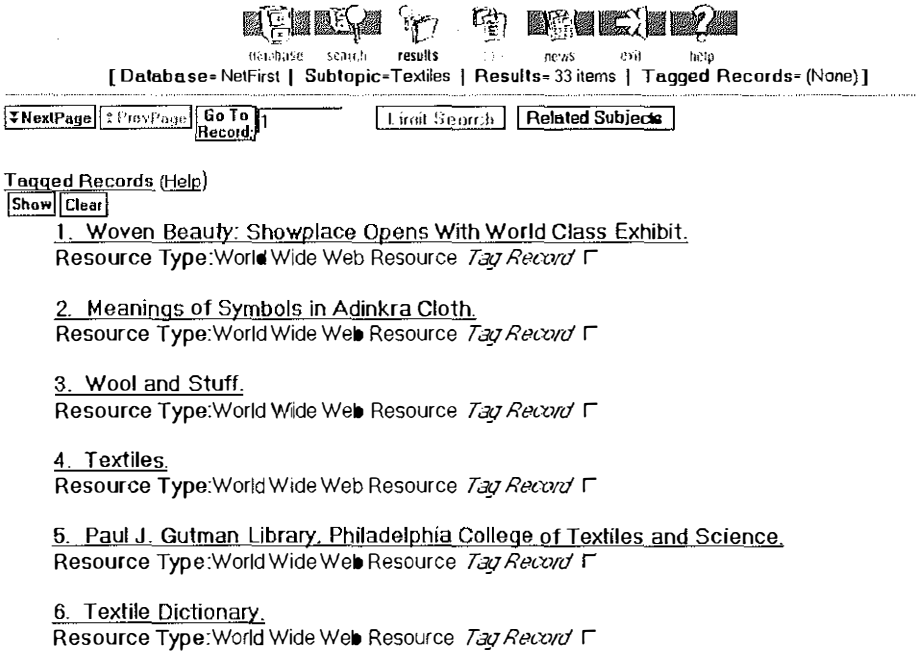

Figure 8: Results of the textile browse

Record 1 is selected:

TITLE:	Woven Beauty: Showplace Opens With World Class Exhibit. (*Type:* World Wide Web Resource)
LINK:	http://www.abqjournal.com/venue/museums/1mus9-21.htm
SUMMARY:	Features a September 20, 1997 article "Woven Beauty," written by Leslie Linthicum and published in "Albuquerque Journal." Discusses the opening of the Navajo Museum, Library and Visitors Center in Window Rock, Arizona and an exhibit of Navajo textiles.
DEWEY INFO:	*Class:* 970.00497209 677.02 020 025.174 *Subjects:* General history of North America. Textiles. Library & information sciences. Library operations.
LC SUBJECT:	Navajo Museum, Library and Visitors Center. Navajo Indians -- History. Textile fabrics -- Exhibitions. Libraries and museums.
DB NO.:	317932
DOMAIN:	com Commercial
DESCRIPTOR:	Electronic publications.
GEOG. INFO:	*Class:* T2-791 T2-73 *Subject:* Arizona. Window Rock (Ariz.).

Figure 9: NetFirst item record

Figure 9 demonstrates how OCLC is adapting Dewey for Internet use. Multiple Dewey codes are assigned to capture the multi-disciplinary nature of the resource (Figure 10):

Dewey code	Traditional Dewey caption(s)	NetFirst Dewey caption
970.00497209	General History of North America; Racial, ethnic, national groups; North American native peoples; Na-Dene languages (including Navajo) Historical, geographic, persons treatment	General History of North America
677.02	Textiles; General topics	Textiles
020	Library and information sciences	Library and information sciences
025.174	Library operations; Administration of collections of special materials; Machine readable materials	Library operations

Figure 10: Decomposition of multiple Dewey codes

The decomposition illustrated in figure 10 reveals the rich underlying hierarchy encoded within these Dewey notations which could be made explicit by the introduction of further captions. Instead, as in the classification of physical items, this detail is subsumed under a more general top-level heading. Thus, although NetFirst has recognised and broken away from the limits of a physical classification by assigning multiple Dewey codes and captions, much specificity remains to be exploited *within* a single Dewey code. This may become particularly important if the lower-level captions are employed for browsing. The beginnings of a faceted approach are displayed in Figure 9 by the addition of separately searchable geographic area codes from Dewey's Table 2. OCLC recognises the potential of such decomposition since it comprises one of five key areas in its research agenda. However, seemingly little work has been done in this area as yet.

3.2 The unrealised/untapped value of classificatory structures at the OPAC interface

Despite this growing recognition of the value of classificatory structures for subject searching on the Web, commercial OPACs, many of which are now suitably within a hypertext Web environment, surprisingly seem set to remain in splendid isolation. As early as 1964, Swanson (1964) was envisaging an 'automated catalog' with a subject search that used classificatory structures and later as the 'first generation' (to use Hildreth's typology) of OPACs emerged other proponents followed, often demonstrated by the implementation of experimental OPACs (e.g. Markey, 1989; Cochrane, 1982). More recently, Hildreth (1995b) has recommended that 'third generation' OPACs:

> "break out of the query-oriented, Boolean mindset, we need to turn the conventional query-first-then-browse paradigm upside down. Searching by exploration, recognition, and discovery in a well structured bibliographic space should be the primary search interface provided to information seekers... "

Hildreth includes provision for 'classification-based searches' within this browsing/exploratory paradigm. Although, still assuming a query-based system for subject searching, the IFLA guidelines on OPAC displays (Yee, 1998) do recommend that a user be shown the cross-disciplinary classification categories for a specified search term and be able to 'right click' on the mouse to view the hierarchical context of any classification number.

In commercial systems, however, the DDC is used for the physical purposes of *directing* users to items on shelves, with access to the hierarchical knowledge structures behind classification schemes remaining the sole preserve of the librarian. We agree with Cochrane (1982) that this seems "a horrible waste", particularly when users have been found to experience many problems when bringing vague and evolving information needs to query-driven subject search options (Markey, 1984).

> Information retrieval (IR) is only effective when the people involved share common knowledge structures. The match of concepts among user, intermediary, and system then has some chance of success. (O'Brien 1994, p.219)

If the user had visual access to the knowledge structures used by the library system then perhaps subject searching would be more fruitful, and browsing encouraged, due to the accommodation of search by selection as well as specification.

Making the knowledge structures of Dewey visible at the OPAC interface may also serve to have positive impact on students' learning experience. Jack Mills is an ardent advocate of the educational value of classification:

> ... classificatory structures assist seekers of information to realize the connectedness of concepts in a store of information, this must have implication for the education of those seekers...
> The value of a classification in formulating a search prescription is the result of its systematic presentation of the connectedness of concepts. It presents a clear picture not only of the concepts involved but also of their generic contexts and their syntactic relation.
> Mills (1997, p.9).

Mills cites the work of Lilley's 1954 study in which students were asked to specify subject headings for a particular search area. The results "in many cases demonstrated a serious absence in their education of any sense of an overall structure of knowledge or of any sort of connectedness between different fields which might give them some sort of perspective when searching for information" (p.10). The informative context of classificatory structures for searching and browsing in a library OPAC may go some way towards improving a situation which Mills believes has changed little since completion of Lilley's study.

Understanding this 'connectedness' is particularly important in a University such as Huddersfield, where many students can take modules from a variety of disciplines, courtesy of the CATS Scheme (Credit Accumulation and Transfer), to create their own study 'pathways', for example computing can be studied alongside a modern european language. For these students, therefore, understanding how their chosen subjects represent knowledge (i.e. terminology) and relate to each other is very important in information seeking and for a coherent grounding to their educational work. Classification is an available tool for achieving this end. The library OPAC should be conceived beyond a locating tool, as also a means of exploring and discovering information (Bates, 1986; Hildreth, 1995a; Kulthau, 1999).

3.3 DDC in a view-based searching OPAC

We are currently investigating the value of classificatory structures for searching and browsing in a project to apply the Dewey Decimal Classification Scheme at the interface of a view-based searching OPAC. As exemplified earlier, the DDC is a scheme predicated on the physical organisation of library materials and thus in its traditional application is not optimally suited for providing electronic access. As such, the project seeks to build on work of the OCLC *NetFirst* team by tailoring Dewey usage for this electronic domain, addressing and alleviating the issues of disciplinary scatter, strict citation order and choice amongst subject facets.

We are exploring such issues in a technique called view-based searching. Underpinned by a relational database model, view-based searching utilises knowledge structures in navigable views, reflecting facets of the objects in databases with implicit Boolean searching and mutually constraining views. These techniques have been previously applied to a directory of hotel information, the business database of the European Parliament (EPOQUE), university student records and a 600,000 subset of the EMBASE bibliographic database (Pollitt, 1997; Treglown et al, 1997). View-based searching interaction is characterised by selection rather than specification, thus obviating the complexities of query formulation found in the command-line and more recent form-based interfaces of today's commercial OPACs, particularly important when users have difficulty articulating vague information needs (Belkin, Oddy and Brookes, 1982; Taylor, 1968).

3.3.1 The OPAC interface and searching capabilities

View-based searching provides both direct search and browse capabilities using the hierarchical structure of the DDC. Currently, the collection can be browsed by four simple facets:

- Subject (Dewey)
- Type of Publication
- Year of Publication
- Geographic Setting

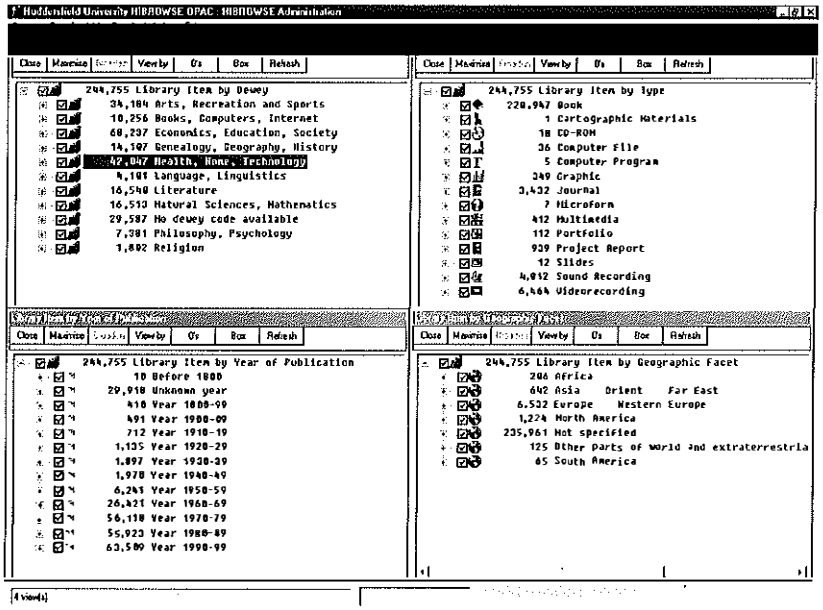

Figure 11: Four facets in the HIBROWSE for Dewey OPAC

Figure 11 shows four views onto the entire holdings of unique titles in the University of Huddersfield library. Owing to their 'fitness for purpose', the *NetFirst* summary captions rather than the traditional paper-based counterparts were chosen for use in development of the view-based searching OPAC.

The user here is selecting the Dewey caption 'Health, Home, Technology' in the 'Library Item by Dewey' subject view. With the exception of the 'Library Item by Type' view, each of the descriptors in the views have further narrower descriptors beneath them when selected.

The *subject* view (Library Item by Dewey) shows the ten main Dewey classes and the number of titles for each in the library collection. The *NetFirst* revised Dewey captions were re-used in the first development phase of the view-based searching OPAC. These captions comprise the DDC three-level summaries only, so were extended, using headings from the paper-based version in the first instance, in order that the university library collection be fully represented.

The *type of publication* view (Library Item by Type) provides a profile (illustrated by icons) of the particular types of material held by the library. As currently expected for an academic library, the collection is predominantly book-related.

The *year of publication* view gives another interesting profile of the collection, notably with regard to its history. This chronological hierarchy by decade can also be broken down further into a year-by-year account.

The *geographic setting* view was derived by decomposing the Dewey notation. Thus, the digits .093-.099 indicate that a library title has a geographic aspect which has been added to the main Dewey notation from the auxiliary tables 1 and 2. These .093-.099 notations are added at the discretion of the classifier. However, a geographical aspect can also be encoded upon direct instruction in the Schedules, without the interpolation of the .09 notation from Table 1. Such cases are more difficult to identify and extract; however a proposed solution is to search the electronic Dewey (*Dewey for Windows*) for those notations having references to Table 2, and then search for .3-.9 notations within these. The remaining issue is whether and how to reconcile this geographic facet with the main geography schedule.

A user can enter the view-based searching OPAC by either a direct search or a four facet browse view. Upon selection of the 'Search' option, the user can search for a specific term in a Dewey caption and then browse the retrieved hierarchical view to tailor, modify or refine their search. When browsing, navigation is enabled by a 'point and click' interaction with the user moving up and/or down each hierarchical view. The views are linked, mutually constrained by an implicit Boolean 'AND' query, and allow the user to select persistently the degree of specificity for each particular facet, simultaneously observing the effects across all facets. This technique was applied in previous view-based searching applications, notably HIBROWSE for EMBASE (Treglown et al, 1997).

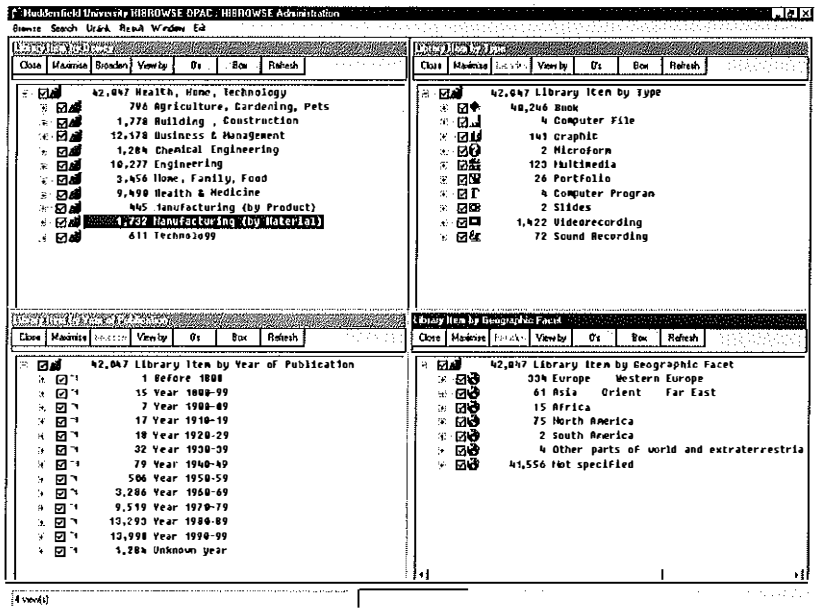

Figure 12: Browsing 'Health, Home, Technology' in the Dewey subject view

In figure 12, the user has selected 'Health, Home, Technology' and the search is immediately refined from 244,757 to 42,047. Since the views are mutually constraining, we can also simultaneously observe the number and type of material, particular year of publication and whether or not there is a geographic focus. The user moves down the Dewey hierarchy and selects 'Manufacturing (by material)':

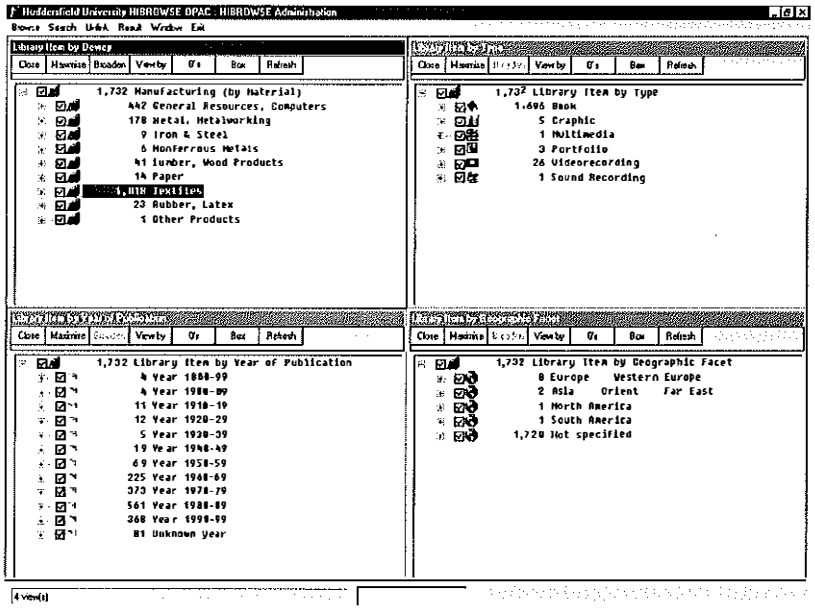

Figure 13: Browsing 'Manufacturing (by Material)

Selection of 'Manufacturing (by Material)' narrows the search further from 42,047 to 1,732. Again, we can view the effects of this across the other facets. The interaction in figure 13 shows that the library has 1,732 titles classified to the subject of 'Manufacturing (by Material)', their type (for example, 1,696 books and 26 videorecordings), year of publication in decades and that there are just 12 titles which examine the subject from a particular geographic standpoint. The user selects the 'Textiles' caption for which the library holds 1,018 titles.

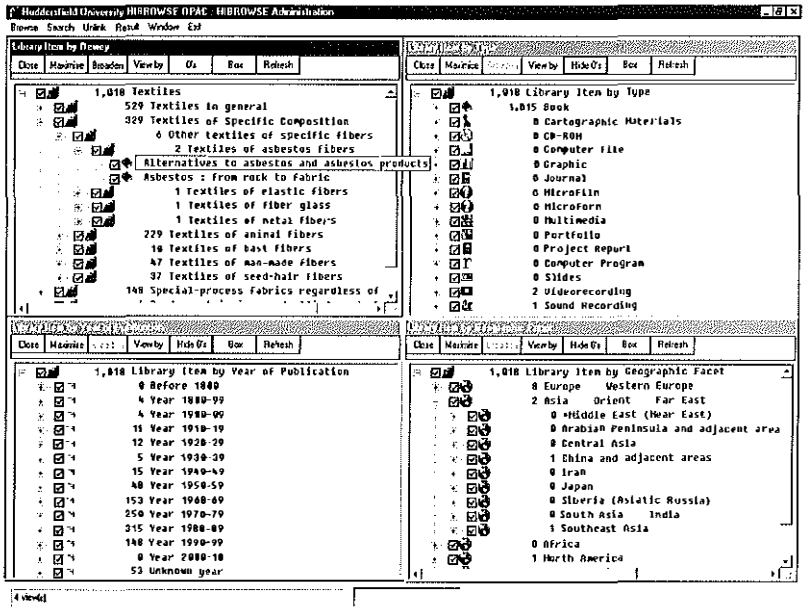

Figure 14: Browsing 'Textiles'

The user has now focused in on 'Textiles'. Figure 14 shows that titles classified to this subject are predominately books (1,015); that the library holds mainly recent (peaking in the 1980s) yet some historical material, although nothing in the main collection before the year 1800; and that very few titles are held or have been classified which relate to a particular geographic area. The user examines the 'Geographic Facet' for titles about Asia, discovering that the two titles focus on 'China' and 'Southeast Asia'. The zero statistics give explicit clarification as to what has *not* been classified to a particular subject. The user returns to the 'Textiles' hierarchy and expands 'Textiles of asbestos fibers' to reveal two titles which happen to be books. The icons here at the record level are valuable for indicating at a glance the type of material sought.

This example interaction has demonstrated the flexibility and visibly inclusive nature of the view-based approach. The persistent query enables a user to successively broaden or narrow a search by adjusting the specificity of each facet whilst observing the effects across the remaining facets until the search meets their own individual information needs.

3.3.2 Future work

Further development will see the incorporation of free text searching and additional facets such as 'Author' and 'Institution'. The 'Search' facility currently searches only on terms contained in Dewey captions; addition of the relative index with its synonyms, quasi-synonyms, narrower terms and related terms will increase the number of entry points into classified views. Circulation data and copy details have yet to be included, but such data could facilitate the creation of views for library management.

In addition to the standard Dewey categories, further *subject* facets can be derived from the Dewey Tables and via decomposition, as illustrated in Section 2.2., to identify notations frequently re-used in the Schedules and thus common to several subject disciplines. This will create a more powerful search interface with greater specificity in subject searching and collocation of recurrent concepts which in a physical application of the DDC are often buried deep within Dewey notations and enumerated across the Schedules. Applying Dewey in this faceted form at the electronic interface means that a citation order is no longer required, which provides the user with the flexibility to select facets in any order, somewhat comparable to Miksa's notion of "moveable or interchangeable facets".

As discussed in Section 3.1, the disciplinary scatter and information loss incurred by the physical constraint of single relative location could be resolved by assigning multiple Dewey captions. This will enable users to browse a subject such as 'Textiles' in all its scattered Dewey locations, and make links among textile facets using the implicit Boolean searching of view-based searching described above. As illustrated in Section 2.2, these additional descriptors could be derived from Library of Congress Subject Headings and subsequently mapped to the appropriate Dewey captions. OCLC has already done much work in making such Dewey-LCSH links.

4. Conclusion

This paper has highlighted the problems inherent in the physical linear arrangement of library collections and presented the multidimensional access provided by view-based searching as one possible solution. The library OPAC is a vital tool, able to compensate for and transform the limited searching and browsing available at the library shelves. The knowledge structures of classification schemes such as Dewey are currently under-utilised for subject searching in OPACs. Library classifications, however, should not be constrained by their original 'shelf motivated' form but be applied in a way which takes full advantage of the electronic environment. In so doing, a more complete and individual access to library collections will be provided which aims to satisfy the diverse needs and degree of query specificity users may bring to a system.

5. Acknowledgements

The View-based Searching and Dewey Classification project is funded by the University of Huddersfield and benefits from collaboration with OCLC Forest Press, OCLC Office of Research and the British Library. The team at Huddersfield is grateful for the support and encouragement given by Joan Mitchell, President of OCLC Forest Press and editor of DDC, and her colleagues at OCLC.

Notes

* This paper is based on a presentation given at the British Library Seminar on "The British Library and subject access: investment in Dewey Decimal Classification" in February 1999.

References

Bates, M J (1986) Subject access in online catalogs: a design model. *Journal of the American Society for Information Science* 37(6), 357-376.

Beghtol, (1998) Knowledge domains: multidisciplinary and bibliographical classification systems. *Knowledge organisation*, 25(1/2), 1-12.

Belkin, N J, R N Oddy & H M Brookes (1982) ASK for information retrieval: part I. Background and theory, *In:* K Sparck Jones & P Willett, eds. *Readings in information retrieval.* San Francisco: Morgan Kaufmann, 1997.

Davies, R (1989) The creation of new knowledge by information retrieval and classification. *Journal of Documentation* 45(4), 273-301.

Chan, L M et al (1996) *Dewey Decimal Classification: a practical guide*, 2nd ed. Albany, New York: Forest Press.

Cochrane, P A (1982) Classification as a user's tool in online public access catalogs. *Studien zur Klassifikation*, 11, 260-268.

Forest Press (1998) Introduction to the Dewey Decimal Classification – print version

[21st ed.]. *Dewey for Windows* [CD-ROM]. Albany, New York: OCLC Forest Press.

Foskett, A C (1982) The subject approach to information, 4th ed. London: Clive Bingley.

Hildreth, C R (1995a) Browsing and exploring: a new paradigm for IR/OPAC system design. *In:* C R Hildreth, *Online catalog design models: are we moving in the right direction? A report submitted to the Council on Library Resources August, 1995.* (http://www.ou.edu/faculty/H/Charles.R.Hildreth/clr-five.html), 25 October 1999.

Hildreth C R (1995b) Conclusion: outline of the next generation of online catalogs. *In:* C R Hildreth, *Online catalog design models: are we moving in the right direction? A report submitted to the Council on Library Resources August, 1995.*
(http://www.ou.edu/faculty/H/Charles.R.Hildreth/clr-six.html), 25 October 1999.

Kuhlthau, C C (1999) Accommodating the user's information search process: challenges for information retrieval system designers. *Bulletin of the American Society for Information Science* 25(3), 12-16.

Markey, K (1984) Subject-searching experiences and needs of online catalog users: implications for library classification. *Library Resources and Technical Services*, Jan/March 134-149.

Markey, K (1989) Subject searching strategies for online catalogues through the Dewey Decimal Classification. *In:* C R Hildreth (ed.) *The online catalogue: developments and directions.* London: The Library Association.

Marsh, E (1999) Improving communication and classification in the next century. *OCLC Newsletter*, 237, January/February.

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

Mills, J (1960) *A modern outline of library classification.* London: Chapman & Hall.

Mills, J (1997) Introductory address, *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, Netherlands, FID, 1997, pp.1-11.

Mitchell, J S (1997) Challenges facing classification systems: 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, Netherlands, FID, pp. 85-89.

O'Brien, A (1994) Online catalogs: enhancements and developments. *In:* M.E. Williams (ed.), *Annual Review of Information Science and Technology, 29.* Medford, N.J.: Learned Information, pp.219-242.

Pollitt, A S (1998) The application of Dewey Classification in a view-based searching OPAC. In: W Mustafa el Hadi, J Maniez & A S Pollitt (eds.), *Structures and relations in knowledge organisation. Proceedings of the fifth international ISKO conference 25-29 August 1998, Lille, France.* Wüzburg: Ergon Verlag, pp. 176-183.

Pollitt, A S (1997) Interactive information retrieval based on faceted classification using views, *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, Netherlands, FID, pp. 51-56.

Pollitt, A S (1989) Information storage and retrieval systems: origin, development and applications, Chichester: Ellis Horwood.

Pollitt, A S, A J Tinker & P J Braekevelt (1998). Improving access to online information using dynamic faceted classification. *In: Online information '98, 22nd international online information meeting proceedings 8-10 December, 1998, London.* Oxford: Learned Information.

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

Searing, S E (1996) Meeting the information needs of interdisciplinary scholars: issues for administrators of large university libraries. *Library Trends* 45(2), 315-342.

Swanson, D R (1964) Dialogues with a catalog. *Library Quarterly* 34, Jan 1964, 113-125.

Taylor, R S (1968) Question negotiation and information seeking in libraries. *College and Research Libraries* 29, 178-189.

Treglown, M et al (1997) *HIBROWSE for bibliographic databases: a study of the application of usability techniques in view-based searching. British Library Research and Innovation Report 52*, Huddersfield: University of Huddersfield, 1997.

Vizine-Goetz, D (1996) Using library classification schemes for Internet resources. OCLC Internet Cataloging Project Colloquium Position Paper (http://www. oclc.org/ oclc/man/colloq/v-g.htm), 25 June1998.

Vizine-Goetz, D (1997) OCLC investigates using classification tools to organize Internet data. OCLC Newsletter, March/April, No. 226 (http://www.oclc.org/oclc/new/n226/research.htm), 25 June 1998.

Yee, M (1998) Guidelines for OPAC displays prepared for the IFLA Task Force on Guidelines for OPAC displays November 24, 1998. (http://www.ifla.org/VII/s13/guide/opac.htm), 1 November 1999.