

'Some textbooks are so fundamentally necessary for a study of their subject area that they have attained a classic status and are to be found in multiple, well-thumbed copies on library shelves . . . A. C. Foskett is to be congratulated on the production of this the fifth edition and to be heartily thanked by all teachers and students within the information retrieval field for . . . ensuring the continuity of his work . . .'

LIBRARY REVIEW

'the additions ensure that the book will retain its pre-eminence as an indispensable standard text.'

JOURNAL OF DOCUMENTATION

' . . . this new edition is very welcome, and I shall be recommending it to my students. Despite, or maybe because of, the reduced time available for teaching classification and indexing in library schools, a clear and concise account of the principles and the major schemes is invaluable.'

EDUCATION FOR INFORMATION

'a recommended acquisition for college and university libraries, or for any library seeking to organize its resources based on principle, and not accident.'

THE JOURNAL OF ACADEMIC LIBRARIANSHIP

'The strength of this book lies not simply in its scholarship, but also in its unassuming, didactic style of presentation, its easy-to-read, easy-to-learn format, and perhaps most of all its sheer commitment.'

MANAGING INFORMATION

'This then is an eminently practical book dealing with issues of immediate concern, written in a highly accessible style and not without considerable wit. Its reputation as the standard text on the subject is confirmed.'

LA RECORD

The Subject Approach to Information

Fifth edition

A C Foskett MA FLA AALIA

School of Communication and Information Studies,
University of South Australia

*O What a tangled web we weave
When first we practise to retrieve . . .*



Library Association Publishing
London

References

- 1 Grolier, E. de, *A study of general categories applicable to classification and coding in documentation*, Paris, Unesco, 1962.
- 2 In the rules for main classes in CC6.
- 3 Palmer, B. I., *Itself an education*, London, Library Association, 1971. Chapters 2 and 3.
- 4 See Chapter 19 for details of the writings of H. E. Bliss.
- 5 Toffler, A., *Future shock*, London, Bodley Head, 1970.
- 6 Austen, J., *Pride and prejudice*, 1813. Chapter 20.

Library resources and technical services has a literature review each year with the subtitle 'the year's work in subject analysis'. Other periodicals to watch include *Knowledge organization* 1993– (previously *International classification*, 1974–1992), and *Cataloguing and classification quarterly*. The appropriate chapters in *British librarianship and information work* are valuable summaries.

Chapter 11

Notation

Unlike alphabetical arrangement, systematic order is not self-evident, and we may find that there are differing views as to the best (i.e. most helpful) arrangement at any point. It would be most unhelpful if we had to work our way through the schedules every time we wished to find a particular subject on the shelves or in the catalogue, even if the guiding was of a much higher standard than is usually found. To make systematic arrangement a practical proposition we must add to the schedules a set of symbols – a notation – which *does* have a self-evident order; we can then use the notation to find the subjects we want on the shelves or in the catalogue in a clearly organized order.¹

There are two important points here. The first is that the notation is something *added* to the schedules, and only when we have decided on the arrangement can we begin to think about the notation. It is an unfortunate fact that notation is often taken to *be* the systematic arrangement, and classification schemes have been criticized for poor arrangement when it has been the notation which has failed, not the schedules. The notation cannot turn a bad schedule into a good one, but it can so hamper the use of a good arrangement that it becomes unacceptable to its users. To quote H. E. Bliss, 'notation . . . does not make the classification, tho it may mar it'.²

The second point is that the notation has to show the order: that is its function. The notation itself must therefore have a self-evident order, otherwise it will not serve its purpose. The order must be self-evident not only to the professional information handler, but also to the general user, who cannot be expected to appreciate results which are not immediately obvious, no matter how intellectually satisfying they may be to the compiler.

There are two sets of symbols which have a widely recognized order: Arabic numerals, used world-wide, and the roman alphabet, understood wherever a Western European language is used. Using letters, we have the choice of upper and lower case (capitals and small letters) which means in effect that we have three sets of symbols we can use rather than just two. A notation which uses only one set of symbols is called a *pure* notation, while one that uses more than one kind is known as a *mixed* notation. It is clear that only a pure notation will give us the completely self-evident order we have stated to be necessary, but other factors enter into the picture which may make mixed notation superior to the extent that it may be worth while accepting the loss of consistency.

Memorability

Notation is the means by which we get from a subject expressed in words in an alphabetical listing of some kind to that same subject in context in the systematic arrangement. It has to appear in catalogue entries, on the backs of books, in stock records, shelf guides – anywhere that we need to find our way around the systematic arrangement. We must therefore be able to carry it mentally with ease, write or key it without error, inscribe it on book spines which may be relatively narrow. It must also lend itself to the maintenance of the desired systematic arrangement, for example in the shelving of books by non-professional staff in a busy library. The notation must be easily used for all of these purposes; to do so it must possess certain qualities which between them add up to what is called *memorability*, but might also be denoted by the ‘in’ term *user-friendliness*.

The first quality is *simplicity*, by which we mean that it must be easy to grasp mentally. Consider the following ten digit number:

6183022262

This looks long and clumsy, and most people would find it difficult to grasp as a whole, but if we split it up into three shorter sections:

618 302 2262

it at once becomes much simpler, because we can recognize the structure of a telephone number. By splitting the number up, we have increased its length by two digits (counting each space as a digit), but this actually makes it easier to grasp despite the increased length.

This leads us to a consideration of the ease of use of different kinds of notation. The following pieces of notation are all of much the same length, but clearly some of them are easier to grasp than others:

1	7382159142	
2	738 215 9142	[telephone number]
3	JVG XBF 8EAD	[BC2]
4	Z695.1.E5E5	[LCC]
5	921,52,15,76	[BSO]
6	621.312 424	[DDC]
7	621.315.5:669.14	[UDC]
8	Si(61)NoHm+Hf	[CRG classification of library science]
9	ntx.city.unisa.edu.au	[Internet node]
10	O,111,2J64,HE+8	[CC7]

We find that on the whole devices which normally act as separators – punctuation marks, spaces – are psychologically acceptable for this purpose in notation, though because separators are *empty digits* which convey structure but not meaning they lengthen the notation. Mixed notation *may* be easier to grasp than pure notation – but only if we can grasp the pattern to the mixing; and numbers are for most people more acceptable as a notation than letters. In practice, we also find that familiarity

is a great help; if we regularly use the notation of a scheme, we quickly recognize the patterns even if they are not immediately obvious.

The second quality which is important is *brevity*. Other things being equal, a short piece of notation is more easily grasped than a long one; as we have seen, other things are not always equal, but there is no doubt that brevity is important. For example, it is difficult to put a long piece of notation of the spine of a book for shelf arrangement, unless we can split it up into shorter units, and the longer the notation the less likely it is to be memorable. Brevity depends on two factors: the *base* of the notation, and the *allocation*. The base is simply the number of symbols available in the system: for numbers this is ten (0/9) or nine if we ignore the zero; for letters it is 26. If we mix the notation by using both numbers and letters we may have 35 (there are dangers in using both O [capital letter o] and 0 [zero], as can be seen in CC), while if we use both upper and lower case letters and numbers we will have about 60. (There is the possibility of confusion between 1 (one) and l (lower case l), between i and l, and hand-written b and 6, as well as O and 0.) If we use numbers, we shall have longer symbols than if we use letters. For example, if we have about 2,000 items in our schedule and need to show their order, we have to use up to four figures but only three letters. The longer the base, the larger the number of items that can be arranged by a given length of symbol; mathematically, if the base contains x symbols, then by using up to n digits we can construct

$$x^n + x^{n-1} + x^{n-2} + x^{n-3} + \dots + x^3 + x^2 + x$$

different notational symbols. The general preference for numbers has to be set against the fact that letters will in general give shorter symbols.

Another factor affecting brevity is the way that the notation is allocated. Some subjects are static: they have not developed much in recent years. Others are dynamic, and develop steadily, or sometimes rapidly, over the years. When we allocate the notation for a classification scheme we should try to make sure that we give a large share to dynamic subjects, even if this means relatively long notation for static subjects to begin with. After a few years, the notation for static subjects will not be any longer, while that for dynamic subjects will inevitably have grown. Of course, we cannot tell in advance which particular subjects are most likely to grow in years to come, but can at least make some sort of intelligent guess, bearing in mind that if we could indeed foretell the future the construction of classification schemes would probably not be our chosen profession.

In his first edition, Dewey gave the same spread of notation to Logic as he did to Engineering: ten three-figure numbers. (It must be remembered that this was the age in which a US Senator could recommend the closure of the Patent Office on the grounds that everything of use had already been invented!) As a consequence, in DDC21 we still find three-figure numbers in Logic (which has been static now for the best part of 2,000 years) but in Engineering, particularly those branches which have had to be inserted since the scheme was first drawn up, we find that six digits are common, and ten digit numbers are by no means uncommon. What makes the situation worse is that most libraries have a lot of material on Engineering but relatively small collections on Logic – so the short notation is rarely used.

It must be remembered that, although we can make some provision for the growth of dynamic subjects, it is not possible to make sure that they will retain a brief notation indefinitely; as we have seen, any systematic arrangement will need revision over the years to keep pace with the growth of knowledge, and we are unlikely to be able to keep pace with this growth and still retain a convenient notation indefinitely.

A further factor affecting brevity is *synthesis* of notation. We have seen the contrast between enumerative and synthetic classification schemes; in the latter, only simple subjects are listed, and the classifier has to select the appropriate ones for any subject in hand and combine them according to the specified citation order. In practical terms, the notation for the individual elements is combined to reflect the composite subject, and this will usually lead to longer notation than if the symbols had been evenly distributed over all the required subjects enumerated, simple or composite. For example LCC uses two capital letters and four figures for most subjects in its schedules – though it is fair to point out that many composite subjects are not listed and must be classified with one of their elements.

UDC has often been criticized for the length of its notation. Because it was based on DDC5, and has been developed in areas of technology where Dewey's allocation of notation was inadequate to begin with, it frequently has long notation for single concepts; with synthesis built on to this base, the results can often be clumsy, repeating certain sections of the notation. For example, at one time the notation for the subject 'Power supplies for the electromagnet of a proton synchrotron' was:

621.384.61:539.185:621.318.3:621.311.6

Nobody can claim that this is brief or user-friendly, and it repeats 621.3 (Electrical engineering) three times. It was however specific, and UDC was the only classification scheme detailed enough to specify this subject and others like it. If we want specificity, that is, high relevance, we have to accept that the consequence will often be long notation. If the allocation of notation was poor to begin with, the situation is likely to be compounded. Eventually we reach a stage where the only solution is to revise the scheme, and this is discussed in Chapter 13.

It is often suggested that *mnemonics* in notation are an aid to memorability – as indeed they should be! *Systematic mnemonics* are found where the same concept is always denoted by the same piece of notation: for example the use of (410) for Great Britain in UDC, which has many mnemonics of this kind. In DDC we find mnemonics which fall into this category, but only in a limited way; for example, in Literature, Drama is always denoted by 2:

English literature	820	English drama	822
German literature	830	German drama	832
French literature	840	French drama	842

However, 2 does not always mean Drama, even within Literature. England is often denoted by 0942, but it may be shown by 942, 42, 042 or even 2, and while 0942 does nearly always mean England, the other symbols do not. We do not have the consistency needed for a piece of notation to be truly mnemonic.

Literal mnemonics are associated with the use of letters for notation; the theory is that by using the initial letter of a subject for its notation we shall find it easier to remember. Thus in BC1 C is Chemistry (but Physics is B); in LCC Music is M (but Fine Arts is N). This kind of mnemonic is so haphazard that it is of little value, and it certainly should not be used to affect the systematic order, as appears to be the case in the Generalia class in LCC.

In general, mnemonics are of minor importance. The general reader will not normally be aware of them, while the classifier will have little difficulty in remembering large amounts of the notation of a scheme regularly used, mnemonic or not. However, the use of systematic mnemonics takes on a new importance in searching computer-held files, since they can then be used to carry out searches which would be impractical with manual files. The MARC format for classification (Chapter 15) will make computer searching for notational elements much simpler, though we shall still have to be wary of unexpected pitfalls with schemes such as DDC.

Memorability is important, and the factors contributing to it must be carefully weighed when selecting a notation for a classification scheme. There is no doubt that much of the success of DDC is owed to its simple, easily understood and widely known notation, rather than to any theoretical excellence in its schedules. Despite this, it is important to reaffirm that notation is subsidiary to the needs of the schedules, and that it is possible to worry too much about the difficulties caused by long or complex symbols. Far more important is the need for the notation to possess other qualities, of which the most important is hospitality.

Hospitality

Notation shows the order of the schedules, but the schedules are merely a helpful way of listing subjects; since knowledge is not static, our schedules cannot be static; we must be prepared to add new subjects as they arise, in the correct place (as far as we can see it) in the overall order. The notation must therefore also be able to accommodate insertions, at any point where we may find it necessary to make them. We will most often need to insert a new focus in a facet, but we may occasionally have to accommodate a new facet, or even new basic classes. The notation must allow us to insert a new subject in the correct place: it must be *hospitable*.

If we are using Arabic numerals, we may use them as integers (whole numbers) or as decimals. Integers give a clear order which is known to everybody: 12 comes later than 2 but earlier than 115, for example. But suppose we have a series of foci in a facet, and we give them the numbers 1 to 7; if we now need to insert a new focus between the third and fourth, we cannot do so, for there is no whole number between 3 and 4. One solution is to leave gaps when we are allocating the notation originally, as is done in LCC, but of course this is merely postponing the time when we run out of places; there is also the temptation to insert new subjects in the schedule at points where we have left gaps in the notation, rather than in their correct, systematic places in the schedules. And since it is very difficult to foresee where new subjects will arise, we shall often leave gaps in the wrong places, but none where they are needed.

If, however, we use numbers as decimals, we can insert new symbols at any point in the sequence. Between 3 and 4 we can insert 31, 32, 33 . . . 39; between 33 and 34 we can insert 331, 332, 333 and so on. Now there is no longer any need to worry about leaving gaps in the right places, or to waste notation by leaving gaps in the wrong places. The facility of decimal numbers to incorporate new symbols at any point was seen by Dewey, and proved to be one of the most vital parts of his scheme – indeed, it gave the scheme its name.

The term decimal applies to Arabic numbers and relates to division by ten. The idea can of course be applied equally well to letters, where it will mean division by 26, so the term decimal is strictly speaking not correct, and instead we should speak of ‘radix fraction’. As the word decimal is widely understood, it is used in this text to apply to letters as well as numbers. In a letter notation, between B and C we can insert BA, BB, BC . . . BZ; between BB and BC we can insert BBA, BBB, BBC and so on. If we wish to have complete hospitality at all points, we must never finish a piece of notation with the first symbol of the base: 0, A or a; unless we follow this simple rule we shall find that we cannot insert new subjects at the beginning of a schedule. For example, if we use all ten digits 0 to 9, between 3 and 4 we can insert the ten numbers 30 to 39, but we then cannot insert anything between 3 and 30. If we do not use the first number 0, then we have nine left, so we can insert 31 to 39 between 3 and 4; if we want to insert a new focus between 3 and 31, we can use 301 to 309 and still retain complete decimal hospitality. We reduce the base by 1: 1 to 9, B to Z, b to z; in return, we gain hospitality at all points. Lack of hospitality is likely to lead to distortion of the schedules – notation dictating order – which we must avoid.

Expressiveness

Another quality which notation is often expected to have is *expressiveness*. This means that the notation reflects the structure of the schedules, and such a notation may be *structured* or *hierarchical*. A hierarchical notation reflects the genus-species structure of each hierarchy in our schedule of single foci; narrower terms have a longer notation than broader, while coordinate related terms have the same length of notation. A structured notation reflects the syntactic relationships in composite subjects, which may well involve the addition of *facet indicators* to our notation, as will be demonstrated shortly.

A hierarchical notation has the advantage of showing the structure of the classification. It is important to remember that the arrangement of books on the shelves, or items in a printed catalogue or bibliography, is a linear order which cannot show any structure other than the one-dimensional. We should use guiding to help users find their way around the arrangement, but a hierarchical notation can help. It also has the advantage that computer searching is made more obvious: to get greater recall, we shorten the notation we are using for our search, while to increase relevance we lengthen the notation. For example, we might begin a search in an OPAC by looking for English drama in 822; if we do not find what we want, we might broaden the search to 82, which will enable us to scroll through the whole of English

literature, substantially increasing recall, though at the cost of relevance. Alternatively, we might decide that we want Elizabethan drama, and lengthen our search notation to 822.3, decreasing recall and hopefully increasing relevance. An expressive notation facilitates this kind of search strategy, but it is not necessarily excluded by a non-expressive notation. For example, we may do a search for material on Swedish language at 439.7; if we decide that this does not find enough material, it is not too difficult to move up the hierarchy, but we have to know that Scandinavian languages are at 439.5, *not* 439. Again, the MARC format for classification will make searching easier by showing the steps of a hierarchy, whether these are reflected in the notation or not.

Unfortunately, we find that hospitality and expressiveness are mutually exclusive; sooner or later one or the other breaks down. The reason for this becomes clear if we consider a practical example such as the schedule for Engineering in DDC. In the first edition, Electrical engineering was not included, but it found a place in the second edition as a subdivision of Mechanical engineering. This might perhaps have been considered an acceptable subordination at the time, but it would certainly not be so now. We have also seen various other branches of engineering develop, all of which might lay a claim to be of equal status: nuclear engineering, aviation engineering, control engineering, car engineering, for example. Dewey realized that new branches of engineering might develop, and allocated 9 for ‘Other . . .’:

600	Technology
620	Engineering
621	Mechanical
621.3	Electrical
622	Mining
623	Naval
624	Civil
. . .	
628	Sanitary
629	Other

Clearly, car engineering and aviation engineering are similar to mechanical engineering rather than civil or sanitary, while control engineering, as one of the more theoretical studies applicable throughout engineering, surely belongs at the beginning of the schedule with the other theory subjects rather than with the more practical branches of engineering listed later. However, there is no notation left to show the status of these subjects, nor can they be slotted into the right place in the schedules; control engineering, car engineering and aviation engineering are found in 629, as ‘other branches of engineering’, while nuclear engineering is found in 621.48, which may be approximately the right place in the schedules, but hardly reflects the relative significance of this new basic class.

The fact is that immediately we start to require our notation to be expressive, we limit ourselves to an integral use of the final digit; in the above example, Dewey only had the numbers 1 to 9 to list all the branches of engineering and at the same time show their equal status. As we have already seen, an integral notation cannot

be hospitable; for the same reason, an expressive notation cannot be hospitable. Hospitality is more important than expressiveness, because it is the quality which allows us to govern the notation according to the needs of the schedules, instead of having notation dictate the order in the schedules. Ranganathan sought to overcome the problem by making the final digit of the base, 9 or Z, an *empty* digit; it could be used to extend the base by making 91, 92 . . . 991, 992 and so coordinate with 1, 2, 3 etc. This is of course a rationalization of Dewey's use of 9 for 'Other'. In CC7 it does enable us to insert new notation wherever we wish, but at the cost of losing any semblance of expressiveness; it is difficult to recognize that 2, 4, 91 and 991 are all coordinate pieces of notation.

We must also ask whether we *ought* to seek expressiveness in the notation. The purpose of the notation is to show the order of the sequence, of books on the shelves, or items in a bibliography, say; should we in addition expect it to show the structure which the sequence cannot show? Furthermore, we may find that the structure itself may cause difficulties by concealing the fact that certain foci are coordinate; consider the following schedule:

MUSIC

Individual instruments and instrumental groups arranged according to their basic mode of performance

Keyboard instruments

Piano

Organ

String instruments

Bowed

Violin

Viola

Plucked

If we apply an expressive notation to this schedule, Violin and Viola will have the same length of notation, which will be three digits longer than the general heading; Piano, Organ, Bowed and Plucked string instruments will all have a notation two digits longer than the general heading; while Keyboard and String instruments will have notation one digit longer. We can see that because of the number of steps in the hierarchy taken to define them, individual instruments – which presumably ought to be coordinate – will have differing length of notation. The problem is inherent in hierarchical classification, and we can find other examples in DDC; for example, in Metal manufactures 673, Tin 673.6, Mercury 673.71, and Magnesium 673.723 are all individual metals and are thus coordinate, but are reached through different steps of division.

We saw earlier that expressive notation can be useful in searching a computer database, since it enables us to move up and down the hierarchy, to broaden or narrow our search. However, because of the problem of accommodating more subjects than is allowed by the notational base, and also the different hierarchies which take us to subjects we expect to be coordinate, we find that problems may also arise with synthesized notation. For example, we saw that in DDC, 'drama' is represented by

2 in Literature: English literature is 820, English drama is 822. If we try to search for drama in whatever language by using an internal wildcard 8?2, this will be successful for the major languages as identified by Dewey: American, English, German, French, Italian, Spanish, Latin and Greek. It will not work for the rest, which DDC includes in 890; 892 takes us to Afro-Asiatic literatures. Looking at the schedules, we find that Russian literature is in 891.7, Russian drama in 891.72, so that our wildcard search has to be 8???? or 8?2 OR 89???2. 89?2 will take us to Sanskrit, 891.2, and Vietnamese, 895.2, literatures; Japanese drama, including Noh theatre, is at 895.62. The problem can be avoided in a scheme which uses facet indicators, such as UDC, where drama is -2; to search for all drama we can search for 8 AND -2 if our OPAC allows this.

Although the lack of expressiveness may make the overall arrangement that much harder to follow, we can help to overcome the problem by adequate guiding. It is also doubtful whether users are actually aware of the role of the notation in showing the structure; they are much more concerned with following the sequence of books on the shelves or entries in the catalogue.

Synthesis

We have briefly mentioned synthesis as one of the factors affecting brevity of notation, and it is worth re-examining this point in the light of the discussion on synthesis in Chapter 7. We saw there that coordination of single concepts was an extremely important device for improving retrieval performance from the point of view of relevance, and analytico-synthetic classification schemes are one important method of achieving coordination in an ordered fashion, according to a predetermined combination or citation order. By listing single concepts in the index vocabulary, and providing rules for their combination, we can give the classifier a much more powerful tool than the enumerative scheme, which attempts to provide in advance for composite subjects, but inevitably cannot foresee all that are likely to arise. In particular, we have seen that phase relationships are a form of coordination which cannot be predetermined, and *must* therefore be provided for by synthesis at the time of classification. The implication of this is that each single concept must have its own piece of notation, and that it must be possible to combine these pieces of notation – the *code vocabulary* – to specify any composite subject, including those involving phase relationships. We must therefore now consider in some detail the problems that arise if we try to synthesize notational symbols.

If we take the outline schedule for Library science that we constructed in Chapter 9, we can allocate an expressive notation, giving the kind of result that we have in Figure 11.1 column 1. Here we see that *History* (the generalized Time facet) is 3, and *Academic libraries* is 75; so the notation for *History of academic libraries* ought to be 753. But we can see at once that this will not do: 753 is the notation for *Technical college libraries*. We are of course trying to divide the heading *Academic libraries* in two different ways using the same notation: synthetic, *Academic libraries* AND *History*; and hierarchical, *Academic libraries* NT *Technical college libraries*. The same piece of notation could mean more than one subject. We have

to label not only the foci within facets, but also the facets themselves; if we do this, we can combine elements from different facets to denote composite subjects without causing confusion with hierarchical subdivision within the same facet. We shall have *hospitality in chain and in array*.

The problem is one that arises regularly in DDC, which has no specific facet indicators. Where we have a general heading with hierarchical subdivisions, it is not possible to use synthesis at that heading; by contrast, where we have specific subdivisions which are not divided hierarchically, synthesis is possible. In DDC17 the convention of using an asterisk to denote those places where synthesis is possible was introduced; an example will help to make this clear. In Agriculture we have the general headings:

633	Field crops
633.4	Root crops

Both of these are extended hierarchically, 633 obviously to include root crops and other kinds of crops at 633.1, 633.2, 633.4 and so on, and 633.4 to specify particular root crops, e.g. 633.49 Tubers, which is itself extended to specify Potatoes at 633.491*. To synthesize the notation for 'injuries to crops' we are told to add 9 to the notation for the crop, then add the appropriate number from 632 Plant injuries, diseases, pests and their control (the 'Problem' facet). Injuries to crops in general is 632.1, so to specify Injuries to root crops we would take the base number 633.4, add 9, then add 1 (from 632.1) to give us 633.491 – which is the notation for Potatoes! Because the notation is extended hierarchically (*in array*), we cannot extend it syntactically (*in chain*). However, when we get to the end of the hierarchical subdivisions, in this case at 633.491 Potatoes, we *can* synthesize a number for Injuries to potatoes because there will be no notational conflict. To 633.491 we add 9, then 1, to give 633.49191 Injuries to potatoes. The asterisk is used to show that synthesis is possible, and we are instructed not to use synthesis unless the base number has an asterisk. The only kind of synthesis permitted at all points is the use of the common subdivisions from Table 1, introduced by the zero 0.

How can we label facets so that we can synthesize notation unambiguously? We may find different kinds of notation used for different facets; for example, BC1 used lower case letters only for Place, while CC6 used them only for the common facets of bibliographical form and subject. In both cases it is possible to add these symbols directly to another piece of notation without confusion:

BC	Cricket (sports)	HKL
	Australia	ua
	Cricket in Australia	HKLua
CC	Physics	C
	Encyclopedia	k
	Encyclopedia of Physics	Ck

This method is clearly limited by the fact that there are only three kinds of notation we can use. Another method is shown in column 3 of Table 11.1 (*see pp.194-5*); this uses capital letters to denote the facets, with lower case letters for foci within

the facets. We can now combine the notation for the foci in a composite subject without any possibility of confusion. This was the kind of notation used by the CRG *Classification of library and information science*, and seen for many years in LISA. Arbitrary symbols may be used as facet indicators, for example in UDC, where we find (1/9) for Place, (01/09) for bibliographic forms, “. . .” for Time, and the colon : as a general indicator of relationship. In CC we find the comma , used to label Personality, semi-colon ; for Materials, colon : for Energy, dot . for Place and ' for Time. Synthesis is possible in both schemes, but clearly we have to lay down a filing order for these arbitrary symbols, both in relation to each other and in relation to the main notation.

Retroactive notation

The use of mixed notation or arbitrary symbols loses the great advantage of pure notation: its completely self-evident order. It also tends to make the notation more complex. Is it possible to have a pure notation which will nevertheless permit synthesis? We have had a hint of the answer when looking at the example from DDC earlier, when we saw that the subdivisions from Table 1 may be used anywhere, *because they are introduced by the zero 0*. The 0 is in effect reserved to act as a facet indicator, giving us the possibility of synthesis while retaining a pure notation.

If we have a subject with, say, three facets, we may use 1 to introduce the least important, which should file first. We can now use 2 to introduce the second facet, and combine notation from the two facets according to the citation order, *provided that we never use the figure 1 in the notation for the second facet*. Similarly, we can use 3/9 to introduce the third facet, the primary facet in this simple case, and still achieve complete synthesis, but we cannot use 1 or 2 in the notation for this facet. The penalty that we have to pay for the ability to synthesize within a pure notation is the progressive diminution of the base available. In the second facet above, the base is no longer 1 to 9 but 2 to 9; in the third it is 3 to 9, and so on. If we have nine facets, we might finish up with 9 as the whole of the base, giving 9, 99, 999, 9999 etc as the only pieces of notation possible in the primary facet! This is clearly unacceptable; we must begin by allocating an adequate amount of notation to the primary facet and work back from there. Further, because letters have a larger base than numbers, the method is likely to be more successful with notation using letters, and this is the kind of notation used in BC2. Because the elements must be combined in order working backwards (i.e. following the Principle of inversion) it is known as *retroactive notation*. Column 4 in Table 11.1 illustrates how the method may be used, and further examples will be found in Chapter 19. We may also find examples in DDC, where more than one zero is used to permit synthesis while retaining the pure notation. However, because of the length of the potential notation resulting from the use of this device, we are warned in the Introduction not to use more than one level of synthesis in such situations. An example from DDC20 is 350 Public administration, where 354 is for central governments other than the United States; we find that the facets are as follows:

0001 - 0009	Standard subdivisions
001 - 009	Administrative activities
01 - 09	The executive
.3 - .9	Specific countries from Area Table 2

The primary facet needs no facet indicator, and the three less important facets are introduced by one, two and three zeroes respectively. Theoretically, we could synthesize notation for a bibliography of legislation introduced by the Attorney-General of Australia - but with some 16 or more digits it would hardly be practical!

Table 11.1 Possible notational systems for the library science schedule

The four columns show how various kinds of notation might be allocated to the schedule for Library science. Since the schedule itself is tentative, so are the attempts at the allocation of notation. Column 1 is a simple expressive notation (cf DDC). The facets need indicators to permit synthesis. Column 2 is a non-expressive notation which tends to assume that the schedule is now fixed; it is usually shorter than 1, but less accommodating (cf LCC). Column 3 uses capitals for facets, lower case for foci, and is non-expressive (cf CRG). Column 4 is a non-expressive retroactive notation using letters; it does not need facet indicators (cf BC2).

<i>Schedule</i>	1	2	3	4
Bibliographic forms	1	10	A	B
Common subjects	2	20	B	C
revision	21	21	Bb	CC
research	22	22	Bf	CF
standards	23	23	Bj	CJ
automation	24	24	Bm	CM
economics	25	25	Br	CP
Time	3	30	C	D
Place	4	40	D	E
Operations	5	51	E	EZZ
administration	51	52	Eb	F
selection	511	53	Ec	FG
acquisition	512	54	Ed	FK
circulation	513	55	Ee	FM
technical services	52	56	Eh	FZ
cataloguing	521	57	Ei	G
catalogues	5211	58	Ej	GG
by physical form	52111	59	Ejz	GL
book form	521113	60	El	GP
classification	522	61	En	H
schemes	5221	62	Eo	HJ
UDC	52215	63	Er	HP
cooperation	53	64	Ew	J

finance	54	65	Ex	K
funding	541	66	Ey	KL
federal	5413	67	Ez	KM
Materials	6	69	F	LZZ
books	61	70	Fb	M
serials	62	71	Ff	N
periodicals	621	72	Fj	NN
newspapers	622	73	Fm	NP
non-standard	63	74	Fpz	NZZ
maps	631	75	Fr	O
records	632	76	Fw	P
Libraries	7	78	H	QZ
by subject	71	79	J	R
by mode of use	72	80	L	RZ
reference	721	81	Lb	SS
by population served	73	83	N	SZ
children	731	84	Nb	T
hospital	732	85	Nc	TT
handicapped	733	86	Nd	TU
blind	7331	87	Ne	TV
by kind	74	88	Q	TZZ
special	741	89	Qb	U
government	7411	90	Qc	UU
industry	7412	91	Qe	V
academic	75	92	Qh	W
school	751	93	Qi	WR
technical college	753	94	Qm	WV
university	757	95	Qr	X
public	76	96	Qu	Y
municipal	761	97	Qv	YR
county	762	98	Q	YS
national	77	99	Qx	Z

The schedule for particle accelerators introduced into UDC in 1961³ used retroactive notation, though this was not made plain at the time in deference to the large number of users who preferred to use UDC in the conventional way. Those who wished could use the schedule in the usual way, with the colon to link the notation for foci from the various facets, while those who wanted shorter notation could use it retroactively. This tentative experiment did work, but made no impression on UDC practice in general.

Though there is normally no need for a facet indicator for the primary facet, this may alter if we need to denote combinations of foci from subfacets; for example, a Children's reference library on art might have the notation 731,721,71(Art), where the comma enables us to combine the three foci satisfactorily.

We can see how the four systems compare by classifying two of the titles from the list, using the facet indicators from CC where necessary:

- 2 Baltimore County Public Library initiates book catalog
12 LaRoche College classification system for phonograph records

Notation	Title 2	Title 12
Column 1	762:521113;4(Baltimore)	757;632:522.4(LaR)
Column 2	98,60,40(Baltimore)	95;76:61.40(LaR)
Column 3	QwEID(Baltimore)	QrFwEnD(LaR)
Column 4	YSGPE(Baltimore)	XPHE(LaR)

These examples show that allocation has an important effect on length of notation, and that length itself is not the only factor involved in ease of use. In a fully developed scheme we would have a schedule for place, or perhaps 'borrow' one from a general classification. We may also use *identifiers* such as (LaR) if this is helpful.

Flexibility

If we use arbitrary symbols as facet indicators, we have to lay down a filing order for them, since there is no established sequence. This obviously has disadvantages, since the general user will no longer be able to follow the sequence unaided. On the other hand, it can have the advantage of allowing us to alter the citation order. As mentioned in Chapter 9, it is not always possible to find a citation order which will suit everybody, as shown by the two approaches to Literature exemplified by DDC (Language – Literary form – Period – Author) and LCC (Language – Period – Author – Literary form). If we use arbitrary symbols, as in UDC, we can use whichever of these citation orders we please, and we can even alter the schedule order to preserve the Principle of inversion, since we shall be determining the filing order of the symbols introducing the facets. This ability to change the arrangement by altering the citation order is one aspect of a feature of notation known as *flexibility*. Many schemes provide alternatives in specific schedules; for example, DDC permits us to shelve bibliographies with the subject, or all together, and to decide which arrangement we prefer for our Law books. In each case, there is an editor's preference, and it is this which determines which notation appears in the MARC records, but we may use the alternative if we wish. BC1 had a number of alternative schedules to give classifiers the flexibility to decide which they preferred. However, we should not forget that flexibility is a transient phenomenon; once we have decided which arrangement we prefer, all others must be excluded. We cannot vary our practice from week to week, and a change of citation and schedule order must be a rare event. Flexibility is not a primary consideration in notation; hospitality is far more important.

A classification of notations

We may work out a small matrix to clarify the various kinds of notation, ending up with a tabulation as follows:

Hierarchical but not structured	shows genus-species division but does not allow synthesis (e.g. DDC in parts)
Non-hierarchical and structured	permits genus-species division and synthesis, but does not display hierarchy (e.g. BC2)
Hierarchical and structured	permits genus-species division and synthesis and displays both (e.g. CC)
Non-hierarchical and non-structured	permits genus-species division but not systematic synthesis, and shows neither (e.g. LCC)

(A scheme like LCC does permit composite subjects to be included, and some are enumerated on nearly every page, but they must be fitted into the existing sequence, not denoted by notational synthesis. Hierarchies are shown by layout but not by the notation, except where it has been necessary to introduce decimal subdivision to accommodate new subjects.)

Because a synthetic scheme lists only single concepts, and has to provide means for notational synthesis to represent composite subjects, it makes the distinction between hierarchical and structured notation in the above table very obvious. We have discussed semantic relationships in Chapter 6 and syntactic in Chapter 7; we expect the notation of a classification scheme to accommodate both, but it is not possible for it to display both indefinitely. As we have seen, hospitality and expressiveness are mutually exclusive in the long run, and although CC claims both, it is at the expense of a number of devices which make the notation of the scheme much less practical than we would wish.

A cautionary tale

It is a fact of life that most people are not mathematically minded to any great extent. We have been discussing notation as a device to reflect the order of the schedules in a classification scheme, but for most library users it is solely a locating device. A piece of notation is found in the schedules or the catalogue, and the user can then go to the place in the classified sequence and find what is wanted by browsing the shelves, using authors and titles rather than notation. The kinds of notation which are accepted by people in general are telephone numbers and car registration numbers; in neither case are we concerned with order, but with *identification*. The only sequences which are readily accepted are alphabetical, as in a dictionary or telephone directory, and integral, as in house numbers. It is easy to look at the desirable properties of a notation from a theoretical point of view, but much harder to get users to make practical use of the result.

When *Library and information science abstracts* (LISA) began in 1969, the abstracts were arranged by the CRG classification of library and information science, as discussed in Chapter 9. In 1971 the citation order was changed, but the notation still consisted of upper case letters for facets and lower case for foci. Many users found it unhelpful, and the indexing was amended to give only a short notation sufficient to take the user to the general area, while the annual indexes, which

were not under quite so much pressure of time, indexed direct to abstract numbers. In 1976, computerization made it possible to index direct to abstract numbers in the bi-monthly issues, and in 1993 the notation was dropped completely. The abstract numbers are of course in a simple integral sequence beginning at 1 each year, which is familiar to everyone. We would expect the users of LISA to come very largely from the ranks of qualified librarians, who might be expected to cope with any kind of notation with equal ease, but this appears to be over-optimistic! If librarians find themselves ill at ease with a mixed alphabetical notation, we cannot reasonably expect the general public to find it acceptable.

A similar situation was found with the classification devised by E. J. Coates for the *British catalogue of music*.⁴ This was used from the beginning of the service in 1957 until 1982, when it was replaced by the 'Proposed revision of 780 Music' published as a draft phoenix schedule for DDC and formally adopted in DDC20. The new schedule was firmly based on the BCM *Classification*, adapted to fit into the general structure of DDC, using the more familiar decimal notation rather than the non-expressive alphabetical notation of Coates' scheme. Its adoption by BCM was part of the internationalization of BL services, but also a recognition of the fact that other kinds of notation are not welcomed by the users at large.

Summary

Notation is a device to mechanize the use of classified arrangement: to serve as a set of convenient ordering symbols, and to enable users to get easily from a subject expressed in words to the same subject slotted into its place in what is intended to be a helpful sequence. To get a notation which will give us hospitality in both hierarchical (semantic) expansion and in structured (syntactic) synthesis, we find it necessary to introduce some modifications to the kind of simple notation that is most easily accepted by users, but this conflicts with the need to provide users with the most simple access to information possible. With abstracting and indexing services, the computer can make the notation transparent to the user, giving a detailed classified arrangement with access through a separate easily grasped sequence of integers. We still have the problem of shelf arrangement, where simplicity is still essential if shelving is to be carried out quickly and accurately by clerical staff, who cannot be required to understand the niceties of a complex notation. It seems that we must accept a loss in specificity and thus a poorer relevance performance in order to have a practical shelf arrangement notation. However, the user does not find it too difficult to scan a shelf full of books, so perhaps concerns on this score are not acute. It is clear from experience that the majority of attempts to provide greater specificity by using an unfamiliar notation have been at the least unwelcome to the majority of users, to whom, in a reversal of the old proverb, unfamiliarity breeds contempt. With the development of the digital library, the computer may solve the problem by taking over the role of shelf arrangement as well as that of detailed subject access.

References

- 1 Most of the work on notation was done in the 1950s, and recent discussions have related to the use of class numbers in online retrieval, discussed in the chapter on OPACS. The following references cover all the essentials:
Coates, E. J., 'Notation in classification' in *Proceedings of the International Study Conference on Classification Research*, Dorking, 1957.
Foskett, D. J., *Classification and indexing in the social sciences*, London, Butterworths, 2nd edn, 1974, Chapter 4.
Vickery, B. C., *Classification and indexing in science and technology*, London, Butterworths, 3rd edn, 1975, Chapter 9.
- 2 Bliss, H. E., *The organization of knowledge in libraries*, New York, H. W. Wilson, 2nd edn, 1939, Chapter 3.
- 3 Coblans, H., Sabel, C. S. and Foskett, A. C., 'Proposed schedule for Particle accelerators 621.384.6', adopted by UDC 1961.
- 4 Coates, E. J., *The British catalogue of music classification*, London, Library Association, 1960.