Faceted Classification and Logical Division in Information Retrieval

Jack Mills

Abstract

THE MAIN OBJECT OF THE PAPER is to demonstrate in detail the role of classification in information retrieval (IR) and the design of classificatory structures by the application of logical division to all forms of the content of records, subject and imaginative. The natural product of such division is a faceted classification. The latter is seen not as a particular kind of library classification but the only viable form enabling the locating and relating of information to be optimally predictable. A detailed exposition of the practical steps in facet analysis is given, drawing on the experience of the new Bliss Classification (BC2). The continued existence of the library as a highly organized information store is assumed. But, it is argued, it must acknowledge the relevance of the revolution in library classification that has taken place. It considers also how alphabetically arranged subject indexes may utilize controlled use of categorical (generically inclusive) and syntactic relations to produce similarly predictable locating and relating systems for IR.

1. Introduction

As a memorable aphorism prefacing his novel *Howard's End,* E. M. Forster gave simply "Only connect." It could claim to be the finest, even though briefest, definition of intelligence we have. To understand anything, whether it is the operation of a complicated mechanism or the complex social factors that underlie almost any human situation, understanding it means seeing the connections. The basic intellectual instrument we use to do this is classification. It is appropriate that libraries, which seek to organize ev-

erything in the way of recorded human knowledge should find explicit classification as central to their organization.

1.1. Indexing and searching

Indexing and searching are the two fundamental operations in retrieval. The usual situation in the library is that the librarian prepares the scene for retrieval by indexing each document (assigning to them retrieval handles such as classmarks, subject headings, etc.). Searching may then be done directly, by examining the documents on the shelf or vicariously via their surrogates in the catalog. Although the term "indexing" is used with various connotations, especially ones involving terms in alphabetical order, the central meaning of pointing out or indicating describes exactly what librarians do when, in response to any enquiry, they indicate where the inquirer may best begin looking and, perhaps, where they might next look should the first search prove inadequate. This function is neatly summarized in cataloging theory as one of locating and relating.

1.2. Classification

This is the most fundamental operation in indexing. In its broadest sense, it is the action of recognizing and establishing groups of classes of objects, the subclasses and members of which all manifest (even though in different ways) a particular characteristic or set of characteristics. The different kinds of shared characteristic(s) used to define a class for retrieval have been called index devices (Cleverdon et al., 1966). Library classification, via shelf order and the classified catalog, uses a number of different devices; two of these reflect the sort of class definition usually understood by the term "classification"—those defined by generic and whole-part relations; but coordination (combination), synonym control, role indication (by inclusion of terms in facets defining their relation, such as agent, property), and some confounding of word forms (via their adjacency in the A/Z index) are also prominent. Mechanized retrieval systems developed a number of less direct devices, e.g., an extended confounding of word forms and oblique ways of defining a set of documents sharing the same subject content such as is found in citation indexing. Electronic systems have now extended these oblique forms of class definition (see Section 3.5).

2. What Is Classified in the Library

Library materials physically are the object of relatively rudimentary classification in that significantly different physical forms are separately housed and may be separately indexed. However, in nearly all cases it is their content which is their ultimate justification and the problems of information retrieval (IR) are paramount. Whether this content is best described as information or knowledge is best left to the philosophers. Early writers on library classification tended to use the term "knowledge" as the object of classification and retrieval. A dissident voice at the beginning of the last

century, when Bliss began opting firmly for a knowledge basis, was Wyndham Hulme (1911–1912). Hulme distinguished mechanical classification from philosophical and claimed that library classification belonged to the first kind. He coined a term "literary warrant" and described library classification as the plotting of areas preexisting in literature. This was, in fact, not a bad description of what the Library of Congress was doing in many of its classes but interpreting preexistence as being what they held in stock. When we consider content only, a major distinction is found in all general libraries (and some special) between subject content and what we may call, for lack of better words, "imaginative content." Much discussion of the exact nature of information reflects the unease over the use of the term "information retrieval" when it is clear that the content of a significant class of documents is not defined sensibly as information. The term "knowledge" appears to be somewhat more receptive to the inclusion of imaginative works than the term "information."

2.1. Subject content and imaginative content

The latter has received attention by and large only in respect of fiction (Beghtol, 1994; Hjørland & Albrechtsen, 1999). But the well-established dichotomy between fiction and nonfiction is somewhat misleading. Fiction is only one example of imaginative content; the latter includes also other literary forms (poetry, drama), all musical compositions, and all forms of the visual arts that can form the content of a record (e.g., a folio of paintings). If fiction offers viable characteristics of division whereby it can be organized, the same characteristics, in principle, should be applicable to all of them. The folio of paintings (say) might be classified by creator or by place (French paintings, etc.) or by period (twentieth century, etc.). But the above characteristics represent logical categories that are common to all kinds of record content. Music scores are classified by instrument (vocal, instrumental, etc.) and only secondarily by creator. But some characteristics might be thought to be special to imaginative works. For example, the new Bliss Classification (BC2) (see Section 5.2) includes in its Properties facet of Class W The Arts such terms as "didacticism, parody, sentiment, realism" and in its Elements facet terms like "symmetry, rhythm, symbolism, fantasy." By the process of specification (see Section 7.3), this allows imaginative works to be classified as didactic, parodic, sentimental, symmetrical, rhythmic, symbolic, fantastic, and so on. But many of these could also characterize subject content (in individual behavior, social behavior, technological work, etc.). In practice, much of the classification of imaginative works, especially fiction, is by subject content. But iconographic art (and its opposed nonfigurative or abstract art) inevitably uses the concepts making up a subject classification itself. The Subjects of art facet in BC2, for example, makes direct use of the whole classification, which gives a comprehensive and predictable order. Insofar as the classification of imaginative works

raises problems of cross-disciplinary and cross-cultural influences, it does not differ essentially from subject classification. The rules developed for the systematic handling of such relationships (see Sections 5/7) are as applicable to imaginative content as they are to subjects. Where imaginative content does present a special problem is that the categorization of a given imaginative work by some or many of the characteristics available would most likely be very subjective, and this factor almost certainly limits the degree to which they are practically feasible. But this does not mean that the rules present a rationalistic bias when applied to imaginative works, only that they are essential to the aim of achieving predictability in location, whatever the content of the record.

2.2. Common-sense view

The interpretation assumed in this paper of what exactly is classified may be described, for better or for worse, as the common-sense view of most librarians. The object of attention in library classification is the content of records; they will have embedded in them, to varying degrees, matters of fact (as Hume would say, in a famous phrase that, incidentally, begins with "When we run over libraries . . .") accompanied by considerations of analysis, discussion, prediction, opinion, and other matter (much of which might be considered to fall within the category of "relations of ideas") and other less concrete matter that may or may not be deemed worthy of inclusion in the index description. But if it does appear, it will be susceptible to logical division.

3. Levels of Indexing in the Library

A reader so far may have assumed that the catalog is the form par excellence of an index to the library collection and the prototype of indexes to larger collections and networks. This is not quite true. A library is indexed for retrieval at three levels: the systematic order of documents on the shelves (assuming complete or partial open access), the A/Z index to the classification governing the systematic order, and the catalog.

3.1. Shelf order

This is scarcely ever mentioned in the literature on retrieval, being treated very much as a poor relation, if not a terminally ill one. This is most unfortunate, since it is the very first index to the resources of the library for the great majority of library users and in many cases the main or even only one. Although this level of retrieval may be regarded as small beer and not deserving much attention, the special demands it makes because of its limitation to a single, linear order has had an important effect on the development of the theory of library classification. The limitation to a linear sequence throws into sharp relief a crucial property sought in indexing—that of predictability as to the location of any given class of information. The physical document can only go in one place. But the concepts that

define the class represented by that one place are in most cases multiple, e.g., a class represented by the rubric Bone—Cancer—Therapy—Radiography could legitimately go in any of twenty-four different places, everyone of them making sense. The expectations of users reflect this. A radiographer would like to see it under medical radiography; the cancer specialist would like to see it under cancer, and so on. The implication is clear. The classification must have comprehensive rules governing the order in which the different component parts of a compound subject are to be taken when locating a class. This does not depend in any way on the specificity of the index descriptions given to the documents; even if the classmark locating it is not specific (i.e., reflects "broad classification") the librarian and library user still need to know where it will go—under skeletal system, or therapeutics, or radiography, or cancer.

3.2. The A/Z index to the classification

The relative index that Dewey provided for his classification has been an outstanding example of this indexing component since the scheme was first published in 1876. It intuitively recognized that the central weakness of the classified index represented by the shelf order is that it distributes many subject concepts over many fields according to the rules for combination already mentioned. So, for example, literature on children will be scattered as a result of its subordination to different containing classes—medicine, psychology, education, welfare, and so on. Hence, Dewey's (1985) term "Relative Index" and the general use of the term "distributed relatives" to describe the situation.

3.3. The catalog

This consists of surrogates representing the records themselves, each surrogate containing, to a greater or lesser degree, a bibliographical description and rubrics to act as retrieval handles (indications of its subject, author, title, etc.). It has two central functions: first, as an inventory of the library's holdings; second, it provides for multiple access in searching (by author, title, form, or subject). Accessing by subject presents the central problem, and it is the subject catalog that is considered below.

3.4. Precoordinate indexes

Apart from a few special collections, this was the only form of subject catalog used until the 1950s. The term refers to the handling of compound subjects, which constitute the vast majority in the literature. The constituent terms that in combination (coordination) describe the subject are coordinated in the subject heading or classmark in anticipation of the needs of searchers. Compounding immediately raises the problem of distributed relatives; this problem, absolutely central to shelf order, continues to be central to the organization of the surrogates also, despite their much greater facilities for providing multiple access. How the separate concepts needed

to describe the compound subject are linked depends on the relationships subsisting between them, and these, in turn, determine the search strategies for locating the information sought. The problem of distributed relatives that this poses can be ameliorated (but never completely resolved) by making multiple entries for a document with a compound subject so that a separate entry appears directly under each of its major constituent concepts. For example, the document referred to earlier might get a separate entry under each of the four constituent concepts: Skeletal system, Cancer, Therapy, and Radiography (but omitting separate entries for the other twenty permutations theoretically possible). Such permutation is standard practice in libraries using the Universal Decimal Classification (UDC), whose notation particularly provides for it. Such permutation of multiple entries is rarely found in the alphabetical subject catalog. Notably, most subject cataloging takes as its unit a complete and discrete record (a book or article), and its classification and indexing involve a process of summarization. The subject description is of the record as a whole, and this determines its position.

3.5. Postcoordinate indexes

The development of mechanical aids to indexing (e.g., peek-a-boo, machine punched-cards) from the 1920s onward saw the removal of the need to summarize the overall content in a single precoordinated subject description. Now, only single constituent terms were assigned, and their combination to form a search request for the subject concerned was left to the search stage. This system was called postcoordinate indexing since the coordination appeared after the indexing step, requiring less effort since it moved the burden onto the searcher. The absence of recognized relationships could result in ambiguity, e.g., a search for fertilizers for sugar beet by the simple coordination of Sugar and Beet and Fertilizers would also produce documents on the use of sugar-beet tops as fertilizers. This led to the reintroduction of classification at the indexing stage in the form of role indicators and other devices that are implicit in the precoordinate index.

Mechanical aids were soon supplanted by electronic systems, and a still more drastic change in indexing practice followed. With the development of networks for electronic retrieval, the economic burden presented by the prior indexing of individual records (typically, for services operated commercially) became prohibitive. Now, it was not just a case of abandoning the intellectual precoordination of index terms but the abandoning of preindexing altogether. Reliance was to be entirely on keywords found in the record and recognized by electronic searching. Indexing devices developed by librarians can only be used indirectly, by assisting the framing of requests to search engines. The limited discriminatory powers of keywords, with all their attendant ambiguities in the unruly natural language, were now supplemented by new index devices, with machines operating on the

relatively raw text of the documents. All of them are based on the measurement of relatively artificial characteristics of documentary texts, such as frequency of occurrence of particular words, contiguity of particular words, etc., using statistical techniques and mathematical algorithms. These are deemed sufficiently correlative to conceptual meanings to form classes allowing searches defined conceptually. They constitute new index devices, but they are still classificatory in operation, establishing subclasses of the total store identified by the parameters of the technique used. They are not assigned by an indexer but must utilize the computer programs of the store's service provider.

The shift from IR from stores of limited size, in which trained librarians have prepared the field for searching by the prior indexing of materials, to much larger stores in which there has been only minimal preparation of the field has important implications for the relationship of libraries to information science. The cognitive processes connecting the producers of texts stored and the would-be recipients of the knowledge stored in the texts are the subject of much current research. However, the highly structured maps of knowledge developed by modern faceted classification apparently have considerable potential in assisting these processes.

4. Indexing in the Library Today

The inroads on the librarian's time made by the need to master rapidly developing computer techniques has had a particularly unfortunate effect on the curriculum of library schools, where the study of the organization of knowledge has been eroded just when the need for it has become greater. The information explosion led, inter alia, to the development by librarians of greatly improved index languages, largely based on facet analysis. The relevance of these to the future of the profession assumes two things: first, that the library will continue to be an integral part of our culture and that reports of the birth of the paperless society have been greatly exaggerated; and second, and following from the above, we have an obligation to seek the best possible ways of facilitating its work. The development of logically structured classifications covering the whole of knowledge is still unique in the field of LIS. These provide detailed maps of knowledge to assist in the searching of stores of records and can be used as the basis of, or valuable supplements to, numerous other retrieval languages.

5. The Design of a Modern Library Classification

Two conceptual areas must be distinguished: general classifications covering all knowledge and special classifications restricted to a specific field. The significant developments in classification design claimed above refer primarily to the second area and will be considered in detail under that. Here, some distinctive features of a general classification are considered.

5.1. General classifications

Remember that all special classifications need to draw on a more general one, often extensively. Another reason why IR cannot afford to ignore the concept of a general classification is that it alone can provide a bird's-eye view of the whole field of knowledge, offering a comprehensive context within which searches in a very large store can be framed. How the main classes (a loosely defined but reasonably well-understood concept) are handled within a general classification is the main theme of this paper. But whereas the central feature of the faceted special classification is its rigorous observance of the rules of logical division (see Sections 5.3/7), this cannot be said to apply initially to a general classification. If the first step in establishing what are loosely called its main classes were to be the division of the whole field of knowledge by applying explicit characteristics of division, the only feasible contenders would be of the nature of fundamental categories. The earliest and best-known set of such categories is seen in those advanced by Aristotle. Some of these are ostensibly feasible as constituting the initial divisions of the whole field of knowledge, e.g., substance, quantity, quality, place, time, and action. Such a first step has not been attempted by any of the general library classifications produced since Dewey's annus mirabilis in 1876, although something like it was attempted by the Subject Classification of the British librarian James Duff Brown (Brown, 1939/1906) with its quadruple division into Matter, Life, Mind, Record. Brown's scheme was notorious in its day for its subordination of music to sonics in physics—an example of its attempt to ignore disciplines as a primary level of division. What did emerge, with a relative unanimity that is not really surprising, was an initial division into main classes reflecting the division of labor—intellectual, imaginative, and practical. The division of labor is a fundamental feature of society, which is itself the producer of the knowledge in the records that are the objects of IR. It is manifested in every sphere of society, including academia as well as in the practical production of material wealth. The term "discipline" is frequently used to refer to these specialized fields, but is ambiguous insofar as a truly main class (e.g., the natural sciences) is usually susceptible to logical division into subclasses that are themselves known as disciplines.

The particular notion of the fundamental forms of knowledge that underpin main classes has received significant attention by Langridge (1976), who has drawn extensively on the work of a number of philosophers, particularly that of Hirst (1974) and of Phenix (1964) in the philosophy of education. Of particular significance is the distinction Langridge draws between the forms of knowledge on the one hand and the objects of knowledge (the phenomena they examine) on the other. The order in which main classes might appear became a particular focus of attention in the work of Bliss (1929, 1933), and a modified form of the general order he advocated is considered in Section 5.2.

A common criticism of the viability of any schema of universal knowledge is that the interaction of existing fields tends to dissolve their boundaries. While this interaction and its tendency are indisputable it does not invalidate the search for relatively permanent structures. Work on BC2 (Mills & Broughton, 1977–) has not found the great waves of new specializations an insurmountable obstacle. With enduring principles like gradation and integrative levels, together with highly practical principles such as the subordination of means to ends to reflect the concept of purpose or end-product to determine citation order within a given class (see Section 8.2), the predictability in the location of quite intricately mixed specializations is ensured. For example, modern forensic science draws on chemical analysis, molecular biology, and any number of medical specializations, but the purpose it serves—to validate the evidence in legal processes—determines its location in the law class with high predictability.

5.2. Two modern general classifications

The Colon Classification (Ranganathan, 1960) is not included here; its significance is primarily that it pioneered faceted classification and provided an experimental test-bed for its development. But its main-class order is quite conventional and offers no solutions to the problem of general classifications per se. The Broad System of Ordering (1978), or BSO as it is usually called, was first designed as a switching language—i.e., an intermediary through which other classifications could translate into each other. Its lack of detail stems from the fact that it was initially based on an institutional warrant—i.e., of subjects displaying institutional organizations underpinning them rather than on the much larger literary warrant of library collections. One feature is the break it makes with the generally recognized fields of knowledge, e.g., it has separate general classes for important concepts normally distributed under different contexts, e.g., Communication and information, Management, Human needs. It also has a Phenomena class (see Section 5.1) for works that cannot be accommodated in any of the largely disciplinary main classes, which are in BSO all fully faceted. It has also been very influential in the development of the next system, BC2.

For historical reasons, as well as theoretical ones, the BC2 (Mills & Broughton, 1977–) has largely taken the main-class order of the original Bibliographic Classification (Bliss, 1940/1953). This order reflects the Comptean principle of gradation and that of integrative levels (Feibleman, 1954; Foskett, 1961). The major sequence these give is modified in a few respects, as is shown in the outline in Appendix 1. BC2 has completely restructured all the individual classes, and each class is now fully analytico-synthetic in structure and notation. It is now virtually a new general classification and constitutes the most detailed, fully faceted general classification in existence. For this reason it is used in this paper as an exemplar of faceted structures, which are now (from the work done on it) seen to be appli-

cable to every field of knowledge. Like BSO, it also includes a separate Phenomena class, in which the order of phenomena closely follows the main-class order and uses the principle of unique definition to determine the location of multidisciplinary works on a given phenomenon. An outline of the system is given in Appendix 1.

5.3. Faceted classification of a subject field

This has been the major development in classification for IR in libraries in the past fifty years, although its first formulation was in the work of Ranganathan. Although, curiously enough, Ranganathan never referred explicitly to the fact, the fundamental feature of his Colon Classification is that it divides any given subject in accordance with the rules of logical division. But logical division is not the whole story. The work on BC2, covering every field of knowledge, clearly has shown that the design of a special classification requires recognition of six fundamental steps. These steps must of necessity be taken in the same order, since each step depends on the completion of the previous one. Only the first two use logical division; the other four use extralogical procedures. The steps are easily summarized:

5.4. The six fundamental steps in design are

- Division of the subject into broad facets (categories);
- Division of each facet into specific subfacets (usually called arrays, following Ranganathan);
- Deciding the citation order between facets and between arrays;
- Deciding the filing order between facets and between arrays and the order of classes within each array;
- Adding a notation;
- Adding an A/Z index.

5.5. The role of logical division

Before considering each of these steps in detail, the general role of logical division, which governs the crucial first two steps, must be noted. The rules of logical division, developed more than two millennia ago, are admirably brief:

- Only one characteristic of division should be applied at a time;
- Division should not make a leap; steps should be proximate;
- Division should be exhaustive.

The first and crucial rule is purely one of conceptual analysis and doesn't depend on practical considerations. The second and third rules involve to some extent subjective practical considerations as to the size of vocabulary to be accommodated and the degree of specificity with which compound classes are to be described. They are manifested only at the level of arrays (see Section 7). Observance of the first rule is the hallmark of faceted classification; a classification that fails to observe it rigorously throughout

the system cannot claim to be fully faceted. The operation of distinguishing the subclasses of a genus has been well-described by Broadfield (1946).

6. DIVISION INTO FACETS

The first step is to assign all the terms constituting the vocabulary of the subject into a limited number of broad categories. The use of the term "category" requires some explanation here. The outcome of the classification is an almost infinite number of possible subject descriptions of documents or parts of documents, nearly all of which will be compound classes —i.e., requiring two or more terms to summarize their content. For example, a document on radiographic diagnosis of bone cancer reflects four different categories of concepts in medicine; if the human body is seen to be the entity with which all medicine is concerned, bone is seen to be a Part, cancer a Process (an action internal to the body), diagnosis an Operation (an action performed on the body), and radiography an Agent of the operation. But the notion of Part is not a category in the traditional sense of the term, since it implies being a part of something—i.e., it is a relation, not a unique and independent category. Similarly, Agent is relative to the action it assists—it is a relation. So facet analysis might be said to be the assignment of terms to true categories (Time, Space, Matter, etc.) and to relational categories (Kind, Part, Agent, etc.).

6.1. Categories in subject fields

All or most of the categories will be found in all or most subject fields. Ranganathan was the first to see the need for initial categories. He provided five and called them Fundamental Categories—Personality, Matter, Energy, Space, Time (widely referred to as PMEST). He claimed that this order represented one of decreasing concreteness; so Colon displayed not only a template for logical division but also a citation order (see Section 8.1). The (British) Classification Research Group (CRG), formed in 1952, developed a more detailed set of categories, entirely consistent with PMEST in outcome but aiming to be more explicit—particularly in its interpretation of Personality; the set may be summarized as Defining system or entity, its Kinds, its Parts, its Materials, its Properties, its Processes, Operations on it, Agents of the Processes and Operations, Place, Time, Forms of presentation (of the information in the documents). The sequence above also embodies a citation order (see Section 8.1).

Assigning terms to categories is a deductive approach to concept organization, and it may be noted that one member of the CRG advocated and developed an inductive approach (Farradane, 1950). This he appropriately called relational analysis, since it is the relations between concepts that are at the heart of retrieval and categories are really a first step in recognizing those relations. Classifications resulting from Farradane's system proved to be remarkably similar to those of faceted classification.

6.2. Facet analysis

The operation of logical division in assigning concepts consists in essence of taking the whole vocabulary of the subject to be classified and asking of each concept, represented by a word or words, what category it belongs to in the context of the subject. This assignment to categories is simply another way of expressing how a particular characteristic of division is applied to obtain classes that share that characteristic, although in different ways (as division of objects by color will produce classes of different colors). The process is best explained by considering some examples of subjects and seeing how it handles every kind of concept.

6.3. Classification of "Politics"

When classifying the subject "Politics," a document may be found entitled "The British Nuclear Deterrent: For and Against." Taking Politics as the summum genus, we first decide on an acceptable definition of the class; this may be something as follows: Politics is the process in a social system (not necessarily confined to the level of the nation state) by which the goals of that system are selected, ordered in terms of priority, both ideologically and as to resources allocation, and implemented. Collectively, these functions often are summarized as being the exercise of power within the political system. Bearing in mind the categories already recognized, the title is analyzed to reveal the hidden concepts implicit in it; for the purposes of this demonstration these could be stated in a string: Britain—Foreign relations—National security—Weapons systems—Nuclear-Policy-Deterrence. This string reflects the following category assignments: Britain is a particular state; although it could be assigned to a number of different species of political systems (parliamentary democracies, monarchies, etc.), its logical status (as defining a particular political system) is technically that of a member rather than a species of the genus. Foreign relations reflects the Subsystems, or Parts category; although the term "foreign relations" sounds like a process, it reflects the main concern of an integral part of the wider process of governing the political entity Britain. This analysis is consistent with that distinguishing other major subsystems in politics (e.g., legislative systems) that are defined by the political process. National security in the context of politics is special to foreign relations and is treated as a Kind of such relation. Weapon systems represent an Agent used in the exercise of the process implicit in national security and Nuclear weapons represent a Kind of weapons system. Policy is regarded as one of a number of general activities or operations (in this case defined by the social objectives sought) that may apply at every level of political activity. Deterrence is a kind of policy, applied here to the process of national security.

6.4. Classification of "Medicine"

"Medicine" may be defined as the technology concerned with the actions taken by the human person to maintain their health and treat their

sickness. The definition of the subject leads directly to the primary category (the defining entity, the person), and all the other categories are realized in their relationship to this. The categories disclosed are

- Kinds of human persons (females, males, young, old . . .)
- Parts of the person (anatomical and regional, and physiologically functional subsystems—trunk, circulatory, neurological . . .)
- Processes in the person (normal physiology, pathology)
- Operations acting on the person (health maintaining or preventative, diagnostic, therapeutic)
- Agents of operations (medical personnel, instruments, institutions—hospitals, health services . . .)

So a particular document entitled "Rehabilitation Following Fracture of the Femoral Neck [in old persons]" would get the index description: Old persons (geriatrics)—Bone—Femur—Neck of femur—Fracture—Therapy—Rehabilitation

Medicine also demonstrates a situation where two fundamental forms of knowledge (here, the natural sciences and technology) may be said to merge in response to the demands of a classification for IR. This situation is sometimes said to be one of the signs that the concept of separate disciplines is breaking down. But nothing is new in this situation; whether we like to think, for example, of biochemistry as being a separate discipline or not, the central conceptual relation between the disciplines of biology and chemistry that meet in the class is clear: it deals with the chemical nature of living things. Chemistry here is a field of action serving the purpose of explaining biological phenomena and as such serves primarily the study of biology. It does not exist as a separate discipline outside the old-established two. Medicine as a technology may be defined as the application of knowledge and skills to produce an artifact of some utility—in this case, a healthier human person. It is inconceivable that the biological bases should not be seen as part of it. Such collocations are at the heart of the notion of helpful order that so appositely defines a main objective in indexing.

7. Division of a facet into its arrays

The classes constituting each facet are now organized into more specific subfacets (called arrays by Ranganathan). At the facet level, classes are undifferentiated and in most cases will not be mutually exclusive. An array consists of mutually exclusive classes. To achieve this condition, which is essential for the retrieval of a specific subject with a minimum of noise, these classes now must be differentiated by applying specific characteristics of division. For example, the primary category in building technology is Buildings, the entity reflecting the end-product or purpose of the technology. These are now differentiated by function (to give residences, etc.), by dominant material (timber buildings, etc.), by number of stories and so on. The

classes in the arrays so formed are now mutually exclusive; one cannot have a high-rise single-story building. But in some cases, certain arrays cannot be so easily named. For example, in the large Subject of law facet (substantive law), the first step of division gives three very large subclasses (Private law, Criminal law, Public law), each calling for further subdivision; the array of subclasses of the first includes Conflict of laws, Persons, Obligations, Property, Commercial law—all with numerous subclasses of their own. At this stage, numerous other characteristics still must be applied to distinguish yet more specific arrays; this is clear from the fact that the subclasses are not yet mutually exclusive, e.g., a compound class may be formed for torts of property (in which torts comes from the class Obligations). So the process of subdivision continues until characteristics are so specific that they generate mutually exclusive classes in an array, e.g., Persons by age, Persons by sex.

7.1. Division must be exhaustive

The constituent species collectively must be coextensive with the extension of the genus. The obvious difficulty encountered here is that of our imperfect knowledge. This can be overcome in a technical sense by the process of dichotomy, in which one species is named and all the others are covered by its negative, e.g., the array (Buildings by material) could give just two classes, brick buildings and nonbrick buildings, and this would exhaust the array—no buildings would be missed. In practice, of course, all significant kinds of other materials would be enumerated with a possible residual class for "Others."

7.2. Each step of division should be proximate

Division should not make a leap. Like exhaustivity, this is a counsel of perfection, which in practice is limited by imperfections in our knowledge. The price of failure is the obscuring of relations that might in fact be important in the definition of classes. Division of transport systems into road, rail, sea, and air obscures the relationship of road and rail as being kinds of land transport and of sea transport being a kind of water transport. In this example, more than one characteristic of division has been overlooked, e.g., land and water represent division by the characteristic of natural medium, but road and rail reflect the characteristic of form of track, which is special to land transport.

7.3. Special problems of division into arrays

As a faceted classification moves into more and more detailed analysis of a subject, more and more arrays are disclosed and some of these pose special problems. Several examples have been given already of the situation in which terms appearing in one facet (as properties, materials, parts, etc.) appear also in other facets in a different relationship. For example, the Materials facet in Building technology includes timber; this could qualify

a structural unit (e.g., timber for fencing). But it also could define a unit as being a kind of structure (e.g., timber houses). This relation is called specification (species-making). BC2 now generalizes this situation by assuming the possibility of terms from any facet behaving in this way, and this may be seen as a particular example of the general theory of analytico-synthetic classification. The distinction between qualification and specification was regarded by Metcalfe (1957) as a major feature of the relations found in indexing. At the most general level, it reflects the distinction between the inclusion relation (generic, semantic, hierarchical) and syntactic relations (see Section 12.3). It poses a particular problem in the entity (end-product, purpose) facet (see Section 8.3) but can appear in other facets, e.g., the concept of prefabricated bathrooms (those fabricated off-site and installed in toto in different kinds of buildings) reflects a part of a building (a room) specified by an operation (prefabrication). In BC2, wherever the need is demonstrated, the array reflecting the primary entity in a subject (e.g., in Building technology, the Buildings by function array) is preceded by a number of arrays derived by specification using other facets, for example, Buildings by detachment, Buildings by number of stories. In chemistry, the primary entity array (Substances by chemical constitution—i.e., elements and their compounds) is preceded by a number of arrays defined by concepts from other facets (Behavioral properties, Structural properties) and so on. In nearly all classes these other, derivative arrays appear in the same order as their defining facets appear in the class in general. In this respect, it has been noted (Coates, 1973) that a faceted classification provides a potent medium whereby newly emergent classes can be accommodated in a consistent and predictable fashion.

A further problem exposed by specification is that of dependent concepts. For example, in chemistry, the concept of allotropy might appear in the Properties facet, and by using it as a specifier it could generate the separate class of substances Allotropes. But allotropy is a property special to (dependent on) allotropes and should appear only under allotropes. In BC2, such dependent classes may appear in their basic facet as ghost classes, accompanied by a reference (e.g., Allotropy, see Allotropes). This situation does not occur in the example of (say) an operation like prefabrication; this could be used to specify a number of quite different objects in building technology (e.g., prefabricated bathrooms, as well as prefabricated buildings) and would therefore appear in the Operations facet in its own right.

8. EXTRALOGICAL STEPS IN CLASSIFICATION DESIGN

8.1. Citation order (combination order)

After logical division, this is the most important feature of a faceted classification. It may be defined as the order in which the characteristics

governing division of a class into its facets and arrays are applied. This in turn is reflected in the order in which the constituent terms/concepts (which together summarize the content of a document) appear in an index-description. This is seen most clearly in the rubric (heading) that represents a compound class in a specific alphabetical subject index (see Section 12.2); the designation "specific" here relates to subject headings that seek maximal precision (specificity) in describing a work's subject. Notably, the subject headings in most alphabetical subject catalogs are rarely precise enough to demonstrate this clearly; in a classified catalog, the full rubric for an entry in a medical library catalog (say) might represent a string of terms: Old persons: Bone: Femur: Neck of femur: Fracture: Therapy: Rehabilitation. Usually, in a classified catalog, only the term(s) representing the last steps(s) in the hierarchy are given in the heading, the others being provided for by the headings in the previous steps. The full rubric will appear in the A/Z index to the classified catalog, but in reverse order (see Section 11). Two crucial features of a classification system are largely determined by citation order: First, predictability in locating classes. The citation order decided must be observed consistently if predictability is to be achieved. Clearly, if documents on a disease are sometimes subordinated to the organ affected and sometimes vice-a-versa, the locating of classes becomes unpredictable. Before the appearance of Ranganathan's categories, a measure of consistency was attempted by sets of pragmatic rules, exemplified by Merrill (1939) in his *Code for Classifiers*. The advent of comprehensive category-based rules has now made such selective rules largely redundant. Second, helpful order: This refers primarily to the helpfulness of the collocations it produces—what is kept together and what is scattered by subordination to other concepts. The number of different ways of classifying a subject is so huge that it would be rash to say that one order is better than all the others. But the one decided upon should be one of which it cannot be said that another is better.

8.2. Citation order of facets

The primary facet in a subject represents a summum genus and the other categories at the facet level clearly reflect the different relationships that concepts may have to it. For example, in the class Building technology, the primary facet is that of Buildings. Terms in the other facets always imply the relationship of the concept represented to buildings, e.g., weather resistance in the Properties facet means weather resistance in buildings; sill in the Parts facet means a sill in a building (usually in some kind of opening). These relationships provide a clear and powerful basis for the citation order. Agents serve the operations that may act on the processes or parts or kinds of the defining entity; the processes are inherent in the parts or kinds; the parts belong to the kinds; properties may belong to any of the foregoing and therefore constitute a sort of floating facet, qualifying whichever category they belong to.

The problem of citation order was first tackled by Ranganathan in his Colon Classification (see Section 6.1). His five fundamental categories (PMEST) represented a citation order of decreasing concreteness. While the practical demonstration of the categories and their order in Colon made them reasonably clear, the CRG sought to develop a more detailed set of categories, entirely consistent with PMEST in outcome, but more explicit, particularly in its interpretation of the concepts Personality and Energy; like PMEST, they were presented in a citation order that may be summarized as Defining system or entity, its Kinds, its Parts, its Materials, its Properties, its Processes, Operations on it, Agents of the Processes and Operations, Place, Time, Forms of presentation. In seeking to explain the relations more fully, the defining system came to be seen as reflecting the end-product of the subject in that the other categories are all seen to be features of it or actions directed at producing or sustaining it. The production of this end-product, whether by natural forces or by human actions, is seen as reflecting the purpose of the subject and the overall sequence reflecting the general principle of the subordination of means to ends. Like "only connect," this principle (which may be seen as a species of the first principle), reflects a quite fundamental element in the perception of relationships.

Several other systems have been developed, primarily for specific alphabetical indexes, which incorporate comprehensive rules for citation order, articulated by the relations between the terms in the heading. These are considered in Section 12.3.

8.3. Citation order between the arrays in a facet

The powerful rules for citation order described above operate only to a limited degree when deciding citation order between arrays. This is usually thought to be a weak element in the theory of faceted classification, seen as the essential basis of a fully predictable linear order. But this criticism needs to be qualified by a number of factors, and notably it has not proved to be a serious problem in the comprehensive testing ground provided by BC2. The nature of the compound classes demanding a ruling varies greatly with the subject concerned and would in any case rule out consideration of an immutable rule for arrays in all subjects. The principle of purpose or end-product in the facet formula continues to operate, e.g., in the Buildings facet of Building technology, the array (By function) is cited first; in any Materials facet, the array (By constitution) will cite before arrays reflecting other facets (e.g., By property). The principle of decreasing concreteness leads to the array defined by membership rather than class being cited first (e.g., in many social sciences—politics, law, etc., where the nation state defines the first characteristic of division).

Special (implicit) arrays and derivative arrays. The arrays in a facet usually fall into two groups; those that are special or peculiar to the facet and define it and those that are derived by specification (see Section 7.3), e.g., in Building technology, the first-cited array in the Buildings array is that of

Function, to give houses, prisons, etc.; this clearly defines the purpose and is special to buildings. Other arrays include one characterized by predominant material; this is a derived array, with specification by terms from the Materials facet.

Derivative arrays. These are not implicit; e.g., Prefabricated, as a difference "added" to the species Buildings, to give the subclasss Prefabricated buildings, is not implicit in the species Building. Things other than buildings may be prefabricated—e.g., furnishing units. The concept Prefabricated derives from the operation of prefabrication, which is located in the Operations facet. This feature characterizes all derivative species—they are all derived from other facets. To meet this situation, BC2 now provides classes with the facility to use all the other applicable facets in the role of specifiers. Naturally, the order in which the donor facets are taken will be the order they already have in the facet citation order. But in nearly all cases, these arrays will be cited after those arrays that are special to the primary facet. Similarly, the numerous arrays defining kinds of semigroups in algebra are cited according to the status of their definition in terms of the categories reflected, whose citation order has already been determined by their categorical status. So Semigroups by system (matrix semigroups, topological semigroups, etc.) are cited before Semigroups by property (linear, finite, etc.) and these before Semigroups by relation (inverse, etc.) and these before Semigroups by operation (multiplicative, etc.).

A simple example of how the above problem can occur at any level of the hierarchy is that of Leatherwork in the Decorative arts class. The latter is defined in many cases by the material used, giving silversmithing art, textile arts, etc. This demonstrates the fact that a defining array itself can sometimes reflect another facet, just as when Place features as the primary facet in classes like politics and law. The same principle holds when the array (By kind of leather) is taken as a defining array in Leatherwork, whereas the array (By technique) is derivative, giving, e.g., embossed leatherwork.

8.4. Problems of citation order in array

The absence of a comprehensive general formula for citation order between arrays can, however, present special problems on some occasions; a prominent example is found in the classification of the Arts; in analyzing the literature to determine what categories and arrays to recognize, a document might be found entitled "The Romantic Landscape in 19th-Century British Painting." Assume that a working definition of the arts has already been made: that branch of creative activity concerned with the production of works characterized by imaginative design and expression and in which aesthetic considerations predominate. The concepts in the title, taken in turn, might then be defined in terms such as Romantic designates a movement in art that reflects a commitment to feeling rather than intellectual discipline (and so on); Landscape refers to an art (most often in painting)

defined by its subject matter; nineteenth century defines the art of a cultural period; British defines a society or culture in which the art was produced; Painting defines a particular medium. The trouble is that all these reflect the primary facet of Kinds of art. Romantic is a Kind of art defined by a style, movement, or school; style reflects concepts from the Properties and Elements facets (e.g., didactic, eclectic, realistic, symbolic, fantastic); Movement and School both imply concepts from Place and Period. These three concepts overlap so seriously that they cannot bear the burden of being separate arrays, although provision is made for general works on each of them. The concepts of Landscape art, Nineteenth-century art, British art, and Painting are clearly all legitimate claimants to the status of kinds of art. When it comes to deciding the citation order of the four arrays, several considerations arise. The working definition clearly implies that the work of art produced gives us the entity we start with. Also, the properties characterizing the work clearly imply a human creator, and this facet, the artist, could be construed as the primary one. But consideration of the role of the division of labor in the classification of knowledge (see Section 7) combined with the fact that the vast majority of artists operate in a special medium suggest that the medium should be treated as the primary facet. This is reinforced by pragmatic considerations of helpful order. It is inconceivable, for example, that music should be cited after any of the other arrays. This would mean citing La Mer or the Enigma variations, say, under Subjects of art (landscape, portraiture). But the citation order of the others is less clear. If the artist is seen to define the obvious second array, the importance of the culture in which the artist produced his or her work suggests that place and time also may be serious contenders. Here, the decision that a general classification must make may not meet the demands of all its users, and the provision of alternative citation orders becomes desirable.

8.5. *Alternative citation orders*

The problems posed must be seen in the context of the purpose of library classification, which does not seek to educate the specialist (in the above case, art historians and art critics) in the structure of their subject but to provide an instrument that assists the ready locating and relating of records according to their content. It also emphasizes that the arrangement within a subject in a general classification may not serve the needs of a special collection, e.g., a college library may want its arrangement to reflect as far as possible the curriculum in the subject as taught in that college. The original Bliss classification was notable for its provision for alternative arrangements to meet this problem. BC2 has followed and extended this policy, and it is worth noting that a number of college librarians using it prefer to use some of its alternative arrangements for the very reasons mentioned. In this way, they enjoy the comprehensive analysis, vocabulary, and notation of the general scheme and yet manage to fit it to their special needs.

9. Filing Order of Classes

This is the sequence in which the individual classes, simple or compound, file one after the other in a linear order. It is quite different from citation order. The latter is analogous to the order of constituents in a telephone directory entry—Surname, Forename, Designation (Dr., Sir, etc., perhaps). But whereas the second dimension in the directory (the A/Z filing order of the names) has nothing in common with the first in the manner of its construction, this is not so with the classification, in which filing order is determined to a large extent by the citation order. Filing order has two quite separate components: first, the filing order of the facets and arrays when each facet and each array is treated as a single block of classes, and second, the filing order of the individual classes within each array.

9.1. Facet filing order

This is the order in which the individual facets (each regarded as a block of classes) file, one after the other. It is usually the reverse of the citation order, i.e., the first-cited facet files last, the second-cited facet files next to last, and so on. This is entirely due to its need to observe a general before special order.

General before special (decreasing extension). This principle is quite independent of faceted classification. It is considered here because its implementation requires what is called an inverted order in the filing of the facets and of the arrays within them. It is defined thus: a class that completely contains another class should file before that class. The observance of this rule seems to be almost a universal expectation; perhaps it reflects a folk-awareness of the holistic principle of distinguishing the wood from the trees. For example, a work on marketing is expected to file before one on the specific forms of marketing (retailing, etc.) and a general work on retaining before its specific forms (self-service retailing, franchise, etc.).

The inverted schedule. To observe general before special necessitates a design feature usually referred to as the inverted schedule; we use the layout of the printed schedule here to demonstrate the problem because all librarians are familiar with the situation whereby the classification is laid out in schedules before it is translated into the linear order of classes manifested on the library shelves and in the classified subject catalog. For example, in a medical classification the first-cited facet (Kinds of persons) files last; the second-cited facet (Parts of the body) files next to last, and so on. As a result, a work on the skeletal system of old persons would file not only after old persons in general, but also after the class Skeletal systems in general. If the schedule were not inverted, the special would file before the general.

9.3. Filing order of arrays

In the filing order of the arrays within each facet, the situation is exactly analogous to the filing order of facets; the arrays file in an order that is the reverse of their citation order, e.g., in the Building technology class,

the first-cited array is the array (By function) and it files last; the second-cited array (By attachment) files next to last, and so on.

9.4. Filing order within arrays

This is another problem, quite distinct from the filing order of the arrays (as blocks of classes) themselves. An array results from the application of a characteristic of division so precise that its subclasses are mutually exclusive; so it does not contain compound classes and the problem of general before special doesn't arise. Numerous helpful orders in array have been identified: operations are often given in order of performance (e.g., preparation of soil, sowing, protection of crop . . .); this is really a special example of chronological order, which is a major feature in many classes in the Humanities; geographical (contiguity) order is also a major one in many arrays besides its role in the filing order in the Place facet. For some arrays, no obvious systematic order of its classes is applicable, and these are arranged alphabetically.

10. NOTATION

This assigns to each and every class in the system a symbol (classmark) that possesses or is given an ordinal value; this locates any class mechanically, without the user having to know its hierarchical position. Although this has nothing whatsoever to do with the problems of concept analysis and knowledge organization per se, it is an essential feature of a library classification. Moreover, numerous misconceptions tend to persist that impede the understanding of the conceptual arrangement. So the problems of notation are considered here in more detail than would otherwise be justified.

10.1. Functions of notation

Notation may be defined as a system of ordinal symbols that mechanizes the order of classes in a bibliographical or other linear classification. For example: SL9 H is the classmark in BC2 for the subject Appellate proceedings in common law systems. Assuming that users know the ordinal sequences A/Z and 1/9, the only rule they need to know is that in BC2 a number files before a letter. They can then locate the class exactly in the largest of law collections and can do this mechanically, without knowing the conceptual hierarchy in which the class occurs; in the example above, this is

> Law [S]—Legal systems [SCY]—National systems [SHY]— Common law systems [SL]—Practice and procedure [SL6]—Courts & court procedure [SL6 E] — Actions, lawsuits [SL8]—Trials, hearings [SL8 S]—Trial procedure [SL8 ST]—Judicial decisions [SL9 D]—Remedies [SL9 G]—Administrative remedies [SL9 GV]—Appeal, appellate proceedings [SL9 H].

This example demonstrates several points about notation. First, it in no way determines the order of classes or the location of a particular class. The latter is determined completely by the concepts defining the class and the

rules for citation order and filing order. Notation is simply a servant, using our common knowledge of the sequence conveyed by the basic ordinal symbols used to represent the classes. Second, the quite secondary function of "expressing the hierarchy" is neither necessary nor in a bibliographical system, possible, in that the burden of adding a notational symbol for every step of division in the hierarchy would be quite insupportable. That some systems (DC, UDC, and to some extent Colon) claim to have expressive notations is misleading in that their notations are expressive only up to a point. This is like saying a chain is strong except that some of its links are weak. Whether a given classmark in such systems is truly expressive is quite unpredictable. Third, a major advantage of a nonexpressive notation (often called an ordinal notation because it seeks only to serve the central function of notation) is that it greatly simplifies the allocation of notation and the accommodation of new classes. Fourth, it makes possible much briefer classmarks; this is demonstrated by the example above of Appellate proceedings in common law, in which a classmark of four characters represents a conceptual sequence of twelve hierarchical steps following the main class S Law. A fully expressive notation would require at least thirteen characters.

10.2. Qualities of notation

These are described in detail in a number of textbooks and articles and need only the briefest consideration here. The two basic qualities are simplicity and hospitality. The first depends mainly on the types of symbols used and on brevity, both considered above. The second, hospitality, is the ability of the notation to accommodate whatever number of classes demand a distinct classmark. In a faceted classification, this means the ability to assign a unique position to any compound class called for; theoretically, any class may be combined with any other class other than the mutually exclusive classes in its own array. This implies that the notation must be able to provide for all these. So, just as the conceptual structure is called analytico-sythetic, a faceted notation is called a synthetic notation (the analytico component being the preserve of the conceptual classification). The central problem now is how to provide for the linking of any class with any other while maintaining completely the conceptual order designed for the hierarchy. One way of doing this is to use explicit "facet indicators" as in UDC and Colon (e.g., arbitrary symbols like ( )." ", :, -.).

An alternative method is known as retroactive notation as used in BC2. Since BC2 is used in this paper as the main vehicle for demonstrating all features of faceted classification, a brief account of retroactive notation is given here. The principle was (once again!) first used by Dewey, who reserved the zero to "introduce" three different facets (Bibliographic form, Period, and Place). So the first special subject division in a class is usually given the next digit (one) since the zero is reserved, e.g., 61 is the first sub-

ject division in class 6. In BC2 this principle of reserving earlier symbols that can then be added directly to the classmark they are qualifying is the main device for notational synthesis; it is called retroactive because synthesis in an inverted schedule is nearly always effected by qualifying one class by earlier classes—i.e., working backward (retro) in the inverted schedule, e.g., in Class S Law the following classes are found:

- Damages (from the class Legal actions)   S9M;
- Personal injury (from the class Tort)   SBGQR;
- English law (from the class Common law)   SN.

For the subject Damages for personal injury in English law, the classmark, built retroactively, is SNG QR9 M. Note that (1) A special provision for national jurisdictions allows all the classes in SB Substantive law to be added directly, dropping the two initial letters SB. (2) A convention to assist the easy reading of classmarks is to give the classmark in spaced multiples of three.

10.3. Hospitality to new subjects

This is, of course, an important conceptual problem (see Section 5.1), but it is often considered in notational terms. A classification system, regarded purely as a sequence of terms representing a hierarchy of conceptual classes, has no difficulty in inserting new classes once it has decided where they logically go. Just how the notation can accommodate it exactly at that theoretically desirable point is another problem. Ranganathan described it as one in which "notation brings rigidity." Remember that Ranganathan assumed an expressive (hierarchical) notation in which rigidity is certainly a major problem. In an ordinal notation, the only problem it poses is that of brevity for the new class. It does not have to bother about what the classmark looks like in terms of expressing the hierarchy.

11. The Alphabetical Index to the Classification

The A/Z index to the printed schedule was mentioned briefly in Section 3.2. Here, the relations between the A/Z entries, using the natural language, and the conceptual hierarchies governing the classification itself are briefly considered. The A/Z index performs two essential functions: it provides the user of the classification with a key, linking the natural language terms for the classes to the classmarks that locate them; it complements the systematic display of relations in the hierarchy by showing under any term the distributed relatives.

The main problem is the enormous number of compound classes that are theoretically possible in a faceted classification (or even a largely enumerative one like DC), which makes it quite prohibitive to show all the distributed relatives in the A/Z index. The optimal solution to this is to recognize that the classified index and the A/Z index complement each other and that a fair division of labor is possible between the two parts. This solu-

tion is found in what Ranganathan called chain indexing (see, e.g., Mills, 1960). This has one fundamental rule—that a term in the A/Z index should never be qualified by one of its own subclasses from the classified hierarchy, e.g., using BC2, an A/Z index entry:

    Appellate proceedings   S9H I

would not be followed in the A/Z index by

    Appellate proceedings: Right of appeal   S9H I

because the latter will be found in the classified sequence, following S9H. If it is sought via the latter, it will be found there, at

    S9H I Right of appeal in appellate proceedings

Chain indexing is a highly economical method of constructing an A/Z index, since it does not duplicate work already done in the classified sequence. It is necessary to distinguish here the printed index from the classification schedule and the much fuller index that may be provided to the collection of a given library system or to a special bibliography or national bibliography. So, for example, although no entry will appear in the printed index to Class S in BC2 for

    Appellate proceedings: Scots law   SOB 9H

a classified catalog to a law collection would include the entry if the library had literature on Scots law. The other major feature of chain indexing is that it automatically provides a coherent and predictable order of the terms qualifying the lead term. This order is the reverse of the hierarchy, e.g., in BC2, the entries generated for the specific subject

    Old persons–Femur neck–Fracture–Rehabilitation

in a medical index would be

    Old persons   HXW
    Bones: Old persons   HXW TKX
    Femur: Old persons   HXW TNP
    Neck of femur: Old persons   HXW TNP SR
    Fracture: Neck of femur: Old persons   HXW TNP SRN DL
    Rehabilitation: Fracture: Neck of femur: Old persons   HXW TNP SRN
       DLG TR

This order of terms in each entry may be compared with the order most likely to occur in the natural language statement of the subject as determined by the syntax of the language:

    Rehabilitation [after] fracture [of the] neck [of the] femur bone [in]
       old people.

It is clear that the standard citation produces structures that closely parallel, in reverse order, those of the natural language.

    Before leaving this example, it is worth noting that the rather daunting length that classmarks can reach reflects not on the order or notation of the classification but on the specificity it aims at in subject description. Even relatively broad classifications like DC and LC occasionally reach the length of classmark shown above, but for less specific subjects.

12. Classification and Alphabetical Subject Catalogs

This term is used here to stand for any index to the information content that is alphabetically arranged and is independent of any classified arrangement. This raises the twofold difficulty that catalog searchers have to recognize: Just what are the concepts involved in the subject they seek, and how can they cope with the vagaries of the natural language when phrasing that need for searching? While its basic principle is to give the user known names in a known order (to use Metcalfe's phrase) every practicing librarian knows that this is only the second step. The first requirement is for users to know just where they want to get to; for this they need a map of the subject terrain, showing exactly where the numerous sideroads branching off the main highway lead to.

12.1. Subject headings

The original form taken by these is familiar to all librarians and is exemplified for general collections by the Sears and Library of Congress subject headings. The first feature to be noted is the absence of any serious provision for the specificity demanded by a special collection. This inevitably impairs its ability to locate subjects precisely. The second is the relative arbitrariness in the provision made for the relating function. While the indication of broader and narrower terms inevitably invokes the classification, the choice of terms related in ways other than in generic and partitive hierarchies is usually highly pragmatic and unpredictable.

12.2. Specificity in subject headings

If a subject heading is to provide for the multiplicity of relations considered earlier under faceted classification and that frequently arise now even at the level of books and monographs, predictability in locating demands that comprehensive rules must be observed governing the citation order of components in any given string of terms. The general principle observed is said to be that of immediate access via the sought term. But this only begs the question as to what that term is when faced with even quite simple subjects; e.g., an inquirer looking for child psychology looks under psychology of children; an enquiry for works on the economic history of Britain in the Victorian period poses immediately which of the six likely (or twenty-four possible) combinations of terms should be tried first.

12.3. Relator systems

Since the 1950s, several different systems have been developed, each using their own set of rules for citation order. The term "relators" is often used to describe the conceptual relationships underlying their rules and the symbols that may be used to signal those relationships. The main systems are Farradane's relational analysis (Farradane, 1950), SYNTOL (Gardin, 1965), the British Technology Index (BTI) (1962–; Coates, 1960), and PRECIS (Austin, 1984). The latter (its name standing for Preserved con-

text indexing system) was originally designed as an alphabetical index to the classified British National Bibliography (BNB) but with particular regard for the way in which this might be computer-assisted. The syntactical strings it developed were later applied to free-standing alphabetical indexes. The interaction of classification, categories, and relations is analyzed in a key paper by Coates (1973). The distinction between categories and relations in the context of the classified index, was considered briefly in Section 6.1. The central problem in the case of specific alphabetical subject headings is essentially the same. To achieve predictability in locating, rules for citing the terms in a compound heading must be strictly observed. Clearly, many of the same rules as those described in Section 8.1/3 can be applied. The resulting strings can be seen to consist of a leading term, however arrived at, followed by the other terms according to their relationship to that leading term. A very practical advantage of this articulation of relationships independently of any given classification system is that a special library can set up such an index with minimal recourse to existing index languages (Coates, 1973).

The conspicuous absentee in the alphabetical subject catalog is the inclusion relation, generic or partitive, which is the bedrock of the classified index. Ideally, systems like those above would be supplemented by a comprehensive classification with a structure compatible with the principles of the relator system. Perhaps because, theoretically at least, the specific alphabetical subject heading reflects the natural language more closely than that of the classified index, the terminology used for them also reflects linguistic terms; the relations between the terms are variously called syntactic, syntagmatic, and analytic. The terms used to describe the generic inclusion relation are variously semantic, paradigmatic, generic, and hierarchical; the concept of specification (see Section 7.3) is called predication in SYNTOL and differencing in PRECIS.

12.4. The thesaurus

This is now well-established as an IR tool that provides a controlled language for postcoordinate systems (although it is possible to conceive its structure being accommodated within that of the A/Z index to a classification). Because its use of compound terms (bound terms) is severely limited, the problem of citation order is minimal. The situation regarding connectives is almost identical to that in the conventional lists of subject headings but is usually treated in much more detail. The inclusion relation is covered by BTs and NTs; the scale of provision of other relations (associative relations) is less predictable. The relevance of an attendant faceted classification system is obvious, and this is considered in general terms by Aitchison, Gilchrist, & Bawden (1997) and specifically in relation to BC2 by Aitchison (1986). Fugmann (1994) gives a lengthy and illuminating review of a special thesaurus utilizing classificatory principles.

APPENDIX 1
OUTLINE OF BLISS BIBLIOGRAPHIC CLASSIFICATION (BC2)
* See Section 5.2.
* 13 volumes have been published; 2 are in the press. All other classes have
    detailed drafts in an advanced state, awaiting finalization before publi-
    cation.

Introduction & Auxiliary Schedule

	* Common facets for Form, Time, Place, Languages, Ethnic groups.
2	Generalia
3	(Objects of knowledge, phenomena classes)
	*Subjects treated from a multidisciplinary or nondisciplinary point of view: Properties, Processes, Entities (mainly materials and organisms), arranged by their unique definition.
4	Prolegomena to a universal classification* The field of knowledge itself is the subject. Universe of knowledge.. Methods of enquiry.. Information skills (Forms of knowledge, disciplines)
5	(Operations on information) Data processing.. Computers..
6	Recorded knowledge, library & information science & technology
A	Philosophy & logic... AM Mathematics & statistics
AY	Science & technology in general.. Science.. Physical science..
B	Physics.. C Chemistry.. D Astronomy.. Earth sciences..
E	Biology.. Microbiology.. F Botany.. G Zoology
GS	Applied biology.. Plant & animal husbandry.. Human ecology
H	Human biology.. Physical anthropology..
HH	Applied human biology.. Health & medicine.. I Psychology..
J	Education.. JZ Social sciences & humanities in general
K	Society.. Sociology & social anthropology.. Customs & folklore
L2	Area studies.. Travel & topography..
L6	History.. Biography..
P	Religion.. {Alternative, preferred at Z)
Q	Applied social sciences.. Social welfare.. Crime & criminology
R	Political science.. S Law..
T	Economics.. TQ Management of economic enterprises..
U	Technology.. Materials.. Energy technologies.. Construction technology.. Transport technology.. Process industries..
VV	Household technology & management..
VW	Recreation arts..
W	The Arts.. Visual arts.. Applied arts & design.. Fine arts..
WP	Performing arts.. Music.. Theatre.. Cinema..
X	Philology.. Language & literature
Z	Religion.. The Occult..

Appendix 2

Examples of Hierarchies in BC2

	* Both display inverted schedules and retroactive notation; e.g., Criminal court procedure SBW 6E; Children in primary care HXO ELK (in which a special facet indicator E is used).
HH[1 G]	HEALTH & MEDICINE
HIAP	(Agents) HHG Personnel
HI	(Operations) Technical procedures..
	(Agents) Medical materials.. Equipment..
HJ	Preventive medicine.. Public health.. Health maintenance
HL	Curative medicine
HLK	Primary care.. Secondary care.. Nursing..
HN	Clinical medicine
HNG	Investigation.. Diagnosis.. Treatment, therapy..
HNRE	Physical therapy.. Radiation therapy.. Drug therapy..
	(Processes)
HP	Diseases & pathology.. By process.. By cause..
	(Parts, organs, systems of the body)
HTF	Regions.. HTJ Locomotor system, musculo-skeletal system..
HUG	Cardiovascular system.. HUR Nervous system..
HWE	Respiratory system.. HWI Digestive system..
HWV	Urogenital system..
	(Kinds of persons)
HXD	Males.. Females.. HXO Children.. HXW Aged persons..
S	LAW
S2	Primary materials (Works of law as distinct from works about)
	(Common subdivisions)
S34	Legal profession
S5A	Jurisprudence.. Sources of law.. Formal.. Case law..
S6	Practice & procedure, administration of justice
S6A	Practice of law.. Preparation of documents.. Advocacy..
S6G	Courts & court procedure.. Kinds of courts..
S8	Actions, lawsuits.. Parties to the action.. Proceedings..
S8	Hearings.. Trial procedure.. Evidence..
S9	Judicial decision.. Juries.. Remedies.. Appeal..
S9	(Special kinds of proceedings).. Summary.. Class actions
S9VB	Substantive law, subjects of law
	(By relation of jurisdiction to persons)
S9W	Private law.. Civil law.. Conflict of laws..
SAP	Persons.. Family & kinship.. Corporate persons..
SBD	Obligations.. Liability.. Contract.. Torts..
SBH	Property law.. Commercial law..
SBS	Environmental law.. Social law.. Cultural law..

SBW	Criminal law
SBY	Public law.. Constitutional law..
SCYX	Jurisdictions, systems of law
SD	(By political authority)
SD	International law.. Law of war..
SE	Supranational law.. European Union law..
SH	National law, municipal law
SL	Common law systems.. English law.. Anglo-American..
SR	Civil law systems.. French law..
	(By religious authority)
SWE	Ecclesiastical law.. Christian law, canon law..
SYB	Islamic law, Shari'a

REFERENCES

Aitchison, J. (1986). A classification as a source for a thesaurus: The Bibliographic Classification of H. E. Bliss as a source of thesaurus terms and structure. *Journal of Documentation, 42*(3), 160–181.

Aitchison, J., Gilchrist, A., & Bawden, G. (1997). *Thesaurus construction and use: A practical manual.* London: Aslib.

Austin, D. (1984). *PRECIS: A manual of concept analysis and subject indexing* (2nd ed.). London: British Library.

Beghtol, C. (1994). *The classification of fiction: The development of a system based on theoretical principles.* Metuchen, NJ: Scarecrow Press.

Bliss, H. E. (1929). *The organization of knowledge and the system of the sciences.* New York: H. Holt and Company.

Bliss, H. E. (1933). *The organization of knowledge in libraries and the subject approach to books.* New York: Wilson.

Bliss, H. E. (1940/1953). *A bibliographic classification, extended by systematic auxiliary schedules for composite specification and notation.* New York: Wilson.

British Technology Index (1962–). London: Library Association.

Broad system of ordering: Schedule and index. (E. Coates, G. Lloyd, & S. Simandl, Eds.). (1978). The Hague: International Federation for Documentation (FID).

Broadfield, A. (1946). *The philosophy of classification.* London: Grafton.

Brown, J. D. (1939). *Subject classification for the arrangement of libraries and the organization of information, with tables, indexes, etc., for the subdivision of subjects* (3rd ed.). London: Grafton. (Original work published 1906)

Cleverdon, C., Mills, J. & Keen, M. (1966). *Factors determining the performance of indexing systems.* Cranfield: Aslib Cranfield Research Project.

Coates, E. J. (1960). *Subject catalogs: Headings and structure.* London: Library Association.

Coates, E. J. (1973). Some properties of relationships in the structure of indexing languages. *Journal of Documentation, 29*(4), 390–404.

Dahlberg, I. (1992). The basis of a new universal classification system seen from a philosophy of science point of view. In N. J. Williamson & M. Hudon (Eds.), *Classification research for knowledge representation and organization: Proceedings of the 5th International Study Conference on Classification Research, June 24–28, 1991, Toronto, Canada.* Amsterdam: Elsevier.

Farradane, J. E. L. (1950). A scientific theory of classification and indexing. *Journal of Documentation, 6*, 83–99.

Feibleman, J. K. (1954). Theory of integrative levels. *British Journal for the Philosophy of Science, 5*, 59–66.

Foskett, D. J. (1961). Classification and integrative levels. In D. J. Foskett & B. I. Palmer (Eds.), *The Sayers memorial volume: Essays in librarianship in memory of William Charles Berwick Sayers* (pp. 136–50). London: Library Association.

Fugmann, R. (1994). [Review of *The alcohol and other drug thesaurus: A guide to concepts and ter-*

minology in substance abuse and addiction]. Washington, DC: Department of Health and Human Services.

Gardin, J. C. (1965). *SYNTOL*. New Brunswick: Rutgers, the State University.

Hirst, P. H. (1974). *Knowledge and the curriculum: A collection of philosophical papers*. London: Routlege & Kegan Paul.

Hjørland, B., & Albrechtsen, H. (1999). An analysis of some trends in classification research. *Knowledge Organization, 26*(3), 131–139.

Hulme, W. (1911–1912). Principles of book classification. *Library Association Record, 13,* 444–449.

Hume, D. (1748). *Inquiry concerning human understanding*. XII (iii).

Langridge, D. W. (1976). *Classification and indexing in the humanities*. London: Butterworths.

Merrill, W. S. (1939). *Code for classifiers: Principles governing the consistent placing of books in a system of classification*. Chicago: American Library Association.

Metcalfe, J. (1957). *Information indexing and subject cataloging: Alphabetical, classified, coordinate, mechanical*. New York: Scarecrow Press.

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

Mills, J., & Broughton, V. (1977–). *Bliss bibliographic classification* (2nd ed.). London: Butterworths.

Phenix, P. H. (1964). *Realms of meaning: A philosophy of the curriculum for general education*. New York: McGraw-Hill.

Ranganathan, S. R. (1960). *Colon classification: Basic classification* (6th ed.). London: Asia Publishing House.

Ranganathan, S. R. (1967). *Prolegomena to library classification* (Ranganathan Series in Library Science, 20). London: Asia Publishing House.

Vickery, B. C. (1959). *Classification and indexing in science* (2nd ed.). London: Butterworths.