Notational Expressivity; the Case for and against the Representation of Internal Subject Structure in Notational Coding

Vanda Broughton

University College London

Vanda Broughton is Lecturer in Library and Information Studies at University College London, where she teaches Information Retrieval. She has been involved in Classification Research since 1972, and was for many years Research Fellow and Research Associate at the Polytechnic (later University) of North London School of Librarianship. She has collaborated with Jack Mills in the revision of the Bliss Bibliographic Classification since the early days of that project, and is joint editor of the new edition. Since 1997 she has also enjoyed a close involvement with the UDC, and is currently working on the revisions of Religion and Chemistry. She is a member of the Classification Research Group and the Bliss Classification Association, and is additionally concerned with the Broad System of Ordering.

Broughton, V (1999). **Notational Expressivity; the Case for and against the Representation of Internal Subject Structure in Notational Coding.** *Knowledge Organization, 26(3).* 140-148. 22 refs.

ABSTRACT: The ways in which notation can be used to express the content of documents to which it relates are various. At the most superficial level notation can correspond to the hierarchical structure of the schedules or link to literal components. The notation of compound concepts can express the structure and composition of the compound, and systems exist in which symbols denote the functional roles of the constituent elements and the relationships between them. At the highest level notation can be used to mirror the actual structure of those entities which it represents, as in the case of mathematical systems or chemical compounds. Methods of displaying these structures are examined, and the practicality in a documentary context is questioned, with particular reference to recent revision work on the chemistry class of the Bliss Bibliographic Classification 2nd edition (BC2).

Introduction

Jack Mills writes in his *Modern outline of library classification,* "The vital function (of notation) is to mechanically maintain the sequence of subjects, by giving each term a symbol possessing an agreed ordinal value. This is seen when it is used on the backs of books to maintain shelf-order, or on entries in catalogues and bibliographies to allow a classified order of subject to be maintained. Notation is primarily an ordering device." (Mills, 1960 p.37)

Nevertheless, compilers of systems of classification and indexing have normally felt that it is also a priority to show a correspondence between the schedule structure of the classification and its notation, and it is often the case that a visible relationship exists between classmarks and the information that they represent. This paper will attempt to record the various ways in which indexing languages have used notation to represent not only the intellectual content of a document, but also the relationships between concepts, whether these are hierarchical relationships of the genus-species and whole-part variety, or the functional relationships between compounded terms in an analytico-synthetic system.

The need for expressivity in notation is arguable; on the whole end-users of a scheme do not notice the structure of neither the notation, nor the relationships between class-marks that notation can display. In specific subject searches of an encoded database the objective is normally to retrieve a precise location for a precise subject with no envisaged reference to the general subject context, and when used as an ordering device for a browsing function the notation is probably largely ignored once the reader has located a suitable starting point. It is my experience that readers may not even realise the consistency of classmarks across two or more libraries using the same classification until this is pointed out to them. Expressivity is

primarily of use to the information professional, who can exploit it as an aid to retrieval.

Notational expressivity can operate to a number of purposes:

- to indicate the relationships between classes;
  - of a genus-species, or whole-part nature (hierarchical expressivity)
  - of a functional nature (notated operators and relators)
- to provide mnemonic links between concepts and notation;
  - when corresponding letters are used in the term and in the notation
  - when actual words or symbols for the class are part of the notation (as in UDC Class 546.791.027 Isotopes of Uranium 546.791.027"238 Uranium 238)
- to display the facet structure of a synthesised classmark; (the 0 in DDC, and the apostrophe (') and the colon (:) in UDC)
- to reflect the structure of the entity itself (as in the case of chemical structure notations, such as Wiswesser (Smith, 1968) or SYBYL (Ash, 1977)).

Hierarchical Expressivity

The most obvious example of notational expressivity occurs when a correspondence exists between the length of the classmark and the hierarchical structure of the classification system (in terms of genus-species and whole-part relationships of both taxonomic and other varieties); this is the context in which most users comprehend the notion of expressivity. It is particularly evident in those classifications, which employ a numerical notation, where the overwhelming tendency is to use this decimally (and centesimally where it is more appropriate) to maintain a precise relationship between the number of steps of division and the length of the class-mark.

For example, in the 20th edition of the *Dewey Decimal Classification* (DDC20) (Comaroni, 1989), the following example of hierarchy is found in Class 634:

634 Orchards, fruits
634.1 Pomaceous fruits
634.2 Drupaceous fruits
634.21 Apricots
634.22 Plums
634.227 Varieties and kinds
634.23 Cherries
634.25 Peaches
634.257 Varieties and kinds
Example: nectarines
634.3 Citrus and moraceous fruits

Although the notational classes 634.26 - 634.29 are empty, they are not used to accommodate the class *nectarines*, which is given the longer class-mark 634.257 to express notationally its status of containment by the broader class 634.25 *peaches*.

There is clearly a problem if the next level of division of a class has more than ten sub-classes of equal rank; where ten digits are insufficient to accommodate the sub-divisions of a containing class, centesimal divisions are often used to notationally maintain their co-ordinate status. Hence in the full English edition of the *Universal Decimal Classification* (UDC, 1979), within the class Botany we find:

582.291 ASCOLICHENS
582.292 **Pyrenocarpeae**
.22 Moriolaceae
.24 Epiglocaceae
.26 Verrucariaceae
.28 Dermatocarpaceae
.32 Pyrenothamniaceae
.34 Pyrenulaceae
.36 Phyllopyreniaceae
.38 Trypetheliaceae
.42 Paratheliaceae
.44 Astrotheliaceae
.46 Strigulaceae
.48 Pyrenidiaceae
.52 Mycoporaceae
582.293 **Gymnocarpeae**

The device of centesimal divisions is used here to maintain the appearance of a relationship of equal status between the terms in the class Pyrenocarpeae.

End-users normally find this sort of arrangement helpful in displaying the structure of a subject, but it has two significant shortcomings:

- it is 'wasteful' of notation, since much of the available capacity may remain unused (as in the case of the fruits in the DDC example above);
- topics, that come far down a hierarchy, must be given a lengthy notation, even when they are of major significance and carry large volumes of literature (the classic example here is that of Computing in DDC, which has the class mark 001.64, even though it now has the status of a discipline in most academic institutions).

Mnemonics

Broadly speaking, the phenomenon of structural correspondence between the notational and the

schedule hierarchy is not quite so evident with alphabetic or alpha-numeric notations, where the primary notational function of ordering the classification, and mnemonic value may be placed more highly than expressivity. The first edition of the *Bibliographic Classification* (BC1) (Bliss, 1952) contains the following schedule for Electrochemistry:

CE	Electrochemistry
	General, elementary
CEA	Theoretical
CEB	Experimental and laboratory
CEC	Calculations
CED	Special theoretical subjects
CEE	Electrolysis, chemical studies
CEF	Theoretical, laws of, units
CEG	Experimental, laboratory
CEH	Electrolytic dissociation
CEI	Ionization, chemical studies

Not only is the notation used ordinally rather than decimally (to maintain the order of classes rather than to express the subordination of one class to another) but links are created between the notation and the initial letters of subjects (as CE - Electrochemistry, CEE - Electrolysis, and CEI - Ionization). This phenomenon of *literal mnemonics*[1] is clearly only possible in a letter based or mixed notation, and it is a particular feature of the *Bibliographic Classification*, which has been carried over into the second edition. It is necessary to abandon any consistent correspondence between the schedule hierarchy and the notation if these links are to be exploited, but once the principle of non-correspondence is established, other features, such as brevity of class marks for significant topics with large literatures can come to the fore. This addresses the problem found in predominantly expressive systems such as DDC, where, as mentioned above, topics such as Computing have excessively long numbers because of their hierarchical placing.

The mnemonic value of notation is not only found in the use of initial letters but also in the employment of consistent notation for specific concepts (*systematic mnemonics*); for instance the device of parallel subdivision in DDC and UDC preserves the notation for the same concept used in differing contexts.

For example, in Dewey (DDC20)

at	726.7	Monastic buildings
	726.78	Of other religions (*i.e. non-Christian*)
		Add to base number 726.78 numbers following 29 in 292-299 (*Religions)* e.g. Buddhist monasteries 726.7843 (*from the notation for Buddhism, 294.3)*

All the major classification schemes provide systematic and auxiliary tables in order to maintain a consistent notation whenever concepts such as Period, Place, Language and Form are required. The idea of a persistent and pervasive use of notation linked to specific terms is of fundamental significance in the development of modern classification systems; the idea of generally applicable sets of concepts such as Place and Form needed only a short step to arrive at the BC1 and UDC special schedules and auxiliaries for topical subdivisions of subjects, and thence to the full-blown facet analysis of the *Colon Classification* (Ranganathan, 1933) where notation was absolutely and rigorously consistent for every concept in the system.

Structural expressivity

When we use the term structural expressivity it is usually to mean the presence in the notation of a discernible composite structure reflecting the synthesis of component parts. This may be very evident, as in the use of the colon in UDC to link the classmarks for any two, apparently independent, subjects; for example:

Night photography of animals	77.023 : 59
Christian views of evolution	28 : 575.8

In most of the general schemes some notational device may be employed to introduce a built classmark; in Dewey the 0 is frequently used as facet indicator, and in the UDC, in addition to the general relationship indicator of a colon, various symbols are used to introduce auxiliaries (such as the double quotation marks ("... .."), which indicate the concept of *time*).

It is usually clear in most systems of classification when a systematic table of any kind has been used, since in addition to the consistent notation for these concepts, there is frequently a stylistic element involved in the compounding e.g.

Ecology in France ELf
(BC1, where lower case letters are employed for common subdivisions of place)

A dictionary of chemistry 540.3
(Dewey decimal classification, where the 0 is used to indicate the addition of common form subdivisions directly to a subject class).

Nevertheless these linking devices say nothing in particular about the nature of the relationship between the terms so attached; they simply serve to in-

dicate an association of two or more conceptual elements. It is not altogether clear, for instance, in the example from BC1 above, whether the subject is 'The ecology of the French environment' or 'The study of Ecology in France'.

Representation of more complex structures and internal relations

Some analytico-synthetic classifications have consequently attempted to go further than this in representing the relationships between terms. Two principle schools of thought characterised subject analysis for information retrieval during the fifties and sixties. In the United Kingdom, members of the Classification Research Group (CRG) followed, in the main, the methodology of facet analysis of S. R. Ranganathan and analysed the terminology of a particular subject into *functional categories* (entity, part, process, agent, etc.). The interrelations between terms were inherent in the categorization e.g. an *agent* necessarily acted upon a *patient*; a *part* was contained by an *entity*, etc. The syntax of the system was maintained by the principle of citation order, in much the same way that word order in the sentence maintains the case status of nouns in a generally non-inflected natural language like English.[2]

In contrast to this, research in Europe tended to favour the development of complex systems of notational representation of subject syntax. Typically, these classifications analysed the structure of a subject, in terms not only of the vocabulary, but also of the functional position and interrelations between terms, and were equipped both with notational codes for the terminology and with an encoded syntax which expressed the relational structure of compound concepts. Of particular interest is the work of Jean-Claude Gardin not least because of the complexity of the relationships he identified in a range of non-scientific subjects (Gardin, 1958). Working principally in the field of archaeology, he developed very complex classifications for various artefacts, and for ornamentation and iconography; in the latter he articulated a vocabulary of *representations*, supplemented by those for *attitude*, *viewing angle*, *actions*, *topographic connections*, and for *declensions*, a syntactic category which describes the function (e.g. subject, object, instrumental, locative) of the entity within the representation. All of these operators were notated using up to 20 elementary symbols, operational indicators and geometrical groupings of the symbols. Gardin subsequently applied this system of conceptual analysis to texts, including Mespotamian tablets, the Quran and tribal myths. Further examples of this type of highly detailed conceptual analysis are documented by de Grolier (de Grolier, 1962), and a tabulated analysis of a range of categories and operators is given in an article by Perrault (Perrault, 1965).

Among classificationists in the United Kingdom, J. E. L. Farradane introduced, in his system of Relational Indexing (Farradane, 1952, 1955), the idea of role indicators, which notated the precise relationship of each term to its neighbours. Farradane's thinking was fundamentally different to that of the other members of the Classification Research Group, whose methods of facet analysis are described above. In Farradane's system the individual terms, or *isolates*, had no intrinsic functional value, but were linked to other terms by a notational system that represented the relation between adjacent terms. Hence, the subject summary, or citation string (called by Farradane the *analet*) for 'the prevention of bacterial infection of fruit trees during grafting' would be :

fruit trees/ - grafting/ ; infection/ (bacteria/ - prevention

where the symbols, - , ; , (, and - represent the relations of Reaction, Association, Appurtenance, and Reaction respectively. These relations, to which he gives the name *operators*, are limited in number and are to be derived from the fundamental mental processes of learning. Other operators that he identified are Concurrence, Comparison, Equivalence, State, Distinctness, and Causation.

Farradane's system is unusual in that it builds the finished classification number from the relational associations identified in the analets, working inductively from the documentation of a subject, rather than working deductively from its perceived theoretical structure. Nevertheless, his methodology did not ever gain widespread currency, and nowadays most modern British classification schemes are based on facet-analytical methods, where there is no necessity to express the functional role of terms, since this is inherent in the schedule and in the citation order.

Not least among the reasons for this is the use, in most of the systems described above, of symbols outside the range of the Roman alphabet and Arabic numerals. While we have seen the value of expressiveness used in subject representation, order is still a primary function of notation, and where there is no obvious sequence of symbols, confusion can occur, resulting in a poor rate of retrieval, and a tendency to misfiling. Non-alphanumeric symbols also lack natural expressivity of the sort that characters with value have; even to the uninitiated, 301.42 looks like a subclass of 301.4, while there is no similar obvious relation between an apostrophe (') and a quotation mark (").

ties. Writing in the late 1960s, Arthur Maltby felt able to state with conviction "... the ordinal value of notation cannot be overemphasised. Arbitrary signs which convey no definite order, should rarely, if ever, be used; they must be rejected in favour of letters or numbers" (Maltby, 1967).

Structural expressivity in chemical documentation

One area of documentation in which the expression of structural components and their relationships is of singular significance is that of the literature of chemical compounds. The constant discovery and synthesis of thousands upon thousands of new chemical substances has created a whole industry concerned with their recording and identification. A significant problem is that of nomenclature, with the majority of (particularly organic) compounds carrying various descriptive, structural, functional and brand names. The problems of chemical nomenclature and the need for precise identification of individual chemical entities has led to the proliferation of systems of substance registration and description; these have attempted to notate not only molecular components (atoms, ions, radicals and functional groups), but also bond structures, the positioning of constituent parts, charge, valency and so on.

This situation is only partially resolved by the existence of indexing systems, such as those used in the Merck Index (Budavari, 1989) or Chemical Abstracts Registry which record all new compounds and allocate unique (though random) code numbers to each. However, it is apparent that, in practice, the demand made on databases deployed for the recording of chemical substances is substantially different from that required of databases used for retrieval purposes, or for bibliographic or documentary classification.

> "Database systems which handle chemical information have some special requirements, advantages and disadvantages when compared to conventional information systems ... Since the universal language of chemistry is the molecular structure, the molecular structure itself is uniquely suitable for use as the primary key for entries in a chemical information database" (Weiniger, 1999).

Systems of the type that record and identify individual chemical entities do not of themselves solve the problems of bibliographic description for storage and retrieval, which require the assignment of descriptors, and notational encoding of some sort. For this purpose various descriptor languages linked to molecular structure have been evolved. The best known of the linear systems of chemical description is Wiswesser Chemical Line-Notation[3] (WLN), which reduces the structural formulae of compounds to a linear code. It utilises the normal numerals and Roman letters, plus four symbols (&, -, /, *) to represent functional groups and other components of a molecule; these are combined in a pre-determined order for linear, monocyclic and polycyclic compounds, with notational devices indicating branch, ring and part-ring structures. Other linear notation systems include SYBYL Line Notation (Ash, 1997), an ASCII language that functions in a way similar to Wiswesser.

In the process of recent work on the revision of the main class Chemistry in BC2, it was necessary to establish a method for the treatment of chemical compounds. In the original classification (as in the case of the DDC) virtually no provision had been made for substances in Inorganic chemistry, beyond the enumeration of individual elements, and there was no capacity for systematic synthesis of compounds. It is true that some general classes of salts were provided for at CHR, and that there was a special auxiliary for subdivision of any substance, but neither of these allowed for classmark building based on the molecular components, which was clearly the requirement in a modern classification.

Previous drafts of BC2 Class C (the penultimate draft schedule in the 1970s and subsequent reworkings in 1991 and 1994) had effected classmark synthesis by building from general conceptual classes, some of which represented chemical properties, such as valency, or acid/base properties, and some of which were structural, the components ranging from the constituent elements to ions, radical and more complex functional groups.

It was now necessary to establish whether this method, or an elaboration of it, was sufficient for the subject description of chemical entities, or whether we should go further and attempt an accurate citation string (and hence notational coding) representative of the molecular structure, on a model of Wiswesser or similar linear notations.

A number of considerations conditioned the decision:

- the appropriate level of detail in a bibliographic classification;
- the capacity of the generalist librarian to apply such a system were it to be provided;
- the effects on the overall order of individual compounds, and of their relations to more general functional classes;
- the capacity of the existing notation to accommodate a representational system without recourse to additional symbols beyond the usual alphanumeric base;

- following on from the above, the clarity of filing order for a notation with additional symbols, and the capacity of our indexing and layout software to cope with those symbols (which would have implications for machine handling on a wider scale).

The structure of the current draft schedules for BC2 follows the pattern given below (this being a considerable abridgement of the complete draft):

C	Chemistry
C2	(Common subdivisions)
	(Operations and agents)
C32	Research in chemistry
C36	Practical chemistry
	Materials and equipment
	Materials
	Equipment
	Investigative techniques
C82	Experimental chemistry
C9	Chemical analysis
CA	Physical chemistry
	(Physical properties, conditions, phenomena)
CAC	Chemical combination and structure
CAG	Chemical bonds, valency
CAX	Reaction chemistry
	(Types of reaction)
CDS	Mechanochemistry
CDU	Thermochemistry
CE	Electrochemistry
CES	Photochemistry
CEV	Nuclear chemistry
	(Change of state)
	(States of Matter)
	Solid state chemistry
CGG	Chemical substances
	(By physical and chemical properties)
	Metals (general)
	(By charge and valency)
CGP	Elements (general)
CGU	Compounds (general)
	Molecular compounds
	(by bond structure)
	Covalent
	Ionic, electrovalent
	(by valency)
	Monovalent
	(by behaviour in reaction)
CHA	Radicals
	Positive radicals
	Negative radicals
CHE	Acids ... . bases... salts
	Salts
	Simple salts
	Acid salts
	Derivatives
	Complexes
CI	Inorganic chemistry
	(Specific elements and their compounds)

Using the rules of retroactive synthesis, any and every concept filing ahead of a given term, can be brought down under that term either

(I) to compound directly with it, and/or

(II) to act as a specifier (literally species maker) to generate a class of compounds characterised by that property. So for example, although the general concept of *photoconductivity* is classed at CES BHU, there exists a general class of *photoconductive compounds* at CGH ESB HU (= chemical compounds + photoconductivity) and the concept of photoconductivity can be added to every conceptual class or individual compound.

In this way, classmarks for individual chemical substances would be built by synthesis of the codes for the constituent parts of the compound subject. Bearing in mind that BC2 is used principally for the organization of documents, it was essential to decide on those parts and/or properties of a compound that required specification within the subject string. Various contenders offered themselves, including: constituent elements; number of atoms of each element; valency; bonds and bonding structures; functional groups, radicals and ions; functional properties (electronegativity, etc.).

An additional consideration was the need to provide not only for the description and identification of individual compounds, but also for general chemical and functional classes, and to examine the relationship between individual substances and their containing classes. It is usually the case that the searching of large commercial chemical databases is primarily by specific substance; this is less likely to be the case for users of library and bibliographic databases and catalogues. Bibliographic entities, as opposed to chemical entities, may correspond to an intermediate level of description i.e. they are quite as likely to be concerned with general classes, functional properties or aggregates of classes, as with individual chemical substances. And, whilst there is clearly a need for the chemist to identify uniquely specific compounds, we might question whether this level of precision is necessary (or achievable) within the context of the general scheme of documentary classification.

Provision for optimum (rather than maximum) specificity and the existence of logical intermediate steps in the division (or synthesis) of classes are essential for those libraries requiring only an intermediate level of specificity. It would be ludicrous to require the generation of a 30-character classmark in order to locate a document that is the only member of its containing class.

Bearing in mind that the principal problem for the compiler of a pre-co-ordinate scheme (i.e. one in which the components of a compound subject are co-ordinated before storage), is to produce an acceptable linear order in which, and from which, compound subjects can be predictably stored and retrieved. In a pre-co-ordinate scheme for chemistry, the treatment of individual chemical substances can be approached in two ways:

1. The compound can be regarded as the sum of its structural parts, and a notation assigned which is representative of that structure.

   e.g. sodium bisulphate $NaHSO_4$ can be analysed as
   1 sodium, 1 hydrogen, 1 sulphur, 4 oxygen
   or: 1 sodium, 1 hydrogen, 1 sulphate ion.....
   etc.

- atoms would be cited in order of say, atomic number or electronegativity (simply as an arbitrary ordering device);
- commas, or apostrophes, or some such device would be needed to act as an indicator separating the notation for particular elements;
- notation might have to be provided for linkages, bond structures and positional information;
- no reference would be made to chemical physical or other properties or attributes of the substance.

This approach has the following advantages:

a) it is predictable;
b) the notation is expressive of structure (and needs to be for retrieval purposes);
c) it is applicable with some basic chemical knowledge. The major disadvantage would be the need to introduce non-alphanumeric symbols, with all the attendant problems associated with machine handling, and the lack of an obvious filing order. Additionally the indexer would need access to information about molecular structures and formulae, their relations to names of compounds, and the necessary scientific knowledge to translate the structure into an encoded form.

In this system, compounds will, to all practical purposes, file independently of classes characterised by properties, e.g. acids, salts, etc., and, dependent upon the arbitrary sequence used to order them, they will not necessarily relate to traditional chemical compound groupings e.g. a sulphate of a metal may file with sulphur, rather than with the metal.

2. The compound could be treated in terms of language-based conceptual classes, derived from both substance and non-substance facets in the schedule, with a built in citation order.

   e.g. sodium bisulphate $NaHSO_4$,
   sodium compounds - (monovalent) - compounds with hydrogen,
   sulphur and oxygen - salts - bisulphates

Concepts such as valency, acid/base status, etc., while given a position determined by the facet structure, would normally be enumerated at the appropriate point, to aid the classifier.

Advantages of this approach are as follows:

a) the non-chemist can approach the indexing of compounds on a natural language basis;
b) the resulting notations will group substances with chemically related classes;
c) there will be correspondence between general classes of compounds and specific substances; and
d) the compounding will follow normal retroactive rules, with no need to introduce operators, facet indicators, or non-alphanumeric symbols of any kind.

As with most systems of this kind, there will have to be a cut off point in the addition of terms, particularly properties. For example, the string, *sodium - monovalent - salts of sulphur oxyacids - bisulphates*, is sufficient to define the substance adequately for storage and retrieval purposes. It is not necessary to add extra descriptors to position the compound concept, and to that extent maximum specificity is sacrificed.

A comparable situation had already been encountered in the revision of BC2 Class AM-AX, Mathematics, where the complexity of interrelations between mathematical systems and structures posed very great problems in reducing these to an acceptable linear order. From the user's (indexer's) point of view the subject summary would be more easily done on a linguistic basis. In practice, the results achieved by adopting this approach were very acceptable to mathematicians, and the method was manageable by non-mathematician indexers.

If the classification of chemical compounds were to proceed on this pattern, we would have abandoned

the idea of 'chemical structure' expressivity in the classmark, although we would still have retained mnemonic values (where used), facet indicator/structural components, and the consistency of notation for particular elements (e.g. within Class C sodium is always -KS wherever it appears).

In practice this seems a reasonable, and necessary compromise, for a system that must reduce the complexities of n-dimensional structures to a linear order.

Conclusion

We come to the conclusion that while synthesis of classmarks for individual substances must be logical and systematic, and their location must be consistent with the structure of non-substance facets, a notation based exclusively on structural constituents does not necessarily produce optimum filing order in a pre-coordinate system, intended principally for documentary classification.

Collocation of related topics and browsability may be at least as important as specificity and the provision of unique locations. Ease of use by the indexer is also a consideration for a general system of classification, and notations involving non-alphanumeric notation, and the encoding of complex relationships between components are intellectually demanding and time consuming to implement; they also lack clarity from the end-user's viewpoint, and while increasing the accuracy and specificity of storage, in a number of respects they impede the ease of retrieval.

It is important also to realise that with the advent of the Internet and world wide web, much activity in the area of storage and retrieval is carried out by non-indexers, and searching is almost exclusively on a natural language basis. Surveys of the use of existing classifications on the web (e.g., Project DESIRE, 1997) have criticised them for being too complex, and above all we should be wary of introducing levels of sophistication that are incomprehensible to the end-user, and in a context that does not demand them.

Notes

1 Bliss called this system of corresponding initials "intensive mnemonics", as opposed to "extensive mnemonics", where the same notation is used to represent the same concept (e.g. B Physics, C Chemistry, CB Physical chemistry) and "constant mnemonics" where the notation is consistent across the whole system as in the case of the systematic schedules.

2 A discussion of the similarities between natural language types and the function of role indicators is contained in Costello, J. C. (1964), A basic theory of roles as syntactical control devices in co-ordinate indexes. *Journal of Chemical Documentation* 1964 (4) 116-124

3 There are numerous references to WLN in the literature. A guide to its application can be found in the following: Smith, E.G. (1968) *The Wiswesser Line-Formula Chemical Notation,* New York: McGraw-Hill; Ebe, T. and Zamora, A. (1976) Wiswesser Line Notation processing at Chemical Abstracts Service, *Journal of Chemical Information and Computer Sciences*, 16 (1) 33-35. Koniver, D.A. (1972) Wiswesser Line Notation: simplified techniques for converting chemical structure to WLN, *Science* 176 (30th June) 1437-39.

References

-- (1979) *Universal Decimal Classification.* 2nd English full edition. London: British Standards Institution

Ash, S. et al. (1977) SYBYL line notation (SLN): a versatile language for chemical structure representation. *Journal of Chemical Information and Computer Sciences* 37 (1) 71-9

Bliss, H.E. (1952) *A bibliographic classification.* New York: H.W.Wilson

Budavari, S. ed. (1989) *Merck Index; an encyclopaedia of chemical, drugs and biologicals.* 11th edition. Rahway, N.J.: Merck and Co.

Comaromi, J. *et al.* (eds) (1989) *Dewey Decimal Classification and Relative Index.* 20th edition. New York: Forest Press

Costello, J.C. (1964) A basic theory of roles as syntactical control devices in co-ordinate indexes. *Journal of Chemical Documentation* 1964(4) 116-124

DESIRE (1997). The role of classification schemes in Internet resource description and discovery. *Work package 3 of Telematics for research project DESIRE* (RE1004). < http://www.ukoln.ac.uk/metadata/desire/classification/classtc.htm 14 May 1997. Seen 12.05.1999

Ebe, T. *and* Zamora, A. (1976) Wisswesser Line Notation processing at Chemical Abstracts Service. *Journal of Chemical Information and Computer Sciences* 16(1) 33-35

Farradane, J.E.L. (1952) A scientific theory of classification and indexing. *Journal of Documentation* 8 73-92

Farradane, J.E.L. (1955) The psychology of classification *Journal of Documentation* December

de Grolier, E. (1962) *A study of general categories.* Paris: UNESCO

Gardin, J.C. (1958) Four codes for the description of artefacts. *American Anthropologist* 60(2) 335-7

Gardin, J.C. (n.d.) On the coding of geometrical shapes and other representations, with reference to archaeological documents. *ICSI* Vol. E 889-901

Koniver, D.A. (1972) Wiswesser Line Notation; simplified techniques for converting chemical structures to WLN. *Science* Vol. 176 30th June 1437-39

Maltby, A. (1967) *A manual of classification for librarians* by Charles Berwick Sayers. 4th edition revised. London: Andre Deutsch

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

Mills, J. *and* Broughton, Vanda (1977-) *Bliss Bibliographic Classification.* 2nd edition. London: Butterworth

Mills, J. *and* Broughton, Vanda (1993) *Bliss Bibliographic Classification: Class AM-AX Mathematics, Statistics and Probability.* London: Bowker-Saur

Perrault, J.M. (1965) Categories and Relators: a new schema. *Review of International Documentation* 32(4) 136-144

Ranganathan, S.R. (1933) *Colon Classification.* Madras: Madras Library Association

Smith, E.G. (1968) *The Wiswesser Line Formula Notation.* New York: McGraw Hill

Weininger, D. (n.d.) *A distributed chemical information database system* <http://207.225.60.130/papers/lidpaper.html>. Seen 27.07.99