CLASSIFYING INDEXING CODING

S R RANGANATHAN

Location of relevant entries by machine search requires prior facet analysis. Difficulties in facet analysis regarding consistent choice of sequence of subjects and of facets in subjects in the idea plane and difficulties in the verbal plane are described. Postulational approach and indexing by chain procedure solve the difficulties. Suggestions are made for the machine to be fed with the specific permutation of facets along with connecting symbols for the fundamental categories and the rounds and levels of the facets.

Contractions used:-
CC - Colon Classification
DC - Decimal Classification
PMEST - Personality, Matter, Energy, Space, Time (the five fundamental categories)

Terminology:- The classificatory technical terms used in this paper will be found defined in my Library classification glossary, (Annals of Library Science, 5, 1958, 65-112).

INTRODUCTION

It is nowadays sought to use machinery for the two following major purposes:-

1 Locating, in a store of entries, those relevant to a specific subject — pinpointedly, exhaustively, and expeditiously; and

2 Producing copies of the entries, and/or the abstracts of the respective documents, and/or copies of the respective whole documents.

The latter work does not concern a classificationist. The former does. This paper is, therefore, concerned only with the former.

1 FACET ANALYSIS AND ITS VITAL ROLE

The following five stages may be reconsidered in the work involved in the location of entries with machinery:-

1 Spotting out and marking in the entry of a document all its basic classes, and the isolates in all the facets attached to each of the basic classes;

2 Feeding into the memory of the machine each of those basic classes and isolates, information about the host document or location of the document, and if necessary the name of the author and the title of the document;

3 Expressing any subject brought up for search, in terms of all its basic classes and the isolates in all the facets attached to each of its basic classes;

4 Programming the machine to search the subject in terms of its basic classes and isolates; and
5 Re-translating from the machine language to the natural language the entries of all the relevant documents spotted out by the machine in carrying out the programme.

Of these five stages, the first four fall within the scope of this paper. Stages 1 and 3 are extra-machine stages. They belong to the discipline of classification. Facet Analysis plays a vital part in these two stages. Stages 2 and 4 are intra-machine stages. Facets figure in them too. On account of this involvement of machine search in facet analysis, it may perhaps be helpful in the designing of machine for search, to give an account of the successive difficulties met with in the development of CC and of the facet analysis associated and developed along with it.

2 FACET ANALYSIS AND ITS STAGES OF DEVELOPMENT

2.0 Difficulties Met With

The different stages in the development of CC and of facet analysis were associated with the five difficulties mentioned below:

1. Difficulty in the consistent choice of the sequence of subjects;
2. Difficulty in the consistent choice of the sequence of the facets in a subject;
3. Difficulty in the consistent choice of fundamentally the same sequence of facets in all subjects;
4. Difficulty in the design of a notational system to mechanise the preferred arrangement of subjects; and
5. Difficulty in the design of a notational system to mechanise the preferred arrangement of facets in a subject.

The first three difficulties fall in the idea plane; and the last two fall in the notational plane. Some of the difficulties have been fully solved; and some have been solved only partially.

2.1 Consistent Sequence of Subjects

2.1.1 DC and Single Schedule

In DC a consistent sequence of subjects was sought to be secured by the enumeration of all the subjects - known as well as anticipated in a definite sequence. In other words, DC consciously determined the sequence of all such subjects, as an advance measure. The result was a single or monolithic schedule. Except for the relative index, DC did not provide any aid to analyse the subject of a document into its facets and determine its character - in other words to determine the precise subject of a document. This led to inconsistency - both from library to library and within one and the same library. Generally speaking, it could take care only of old or known subjects. Even here, DC could take note only of some of the relevant facets of a subject. The other facets had to be ignored. This caused a difficulty of its own. Apart from this, the real crux of the problem was the one created by the continuous emergence of new micro subjects almost in every basic class. Their emergence is now very intense on account of the turbulent state of the universe of knowledge. This turbulent state is due to incessant research by large teams all the world over. This incessant research is due to the social pressure caused by population pressure. The new micro subjects gaining literary warrant could not be placed consistently within the sequence of the already existing subjects.

2.1.2 CC and Facet Analysis

CC sought to meet the difficulty with facet analysis. To this end, it attached a facet formula to some of the basic classes. A schedule of isolates was given for each of the facets. The facet formula gave the digit "::" as the connecting symbol between isolate numbers. A synthesis of a basic class number and of any isolate number taken from each of one or more of the schedules of isolates given for the basic class gave the class number of a subclass of the basic class. The ordinal value of "::" was defined to be less than that of 1 and therefore less than that of any digit whatever, forming part of a basic class number or an isolate number. As a result of this arrangement, a class number was polythetical unless it was a basic class number. Different subjects were formed by the different combinations, made of the basic class and any one isolate taken from the schedule for one or more of its facets. The sequence of the subjects so formed was
made automatically consistent by this facet approach, without the need to think out de novo, the filiation of any new subject with the old ones.

22 Consistent Sequence of Facets

22.1 Stage 1: Arbitrary Fixing of Sequence

The first stage was from 1925 to 1933. 1925 was the year of the first design of CC. 1933 was the year of publication of edition 1 of CC. In the first stage, the facet formula enumerated the permissible facets. It also fixed the sequence of the facets. This rigid prescription automatically secured consistent sequence of facets in all the class numbers belonging to any given basic class. However, the fixing of the sequence of facets in a facet formula was done by sheer flair. It was practically arbitrary.

22.2 Stage 2: Change of Sequence

The second stage was from 1933 to 1939. Edition 2 of CC was published in 1939. In the second stage, experience gained during reference service showed that the sequence of facets in some subjects was not as helpful as in others. The facet formula for Chemistry, for example, did not throw the subjects in Chemistry in a sequence felt by the majority of readers to be filiatory and helpful. It was so in Law also. The sequence of facets in their respective facet formulae was therefore changed. Edition 2 of CC gave such changed facet formulae.

22.3 Stage 3: Consistent Sequence at Deeper Level

The third stage was from 1939 to 1950. Edition 3 of CC was published in 1950. Till then, facet formulae had been given only for fundamental sciences, a few major applied subjects, the humanities, and the social sciences. Experience with Chemistry and Law has shown the risk involved in depending on mere flair in fixing the sequence of facets. The risk increased with the number of facets presented by a subject. In the third stage, documentation came into vogue extensively. The emphasis of documentation was on micro subjects. And a micro subject presented quite a number of facets — even exceeding a dozen. The need for the rationalisation of the fixing of facets and of their sequence was increasingly felt. It was not sufficient to secure consistent sequence of facets merely among the subjects falling within a particular basic class. This was rather trivial and it was easily secured. Consistent sequence of facets should be secured also among all subjects, whatever be their respective basic classes. This consistency should be sought at a deeper level in the idea plane. The words used to name the facets of various subjects in the verbal plane had been hiding the equivalence of the facets of different subjects at a deeper level in the idea plane. This was conjectured while writing Library classification: Fundamentals and procedure (1944). This led to the possibility of looking upon any facet of any subject as a manifestation of one or other of the five fundamental categories PMEST.

22.4 Stage 4: Removal of the Rigidity of Facet Formula

The fourth stage was from 1950 to 1957. Edition 5 of CC was published in 1957. During this stage, the PMEST idea was pursued. It led to:

1 the recognition of P, M, E, S, and T facets;
2 the concept of rounds among the facets of a subject;
3 the fixing of the sequence of PME in each round;
4 the fixing of sequence ST in the last round;
5 the concept of levels of each of the fundamental categories in a round; and
6 the design of distinct connecting symbols for the P, M, E, S, and T facets.

This removed much of the rigidity in the facet formula. And yet consistency sequence was secured in the sequence of facets, not only among the subjects of a particular basic class but among all the subjects, as viewed from the deeper level of the five fundamental
categories. This generalisation left the sequence of the subjects undisturbed. And yet, it provided for the mechanisation of the placing of any new micro subject unerringly in one and only filiatory and helpful position among the already arranged subjects.

225 Stage 5: Postulational Approach

The fifth stage is still current. It began about the end of 1956. Its chief feature is the basing of classification on a set of postulates. There is a separate set of postulates for the idea plane, the verbal plane, and the notational plane. There are also principles to regulate the assignment of a facet to a definite round and level. These were first conceived while teaching in the British Library Schools in December, 1956. These were developed in greater detail while lecturing in the Library Schools of USA, Canada, Japan and Ujjain between October 1958 and February 1959.

The postulational approach appears to be of good promise. A definite procedure of seven steps has been evolved for classifying a document on postulational basis, as illustrated in section 5 of my paper, Natural, classificatory, and machine languages. This mechanises facet-analysis and the construction of class numbers. Consistent sequence of facets and a consistent filiatory and helpful sequence of subjects are secured mechanically. No doubt cases arise transcending the principles, enunciated so far, for assigning a facet to the appropriate round and level. These cases occur often in the social sciences. However, there are reasons to believe that these difficulties can be got over by working with a large variety of subjects with several rounds and levels of facets — that is, depth-subjects. The opportunity to solve these difficulties will emerge as and when literary warrant increases in depth subjects. The solution is likely to take the form of additional principles to guide in assigning a facet of a subject to its appropriate round and level.

3 NOTATIONAL PLANE

The work in the notational plane is comparatively easy. This is due to:

1 the generalisation of the concept of ordinal numbers;
2 the use of a mixed notation;
3 the setting apart of the last digit of each species of digits as an octavising digit;
4 the concept of telescoped array;
5 the concept of telescoped facet;
6 the packet notation to implement subject device without introducing homonyms in the classificatory language;
7 the development of the idea of Environmented Entity; and
8 the introduction of schedules of phase and intra-facet relations. With all these features, the notational plane can implement any decision made in the idea plane. This is as it should be.

4 VERBAL PLANE

41 Need for Jargon

Verbal plane is concerned with words and terms in natural languages. It is full of homonyms and synonyms. It is always in a state of flux. Consistency is impossible in that plane. A classification schedule has therefore to improvise its own terminology — amounting to a jargon — free from the defects of terminology in a natural language. Indeed, it is best to have two verbal columns in the schedule — one in jargon and the other in natural language. The meanings of the terms in the jargon column are made clear by the canons of Context and of Enumeration.

42 Need for Alphabetical Index

Readers cannot name their subjects in class numbers. They can only mention them in a natural language. They have therefore to be provided with an alphabetical index to the class numbers. In general, a reader cannot name his subjects precisely. He generally names only a subject of greater extension than his intended one. On account of this, the purpose of the alphabetical index has to be essentially that of giving him the class number of any unprecise subject mentioned by him; this class number will put him into his region of the classified part of the main entries of documents;
there, in the context of the hierarchial display of feature headings, — in the helpful panorama presented by feature headings — he recognises his precise subject. He also finds thrown together under its name and class number — the entries of all the documents relevant to the subject he had in mind.

43 Need for Slimming the Index

A factor in the verbal plane causing difficulty in the building of the alphabetical index, is the necessity to use multi-worded terms for denoting most subjects — particularly depth subjects. Having a different index entry for each of the permutations of the words in the term is suggested at first thought. But it will swell the alphabetical index in the measure of the factorial function. The index part should be slimmed down to make it easy of use.

The Chain Procedure gives a consistent method of slimming. Stated in general terms, it prescribes rules

1. to choose for indexing one and only one permutation of all the words in the term;
2. to have another index entry with a heading got out of the heading in 1 by omitting its entry word alone;
3. to have another index entry with a heading got of the heading in 2 by omitting its entry word alone; and
4. so on, until reaching the index entry with one word alone as the heading.

The result is n index entries only instead of n! entries. Incidentally these are the likely index approaches by a reader mentioning a subject of greater extension than his intended one.

44 Sequence of Words

The sequence of the words in the term index is important. Even in an inflectional language, different sequences of the words often give different meanings. A non-inflectional language has to depend for the meaning of a term solely on the sequence of its words. The terms used in an alphabetical index has to be made of a term made of words in the substantative form without inflection and without the auxiliary words forming substitutes for inflection. We do not go here into the reason for this. The meaning of a term in the index is therefore dependent on the sequence of its words. Each and every permutation of the words may not have a significant meaning. But there are often quite a few permutations with significant and different meanings. The correct permutation should be chosen. "Periodicals (on) Reference Service" is different from "Reference Service (with) Periodicals". In both, the kernel terms are the same. It is their sequence that gives two different meanings.

45 Classification cum Chain Procedure

The postulates of classification and the principles for assigning each isolate to the appropriate round and level throw the isolate terms in a definite sequence yielding a definite meaning. The chain procedure derives the index entries from this definite sequence via the class number representing that definite sequence. Thus the pitfalls in the verbal plane are all avoided by classification and indexing by chain procedure.

5 CODING IN MACHINE LANGUAGE

In machine search, arrangement of entries, — that is, of subjects — may be totally ignored. For, its aim is the mere pulling out of the entries of all documents relevant to the question put, exhaustively. However, the machine has to be fed with all the facets of a subject at the stage of storage of entries. The isolate in each facet should be precise — that is, co-extensive with what it is in the subject. Similarly, at the stage of programming also, the machine has to be fed with all the facets of the subject of search. The facet analysis at the two stages should be consistent with each other. These three requisites in machine search can not be secured successfully if the coding is done directly from natural language. The coding
should be done from the classificatory lan-
guage. The more efficient the classification,
the more helpful it will be in the coding for
machine search.

6 WHAT SHOULD BE CODED

61 Specific Permutation

The next question is whether it is suffi-
cient to code severally the basic class num-
bers and each of the isolate numbers in a
class number. This would amount to coding
only a combination of them and not the speci-
fic permutation of them making up the class
number of the subject. Then all the difficul-
ties mentioned in section 44 will arise. The
questions arising out of this are:

1 Can the machine take into its memory,
as distinct entities to be remembered, each
of the $n!$ permutation of the facet in the $n$
facets of the class number — made up of the
basic facets and the isolate facets?

2 Can the machine pull out on order any
particular permutation needed?

3 Will a machine capable of doing 1 and
2 be far more costly than the ones designed to
handle combinations only?

4 Will it be within economic limits as
determined by management-tests?

62 Connecting Symbols

In the simple conventional machine of
Catalogue *cum* Class Number — provided
by a classified catalogue fitted with an alpha-
betical index having class index entries as
well as book index entries — all the problems
are easily met. This is made possible by the
designing of distinctive connecting symbols
for each of the five fundamental categories
PMEST. But in the application to coding for
machine search, this alone will not be suf-
ficient. There should be distinctive symbols
to indicate all the rounds of every possible
order, and all the levels of every possible
order within each round.

63 Problem for Designers

The designers of machinery for search
may well consider, not only the feeding of
basic class numbers and isolate numbers into
the machine, but also the connecting symbols
for the fundamental categories, and the
symbols indicating their levels and rounds.
Then the machinery for search will become as
versatile as the conventional Catalogue *cum
Class Number*. The problem for future work
in the design of machinery for research
appears in this form, when viewed from the
notational plane. Probably, the designers,
of machinery for search may be able to re-
state this problem in a different way. That
way may achieve the purpose in a more eco-
nomical way. Whatever be the way in which
the designing achieves this purpose, it will
also be available to take care of multi-phased
subjects, for which literary warrant is now
increasing.