Mr. Carey does not, of course, refer specially to indexes and in particular does not mention the modern tendency to confine the use of initial capitals to such main headings as would in any case possess them naturally. Nor, as far as I can trace, does he allude to this topic in his *Making an index*, which forms part of the same very useful series of C.U.P. pamphlets. Thus, to adapt certain entries in the index to the fourth volume of *The Indexer* (which, I am glad to say, does not adopt the new method), we might expect to see:

*How to Catch Trout* (Dickie), index to,
49-50
Hughes, Fielden, 'Introducing the Indexer', 95, 137
hymns, how not to index, 139
hyphenated proper names, 138

Robert Collison, in his *Indexing books* (Benn, 1962), states: 'It is not essential to capitalize the initial letters of ordinary words, but even if lower case is chosen, the printer will probably restore the capitals of his own accord'. Be such an instance of compositorial conspiracy as it may, my own preference, as I have already indicated, is for capitalizing the initial letter of every main heading. This is the old-established method and seems to add dignity and finish to the appearance of the printed page.

It would be interesting to hear the views of other index-makers** on this subject.

G.N.K.

**I have used this old term deliberately, since as a result of my name's appearing in the Buff Telephone Book under the heading of 'INDEXERS' I am being constantly rung up with requests for my terms for constructing those marginal thumb indexes to be found in certain trade publications!

---

**SCIENTIFIC AND TECHNICAL INDEXING**

*E. J. Coates*

In case the title of this paper carries the implication that there is something peculiar or fundamentally different about scientific and technical indexing which marks them off from other kinds of indexing, I should make it clear that I do not think this is so. Very few of the points with which I shall deal will have no application to indexing in other subject fields. The main factors which from the indexer's viewpoint distinguish science from non-science material are questions of degree only. Scientific literature contains a greater number of concepts in toto and a far higher proportion of precisely defined concepts than does the literature of the humanities, so that from the point of view of the multiplicity of concepts scientific indexing looks at first glance formidable, but from the point of view of the battle between words and meanings, the scientific indexer gets off relatively lightly, I say relatively because, despite all essays at standardised nomenclature, there is plenty of ambiguity in scientific, and more particularly technological, terminology. But at least, despite all that is justifiably said about the sheer incompetence of much scientific writing as communication, the scientific author usu-
ally does manage to convey the definite topic about which he is writing. The same cannot be said in all fields.

The second difference of degree concerns the consumer side of scientific indexing. Scientific activity, as you know, has an increasingly serious information problem upon its hands, and a great deal of attention is at the moment being focused upon scientific indexes. Some elaborate experiments are being carried out on the efficiency of scientific indexes, at Cranfield in this country, and elsewhere. Taken all in all it would be fair to say that scientific indexes are under particularly close scrutiny and at the present moment rather more is being expected of them than is the case with other types of indexes.

Some of the interesting things that are being done in connection with scientific indexing are worth considering for the light they throw on the basic purposes and aims of indexing generally. There is, for instance, the KWIC indexing system, a not too distant variant of which is now being used to provide a current subject approach to chemical literature. KWIC was invented by one of the giant computer-manufacturing concerns seeking to develop fresh uses for computers. It is meant for the indexing of scientific papers, though it is possible to conceive how it might be used for book indexing. Briefly, the computer is programmed to take each word in the title, with some exceptions which I will mention shortly, and to print out the title repetitively under each selected word. This is achieved, without disturbing the word order, in this fashion:

For the given title ‘Efficiency and transparency of cheap liquid scintillators’ the following is produced:

```
AND TRANSPARENCY OF  CHEAP LIQUID SCINTILLATORS. EFFICIENCY
LIQUID SCINTILLATORS. EFFICIENCY AND TRANSPARENCY OF CHEAP
TRANSPARENCY OF CHEAP LIQUID  SCINTILLATORS. EFFICIENCY AND
EFFICIENCY AND TRANSPARENCY OF CHEAP LIQUID SCINTILLATORS
```

Now you will see that you get an entry under all words in the title except articles, prepositions and conjunctions. It is of course quite easy to give the computer a list of articles, prepositions and conjunctions and instruct it to ignore these; not so easy, and indeed not attempted, to give it a formula enabling it to reject the valueless entry under CHEAP.

I mention KWIC because it is an outstanding demonstration of indexing at its lowest level of sophistication. All that it requires is that words given in the text (or title, for a bibliographical index) should be picked up and manipulated into alphabetical order. Even at this low level a certain discrimination is needed to reject non-significant words, which no one will ever want to look up. We see that the computer can only discriminate to the extent of rejecting a predetermined list of non-significant words, and that usually this list comprises only articles, prepositions and conjunctions. This question of deciding what is or is not significant is the most pervasive unsolved problem of indexing at every level. When we ask ourselves about significance, we should ask ‘significant to whom?’ It is a fact that the significance problem is in inverse proportion to the extent that we can visualise the needs of the person who is going to consult the index, and in general this means that it is an acute problem for an index covering a wide subject field of interest to many different kinds of specialist
interest, and less acute for indexes covering narrow subject fields. Even the word CHEAP might be significant in a bibliography or book on 'How to sell scientific instruments'. We cannot easily overcome the significance problem, which exists in scientific indexing as in other indexing, though we may be able, as in the KWIC system, to whittle it down just a little.

I pass now to what can be regarded as indexing of the middle or second degree of sophistication. Before leaving KWIC behind, I should like to say that for all its wide-open loopholes which will be obvious to all of you, it is doing a job which at the moment is too big for human indexers, or what is more important, needs to be done more rapidly than can be carried out by human indexers. It has the virtue of its name. It is quick, and speed is of paramount importance in scientific literature indexing. The scientist, particularly the applied scientist, just cannot afford to wait for the leisurely index which comes months or years after the event to which it refers.

The second level of indexing to which I now wish to turn is distinguished from the first in that the indexer is no longer simply a manipulator of words, but someone who tries to isolate the concepts signified by the words. His task is to utilise words to indicate the concepts lying behind the author's words—an invidious task, and perhaps logically doomed to fall short of 100 per cent. success. For this process of signposting the concept behind the word, conventional jargon has given the name of 'vocabulary control'. Vocabulary control means nothing more than that one identifies synonyms, or phrases of equivalent meaning among the index words which one has collected from the text, and having identified a pair of synonyms, one either lists under each all the references already given under the other member of the pair, or alternatively one assembles all the references at one term only and makes a cross-reference from the other.

Of the various features which contribute to the practical utility of a book index, none is more important than synonym control. Two questions arise in this connection to which no easy answers can be given. The first is, how does one spot synonyms, or what can one do to ensure that synonyms are not missed and entered as if they stood for different concepts. The rule of thumb method is that one notices or remembers that there was another word which meant the same thing, but the larger the index the greater the load placed on memory and the more synonyms will slip through the net. The only device for checking synonymy of terms is classification. Classification is simply the ordering of terms into a systematic pattern of likeness of meaning, so that terms that mean nearly the same thing are close together. Each term in the index is assigned to its proper place in the classification and when one finds oneself wanting to put a term into a position already occupied by another term here is the evidence of likely synonymy. It is a formidable task to have to make one's classification and then, as it were, to fit every index term into it—there is however no other way of systematic checking for synonymy. The making of classifications is of course a matter about which librarians have been heavily exercised for a long time, but I do not think that any existing ready-made classification is ideal as it stands, for the purpose of synonym control. The basic technique for constructing classifications is not difficult to acquire, and becomes easier with practice: I suggest that anyone interested would do worse than to read Faceted classification
by B. C. Vickery, in which the author explains very clearly how to approach the classification-making problem.

The second question which arises in synonym control is this: if I decide to prefer one term and make cross-references from the others, which do I prefer? If you are indexing the work of a single author, the author's preference will surely determine yours, but if you are indexing a composite work of multiple authorship, or if you are doing bibliographical indexing, the problem is not always so simply resolved. Attempts have been made in some scientific fields to standardise terminology, but these standards often go against strong conventional preferences. This is particularly the case in chemistry where considerable steps towards systematisation of nomenclature have been made, but where also the briefer so-called trivial names strongly persist in the literature. The names of technical processes often take a long time to settle, and the indexer is frequently called on to take a decision long in advance of a conventional preference. An example which comes to mind is the bulking process applied to yarns to give them softness to the touch, apparent elasticity, and varied thermal insulation properties. The process is variously called bulking, crimping, false twisting, texturing. It recently appeared that texturing was winning the day, but there is now a new variant 'texturising'.

It is not always the logical names which prevail against the more superficial ones. 'Hovercraft' is firmly ascendant in Britain for a certain novel type of vehicle, despite the existence of two other names which refer to more fundamental properties. 'Air cushion vehicle' is logically preferable, because it refers to a more concrete attribute of the vehicle, and anyway there are other vehicles which are not hovercraft which nevertheless hover. Perhaps even more to be preferred is 'ground effect machine', which had a good start in the literature, and formed a pleasant little acronym GEM, but which is rapidly falling out. The direction in which technical advances are made does itself help to establish a conventionally preferred name. I would guess that Hovercraft will eventually give way to Air Cushion Vehicle because the same air cushion principle is now finding applications elsewhere, for instance in the handling of delicate strip material at the strip mills. It is not possible to bring the word Hovercraft into a name for these air cushion bearings, though they have been called Hover Pulleys. There is little more that can be said about correct choice of synonym except that, like nearly every other indexing problem, it becomes harder in proportion to the width of subject field being covered.

I have dealt so far with two levels of sophistication in indexing. At the first level, the essential operation was one of manipulating given words, and at the second the concern has moved on to the listing of concepts, by means of words it is true, but by means of words to which synonym control has been applied, so that the index user's luck no longer depends upon the particular term under which he first thinks of looking. At both of these levels there is also the common underlying question of what is a significant term and what is not.

Both of these levels in various admixtures are the common characteristic of indexes to books by individual authors, and I think we can say that the better sort of book index is normally the one operating on level two. The third level is normally appropriate to indexes to compilative or collective works with contributions from various authors, and to
bibliographical subject indexes. It is distinguished from level two in that it includes connective references between related terms in the index.

It is easy to see why this is generally only appropriate to the indexing of material of heterogeneous origin. In the text by an individual author, the arrangement of the material itself is determined by the relationships between the various parts of the subject that the author thinks important. I doubt if it is part of the indexer's job to indicate other relationships, or even to reproduce those implicit in the arrangement of the text. But you will see that different considerations apply in the case of indexes to heterogeneous material or to the contents of large numbers of individual items such as scientific papers.

Now anyone who sets about constructing a network of relational cross-references quickly realises that there is no end to this process, and the network very quickly becomes a maze. The problem is how to put a limit on the process, and the general answer is to confine oneself to showing one type of relationship only. For most purposes the most useful relationship is that of inclusion. Make connective references from a term to others included within its meaning as from INTERNAL COMBUSTION ENGINE to DIESEL ENGINE and GAS TURBINE or from FISHING VESSEL to TRAWLER, from POLYAMIDE FIBRE to NYLON. But here again we do not go very far before being beset by difficulties. For any given term, there are usually a number of possible including terms. For instance, we might legitimately insist that GAS TURBINE should have a cross-reference from the inclusive term TURBINE, and NYLON might equally well be regarded as coming under THERMOPLASTICS, as under POLYAMIDE FIBRES. So we need not only to limit ourselves to inclusion relationships but we also need a device to help us sort out the inclusion relationships themselves. We find such a device in classification schemes, for these are essentially attempts to lay out multidimensional relationships in linear form.

So we see that classification is relevant to indexing in two ways. First of all on what I called level two as a helpful means of detecting synonyms, and now also as providing a basis for a network of relational references. Once again I have to repeat what I have said in other connections, this particular aspect of the indexing task, like most others, is easier in proportion to the narrowness of the subject field being covered. The wider the harder.

I want now to turn to another aspect of indexing work which in some degree applies to all fields but is especially prominent in indexing scientific and technical material. Most of the concepts which we wish to index are composite in character, that is to say they cannot be expressed in a single word, but require a phrase, sometimes something almost approaching a sentence. For example, DROP FORGING. DROP FORGING HAMMERS. VIBRATIONS OF DROP FORGING HAMMERS. DAMPING THE VIBRATIONS OF DROP FORGING HAMMERS. Let us concentrate on the most complex example:

DAMPING VIBRATIONS OF HAMMERS FOR DROP FORGING

Why don't we index the subject just in that form? The answer is not that the subject is too long to express—we have to accept that subjects of that degree of complexity are sought by people who consult indexes, though they may not start with exactly those words or that order of words
in mind. The phrase as it stands is unacceptable for indexing purposes because it contains prepositions which are always non-significant words. So let us then take out the prepositions: We are left with

**DAMPING, VIBRATIONS, HAMMERS, DROP FORGING**

This composite term entry remains intelligible even when the prepositions are dropped because the order in which the five significant terms appeared in the natural language phrase is still retained. If we start shifting the five terms round very much, we soon begin to lose the sense, for very good logical reasons. The damping is directly related to the vibrations and only related to the hammer through vibrations, so that we have a linear chain of relationships. If we start to tangle up the straight chain, then intelligibility suffers. But you will quite rightly insist that the index should provide a lead-in to this composite concept from each of the significant terms, VIBRATIONS, HAMMERS, DROP and FORGING. How does one manage this and yet at the same time avoid the chain-tangling that I have mentioned? There are two (and perhaps more than two) fairly simple ways of doing this. Can I recall the technique of the KWIC index, and rotate the terms as follows:

**Entry 1**  
DAMPING, VIBRATIONS, HAMMERS, DROP FORGING

**Entry 2**  
VIBRATIONS, HAMMERS, DROP FORGING/DAMPING

**Entry 3**  
HAMMERS, DROP FORGING/DAMPING, VIBRATIONS

**Entry 4**  
DROP FORGING/DAMPING, VIBRATIONS, HAMMERS

**Entry 5**  
FORGING/DAMPING, VIBRATIONS, HAMMERS, DROP

The sense in these cases remains reasonably clear because the linear chain has been broken once only, though I think that entry 5 suggests that breaking between a term and a qualifying epithet may never be justified. A variant of this method that is sometimes used is the following:

**DAMPING**  
Damping, Vibrations, Hammers, Drop forging

**VIBRATIONS**  
Damping, Vibrations, Hammers, Drop forging

**HAMMERS**  
Damping, Vibrations, Hammers, Drop forging

**DROP**  
Damping, Vibrations, Hammers, Drop forging

**FORGING**  
Damping, Vibrations, Hammers, Drop forging

One point to be noticed is that the linear chain may be used equally well in either direction:

FORGING, Drop, Hammers, Vibrations, Damping

In this case to translate the index entry into natural language, one reads the elements backwards, Damping of Vibrations of Hammers for Drop forging. The backwards form is in fact used in the British Technology Index. Where it is an economy to use references in the index instead of direct entries, the problem can be approached in another way. In this illustration I will use the reversed form just mentioned, though the forward form will do equally well. We start off with a full entry:

FORGING, Drop, Hammers, Vibrations, Damping

Next we make references as follows from each of the other four significant terms:

DAMPING, Vibrations, Hammers, Drop forging. See FORGING, Drop Hammers, Vibrations, Damping.

VIBRATIONS, Hammers, Drop forging. See FORGING, Drop Hammers, Vibrations.

HAMMERS, Drop forging. See FORGING, Drop, Hammers

DROP FORGING. See FORGING, Drop.

The point to notice here is that only the first of the above references details the
complete concept. The others are generic references, the import of which is for instance that *something* on DROP FORGING, not necessarily a comprehensive account, is recorded in the index.

This is a very sketchy glance at the problem of ordering the elements of composite concepts which require to be cited in indexes. The practical solution is not always as simple and straightforward as in the illustration given, but on the other hand it is not impossibly complex once its measure has been taken. The indexing in the *British Technology Index* is fairly sophisticated as indexing goes, yet we have been able to reduce most of our problems to a dozen basic situations. I would add, however, that the residual issues are often of very considerable difficulty.

Perhaps I may end with a mention of three persistent danger points which call for the utmost watchfulness.

The first of these relates to homonyms, which abound in technical literature, because of the engrained habit of electrical, electronic, nuclear and even, I think, chemical engineers, of borrowing words from the field of mechanical engineering and assigning them by analogy to new objects and processes. Thus the electrical engineer has taken over FILTER and BRIDGE to mean a particular circuit, the electronic engineer has, of course, TUBES or VALVES only very tenuously related to the tubes and valves met in mechanical engineering. The chemical engineer ‘cracks’ hydrocarbons, and the nuclear engineer contemplates the ‘burn up’ of uranium, though the process is fundamentally different from what is meant by burning in other contexts. The practical problem is to avoid alphabetical inter-spersion of homonyms of this kind:

- BRIDGES, A. C.
- BRIDGES, Arch

BRIDGES, Bascule
BRIDGES, Decks
BRIDGES, Electrical
BRIDGES, Girders
We can improve this by inserting a qualifier:
- BRIDGES, Arch
- BRIDGES, Bascule
- BRIDGES, Decks
- BRIDGES, Electrical
- BRIDGES, Electrical, A.C.
- BRIDGES, Girders

However, we still have blocks of material on structural bridges sandwiching the electrical ones. We can further improve matters by attaching to the introduced qualifier term some punctuation device with an arbitrarily assigned ordinal value greater than Z. But we lose in other directions. Arbitrarily assigned order, even of the simplest kind, leaves wide-open pitfalls to the user who is usually prepared to accept the burden of knowing A to Z order, but not one jot or tittle more. We try to meet the more difficult homonym problems in *BTI* in this way:

- BRIDGES, Decks
- BRIDGES, Electrical. See after the last sub-heading on structural bridges
- BRIDGES, Girder
- BRIDGES, Zurich
- BRIDGES, Electrical
- BRIDGES, Structural See subheadings preceding BRIDGES, Electrical

The second general difficulty to which I should like to refer is that connected with the subject naming of pieces of equipment which are so novel in character that at the time of indexing they have no settled name. It is instructive to consider generally how objects are named. Any object has the basic properties of shape, material function, and sometimes method of working. If you consider established names of man-made articles you will find that usually the name refers to function. As an example, a familiar object comprising a thread of mercury in a graduated capillary tube is called a thermo-
meter in reference to its function. It is not called a mercury capillary, or a mercuric expansion column or anything of that sort. The moral for indexers is fairly clear. A new, as yet unnamed object is to be described primarily by its function, or by its shape if its function is multifarious. Thus a new textile fabric would be designated in BTI as

CLOTHING (Function) Fabrics (Form)
Polyester fibre (Material)

If it is used for a variety of purposes beside clothing it will be entered as

FABRICS (Form) Polyester fibre (Material)

If it is a question of a new material only, which might be made in the form of either yarns, cords, or fabrics, then we enter simply under the chemical name of the material:

POLYESTER FIBRE

The third problem is the most intractable of all. Many technical objects and processes possess properties which are nearly always applicable, taken for granted and therefore not mentioned, until a new development arises in which the object or process turns up with this normally taken for granted property absent. For example, nearly all the literature on welding is actually on fusion welding, that is to say on welding in which solid metal is converted to the liquid state by heating. Unfortunately this fusion property is taken for granted until a new technique is devised which does not involve actual melting of metal. At some stage we have to go back and amend our material under WELDING to WELDING, Fusion. The difficulty is to know how soon, after what delay, to do this. Another concept in technical literature which often illustrates the same difficulty, is that of METAL. There are many names of processes which imply metal nearly but not quite all the time. MACHINING as used in the technical literature nearly always means the machine forming of metal, but it can be extended to wood, glass, and plastics. WELDING and CASTING usually, but again not quite always, imply metal. The printing trade journals often talk of PRINTING when they actually mean letterpress printing, and if one follows this in indexing one is never sure whether material under PRINTING is about printing generally (letterpress, lithography, gravure and so on) or about letterpress alone. Perhaps as a general rule it is preferable in the long run to err on the side of pedantry.

The stratum of snags and difficulties I have attempted very sketchily to explore, is that which arises once you have cleared essential questions of definition, of the meaning of words, out of the way. Because science and technology deal proportionately with more concrete and definite topics than do other fields, this layer of difficulty is reached earlier in scientific and technical indexing than in indexing in the humanities field. I think, however, that it is to be found in indexing of all kinds.

NOMINATIONS for COUNCIL

Members are reminded of rules 7 (a) and 7 (b) of the Constitution that officers of the Society shall be elected at the Annual General Meeting, but that retiring officers are eligible for re-election. Also that the six members of the Council shall retire in rotation, two at each Annual General Meeting, and that they also are eligible for re-election. There is one further vacancy to the Council because of the retirement of Miss A. M. McGrath.

Regarding nominations for office and membership of the Council, attention of members is drawn to rule 7 (c) of the Constitution. Nominations must be received by the Secretary, duly signed, by April 21st, 1966.