AIMS AND METHODS OF THE BRITISH TECHNOLOGY INDEX

E. J. COATES *

This paper deals with an index which has been in existence for less than two years, but the wider problem to which the B.T.I. is intended to be a very partial and limited answer has grown upon us rather insidiously over the past thirty or forty years. This is the problem of communication in science, pure and applied, which has arisen partly because of the volume of scientific literature, which is now so great that the individual specialist is less and less able to keep up with the writing on his own subject, and secondly because of the steady melting away of the hard-and-fast boundary walls which formerly separated one scientific or technical specialty from another. Fundamental advances in knowledge and innovations in the application of knowledge have exposed connections not hitherto suspected, and so we are approaching the situation in which a really up-to-the-moment awareness in one's special field calls for a simultaneous knowledge of what is going on in several other fields. Thus for instance the petroleum chemist working in lubricating oils may find it essential to keep abreast of the latest developments in the design of mechanical bearings, the food processor may be deeply interested in the development of plastic films for packaging, and the plastics specialist may have a considerable preoccupation with the problems of corrosion. There is therefore a great need for indexes and other information-processing compilations which will enable pure and applied scientists to survey not only what is going on in their own field but in marginal fields as well. The problem is, as you can imagine, no easy one and it is made even more intractable by the fact that the literature concerned is written in a variety of languages.

The British Technology Index confines itself to applied science, interpreting that term fairly broadly, though excluding medicine and agriculture. Its source coverage consists of 400 British technical journals. There must be very few British technical journals of importance not covered by this list, but we have to remember that it is still a fraction of the world's output of technical periodical literature. However, it is a sufficiently large fraction to be worth recording. The Index is published monthly and comprises a subject-index to between 2,000 and 3,000 articles per issue, and the articles thus indexed will in most cases have appeared within the preceding seven weeks. This is an extremely short time lag as indexes go, and currency occupies first place in the Index's priorities.

The Index has a counterpart in the United States, but no competitors. Indeed, two earlier attempts to provide a current guide to British technical periodical literature failed for inadequate support. The reasons for these failures are not very clear, but it might not be too far-fetched to suggest that to some extent they are attributable to the arrangement of the entries. These were grouped under broad subject headings corresponding to the organisational diversification of industry, and it was not usually easy to trace highly specific information. It was there, but was not signposted in detail and had to be searched for without the assurance that it

was there. Periodical articles are, of course, not the only form in which technological developments are reported, but they are by far the most important form. One kind of guide to this periodical literature which is perused by many engineers and technologists is the abstracting service. These services provide a digest of each article they record. They cover many, though not all, of the branches of technology. The material they contain is usually broadly grouped by subject, sometimes placed in systematic classified arrangement, and sometimes bear significant catchwords in the margins of each page. The time lag between the publication of an article and its appearance in abstracts is usually three months or longer, but with one or two exceptions detailed subject indexes to the abstracts are published only annually, if at all. Though I emphasise that indexes and abstracts have quite separate rôles, I have probably said enough to indicate the gap that B.T.I. is intended to fill, namely the three months immediately following the publication of the articles when there is no record of them at all, and the longer period before annual subject indexes to abstracts make their appearance. Our purpose is therefore to provide a guide to very recently published technical articles which will signpost specific subject information, without asking the user to read through broadly grouped headings on the chance that what he requires may prove to be there. Values of any index include not only the fact that it tells you what is available, but also that it tells you at once, by implication, what is not there.

To secure the result desired, alphabetical specific subject arrangement is followed. By this I mean that any subject which can be fully described in a single word or phrase is entered directly under the word. Thus an article on the new printing technique FILMSETTING is entered directly under that term and not under COMPOSING or PRINTING. However, there are not many subjects found in the technical literature which can be covered unequivocally by a single word. Most require a combination of words. Composing—the assembly of type in printing—is a case in point. We cannot just use “Composing” alone, because in isolation it is ambiguous, even in an index limited to technology: it could mean composing in industrial design, in photography and possibly in other spheres. We need to add a further word or words to make clear what we mean. We can say either Composing, Printing, or Printing, Composing, and here we run straight into the problem—the primary problem in all subject indexing, namely, the order in which we should cite the various words in such a composite heading. With a two-element heading such as Composing, Printing, there are two alternatives only. But a three-element heading faces us with six alternatives, and there are 24 ways of arranging the individual terms in a four-element heading, of which we find ourselves creating a great number in specifying the subjects of technical articles. So you see that we have a situation which calls for some rigour in deciding element order if we are not going to find ourselves overwhelmed with permutations.

Before going into this further, I should like to draw your attention to one or two points on the general structure of the Index, particularly to the cross-reference structure. Two kinds of cross-references are employed. When we have a subject heading consisting, say, of three elements we do not produce entries under each of the six possible permutations. We choose one, in most cases, on principles which
I shall explain shortly, and we refer from each of the terms used as a sub-heading if we think it is likely to be a sought term at any time. Most of these sub-heading terms are sought terms, but there are some very general ones such as "Manufacture", "Measurement", "Research", "Equipment" which we think will never be used as approach terms. These references from sub-headings are made almost mechanically by reciting the terms of the heading from right to left, inserting the word "see" and then giving the same terms in their left to right order. Thus if we have an article on the Transport of molten aluminium by road, the entry will take the heading:

ALUMINIUM, Molten, Transport, Road

The references will be:

(1) ROADS, Transport, Molten aluminium. See ALUMINIUM, Molten, Transport, Road

Notice that we use the plural form when we place ROADS in the first position in the reference:

(2) TRANSPORT, Molten aluminium. See ALUMINIUM, Molten, Transport

(3) MOLTEN ALUMINIUM. See ALUMINIUM, Molten

You will note that the reference (1) tells the searcher for road transport of molten aluminium that the subject is in the index at once. Whereas (2) and (3) tell the person who comes in at Transport or Molten that it may be worth his while to proceed as directed to the actual entry under Aluminium, where he will see whether the subject required is present. Quite a number of enquirers wanting road transport of molten aluminium specifically will in fact try first Transport or Molten and (2) and (3) represent the provision made for them. Now this method of carrying molten aluminium around in a trailer is, I think, the first attempt of its kind. The idea may catch on, in a few months there may be another article, this time upon transporting molten lead by road. The entry heading will read:

LEAD, Molten, Transport, Road

Now an index of this kind ought to indicate in some way to the searcher when he comes in at LEAD, Molten, Transport, Road, that the earlier material exists. To make a direct reference to it would, if it were the general practice in all such cases, swell the index perhaps four or fivefold and would raise very awkward problems in production. Fortunately such provision is not essential, for we also make the routine references:

ROADS, Transport, Lead. See LEAD, Transport, Road

TRANSPORT, Molten lead. See LEAD, Molten, Transport

MOLTEN LEAD. See LEAD, Molten.

The point is that these references fall next to or very close to the similar set of references for the case of the aluminium, so that the enquirer interested in the road transport of molten lead who is also interested in similar handling procedures for other molten metals, can discover whether the index contains anything about them by looking up in the main sequence the subheading terms MOLTEN, TRANSPORT, Molten and ROADS, Transport, Molten. The alphabetical index can be made to disclose the presence of such related material, provided that the entry
heading is sufficiently specific in the first place and that systematic referencing is carried out.

This is the general picture of the inversion references in B.T.I. In fact the close-up picture is a little more elaborate than that. When we have an element in a sub-heading which specifies the type or kind of the preceding element, our procedure includes a permutation. I give an example:

ALUMINIUM, Molten, Transport, Motor vehicles, Articulated

Articulated specifies a type or kind of motor vehicle and the scheme of references in this case is:

ARTICULATED MOTOR VEHICLES, Transport, Molten aluminium
MOTOR VEHICLES, Articulated, Transport, Molten aluminium
TRANSPORT, Molten aluminium
MOLTEN ALUMINIUM

You will see that in reference (2) here we go first left to right, and then right to left, leaping over the starting term. The second complication is that in a certain limited number of cases we give two entries instead of an entry and a reference. We do this when the entry heading and first sub-heading are in the relation of the whole to part. The point here is that an engineer's interest in a component as such, without regard to its particular application, may be very great indeed. Thus you will find in the Index identical entries under:

SHIPS, Diesel engines

and

DIESEL ENGINES, Ships

We have one other occasion when we use plural entry in the Index. This is when the subject of the article is a proper name, such as the BUILDING CENTRE or DEPARTMENT OF TECHNICAL CO-OPERATION. We give here an entry under the proper name and a further entry under the subject which the proper name entity illustrates.

The second type of reference which we have in the B.T.I. is not derived from the sub-headings, but from library classification. These references direct the enquirer from a general to a more restricted heading. This is a necessary facility in any index of this kind, because quite a high proportion of index users begin by looking up a term with a broader meaning than the topic on which they actually need information. So we need a chain of references modulated step by step down from the broadest to the narrowest headings. The classification scheme from which this system of reference is derived is mainly the Universal Decimal Classification, chosen because it was the scheme most familiar to most of the indexers. It is not entirely satisfactory for our purpose, and our reference structure is bolstered up at particular points on the technological map by other more recently produced special schemes. Improvement in this reference structure probably awaits further fundamental work in library classification. In the meantime it must depend on such classifications as we have, if it is not to proliferate beyond controllable limits. All knowledge is multidimensional in the manner in which the various parts are related. A classification scheme is simply an attempt to plot these multidimensional relationships upon a two-dimensional page.
This question of the citation order of the elements in composite headings on which I have already touched is of importance, because we want our indexing habits to be predictable to users. I refer here not only to consistency in assigning headings to the same subject but also consistency in overall pattern which regulates the analogous treatment of unconnected or remotely connected subjects. While we should not require our users to know what our indexing rules are, they can be expected to absorb unconsciously a pattern that is consistently offered to them, and this is of considerable utility because it enables them to anticipate how the index will handle a new subject. You will appreciate that every month we index such a high proportion of new developments that there is no authoritative list to which we can turn and hope to find answers. We have to construct these headings as we go and it is essential both from the point of view of the indexers, who should know clearly what they are doing, and also of the users, that we construct them according to reasonably definite rules.

In general we arrange our composite headings in order of decreasing concreteness—so that the most concrete term in the compound takes the leading position. However, in technology we frequently require composite headings in which there are two or more elements of equal concreteness, e.g. Corrosion of iron, by wood acids. To meet this situation we have been obliged to go into the question of relationships between the individual terms in the compound. Some of you will remember the highly original paper given to your Society in 1961 by Farradane on Relational Indexing.* Though I do not fully understand all of Farradane's views, nor agree with all that I think I do understand, some of his ideas have helped us a very great deal to formulate systematic rules of attack on subject-heading problems. We think we have isolated 15 distinct varieties of composite headings, of which 10 are proving to be important. In the first five of these ten, one term is a Thing and the other specifies a particular kind or variety of the Thing. I give some examples:

(1) **FABRICS, Foamback**
(Thing₁, Thing₂)
(Fabrics with foambac

(2) **TRIMMING, Presses**
(Function, Thing)
(Presses for trimming)

(3) **BUSES, Garages**
(Thing₂, Thing₃)
(Garages for Buses)

(4) **FABRICS, Elastic**
(Thing, Property)
(Fabrics with elastic property)

(5) **BUSES, Bodies**
(Thing₂, Thing₃)
BODIES, Buses
(Thing₁, Thing₄)
(Bodies of Buses)

The other five important types of headings are concerned with properties and actions upon things and by things. Some examples follow:

(6) **BEAMS, Strength**
(Thing, Property)
(Strength of beams)

These rules take care of most of the problems of term-order in compound headings, though there are some residual points awaiting further experience. As practical indexers you will not need to be told that this is not the end of indexing problems. B.T.I. is fully exercised on the pitfalls in synonym control. This is in one respect made relatively easy (I stress 'relatively') for us, because technical writers have the habit of definition better ingrained than writers in most other fields. On the other hand, B.T.I. covers about 50 recognisably distinct subject specialties, and quite often different specialties use different terms to mean the same thing or action. We take two precautions. We are as generous as possible in referencing from synonyms—for it is the arrival of the previously unreferenced synonym which is liable to trip up an indexer. Secondly we do refer all new concepts to their context in library classification schemes. There is a presumption that two terms which classify in the same place in the scheme may be synonyms.

Homonyms abound in the technical literature, because in applied science new concepts are often named by appropriating a term from someone else's field (usually the mechanical engineers') and investing it with a new meaning. We have not yet found a fully satisfactory solution to this problem which is consistent with mechanical alphabetical order.

Yet another awkward problem which easily produces errors is concerned with what may be called "nearly universal properties". Nearly all tyres today are made of rubber, nearly all workshop practice concerns the making of shapes in metal, nearly all welding is fusion welding. Our policy to date has been to omit what we call "nearly universal" property terms as a contribution to brevity and avoidance of pedantry.

A brief account of the way in which the indexing is organised may be of interest. We have a staff of four library qualified indexers and two clerical workers, and the transcription of titles, author, journal, date and page details is done by the clerks by typewriter on continuous stationery. The actual clerical effort required from indexers at this stage is limited simply to writing the appropriate subject headings on the entry. If the indexer decides that it is a subject which we have not handled before, then he also has the task of setting up the new piece of reference structure. At a later stage there is however a considerable amount of what
could be considered high-grade clerical work which falls upon the indexers rather than upon clerks. Generally these jobs are of the sort in which errors must be kept to the absolute minimum if the indexing routine is to function at all satisfactorily. All in all, each indexer spends about half his or her time actually on indexing; the other half is spent on such tasks as reading the master copy, which is prepared on cards for litho reproduction, and maintaining the authority file of all our decisions. I should mention that the master copy is typed by Varityper (by a contractor) on cards each of which contain one line only of copy. The principle is that the cards thus typed are passed at high speed through a special camera which photographs each line on a film one below the other. If one card gets out of order you get a printer's error of a fairly common kind, but if it is a heading which comes adrift and finally rests upon an entry to which it does not belong then we have the serious matter of the Index giving wrong information. About 18,000 of these cards have to be sorted into as nearly as we can make it perfect order for each monthly issue. When the set of 18,000 have been through the camera for the month's issue they then have to be interfiled with the cards for earlier months of the year to form the annual volume. This interfiling calls for the sorting of about 160,000 cards, and when this is complete they all go through the camera again to form the Annual Volume of about 900 pages.

In the context of this considerable load of clerical work, one naturally thinks of the possibility of further mechanisation of production processes. B.T.I. operates in a half-way house on the road to mechanisation, with advantages in speed of production, but with some slightly bizarre results relative to the division between indexing and clerical labour. Computers can, of course, be used to perform some of the purely clerical operations. Indeed, in America some very large scientific literature indexes are being produced by computer. These rely for their effectiveness upon the appearance of sought subject terms in the title. There is, however, no machine which will, as it were, go behind the title and summarise the actual subject content. We have not yet discovered how to break down the summarising process into the series of simple instructions which a machine can handle. I will not say that no research is being done in this direction, but it is scarcely beyond the stage of experimentation based on word-frequency counts. At the moment, therefore, the essential indexing operation can be carried out only by human beings—and I think can be carried out well only by human beings with the requisite qualities of temperament, such as a methodical mind, coolness and reflection in facing what is initially incomprehensible, interest in the subject matter in breadth, and the ability to remain open to fresh knowledge. These qualities are not very frequently found in combination and their comparative rarity is a little discussed but real issue in the information-retrieval problem.

"The lack of an index completely destroys what little value the book might have had"—John Bulloch, reviewing They call it intelligence: spies and spy techniques since World War II, by Joachim Joesten (Abelard-Schuman) in The Daily Telegraph, August 23rd, 1963.