I want to explain straight away that in talking to you for about thirty minutes on the indexing of scientific books I am not giving a formal lecture and am not intending to be in any way dogmatic. By this I mean that I am not claiming the opinions I shall express are necessarily right or the methods I shall describe are necessarily the best ones, but am putting them forward simply as something for you to shoot at, in order to start a discussion on some moot points which ought to be raised and discussed.

In one of the early numbers of The Indexer, to be precise in Vol. I, No. 4, Autumn 1959, I contributed an article entitled 'Some ideas on indexing' which began with the definition that indexing in the wider sense of the word includes 'any device for discovering or rediscovering in a book, or in a collection of papers or notes, such items of information or passages of text as have a wanted relevance'. By the words italicised here I meant that the purpose of an index is to supply an answer to either or both of two questions:

(1) Does this book (or this file or collection of papers) contain anything that might be useful for me as regards the topic I am now thinking about or wanting to know about?

(2) Knowing or already having reason to believe it contains something I want, on what particular page (or on what particular bit of paper) shall I find that?

I went on to examine the fundamental principles by which provision can be made for answering these two questions, pointing out that essentially an index has to be some kind of classification of symbols which serve to characterise individual items of information contained in the book or file. What the user of an index has to do is first of all to decide which symbol would correspond with what he is looking for, then to locate that symbol by searching in the index. If it is there, he will find after the symbol the page number or the reference number which he should turn up in the text itself. I further pointed out that in principle there are two fundamentally different kinds of classification symbols that can be adopted for this purpose. One principle is to use symbols listed in numerical order (like those of the Universal Decimal Classification) or in a mixed numerical and alphabetical order (as in certain library systems for classifying books) wherein the symbols represent places in a systematic or hierarchical scheme—as it were co-ordinates in a sort of map of the field of knowledge with which the searcher is concerned. The other principle is that which is nearly always adopted for making indexes to individual books, which is what concerns us here tonight. This principle, when you come to think of it, is really a very odd sort of classification: the symbols adopted are descriptive catchwords, and the order in which they are arranged in the index

Adapted from an introductory talk at a discussion meeting of the Society of Indexers on March 25th, 1965.
has nothing to do with the logical interrelationships of the concepts themselves but represents merely the sounds which people utter—the noises they make—each in his own language when they think about them. They are arranged in an arbitrary but conventional order which everyone learns in childhood, known as the alphabet. That is what we mean by alphabetical indexing.

From a practical point of view this has certain advantages and also certain disadvantages as compared with systematic classification. Sometimes, for instance, it is a disadvantage that logically contiguous items are distributed haphazardly from beginning to end of the index according to the letters of the alphabet with which their names happen to begin. I think myself that nevertheless this will continue to be the normal form of indexing within individual books, but it would be interesting to hear if anyone differs from this view. This is a question which depends on what answer is given to the first two of those I have put down in the notice of the discussion we are now going to have, namely the fundamental question, 'For what purposes do readers of scientific books need indexes in them?'—a question we certainly ought to ask ourselves if there is to be any rational basis for deciding how much time and thought and money ought to be spent on the provision of indexes in books—and, closely linked with it, the further question whether indexes in scientific books need to have any different characteristics from those in other books. As I see it myself—but here again I should be interested to hear if other indexers think otherwise—the answers to these two related questions depend on what sort of scientific books we have in mind. To explain what I mean I must make a short digression. The best definition of science I know is taken from a book by T. H. Savory, a biologist schoolmaster. 'Science,' he says, 'consists of organised knowledge in which the facts have been obtained by observation and progress has been directed by hypothesis.' The aim pursued by pure scientists is to clarify and amplify our understanding of phenomena by arranging observed and recorded facts in patterns and trying to discern recurrent relationships among them; when they think they have discerned such a theme they call it a hypothesis and test it by experiment; if the experiment is found to confirm the hypothesis (both qualitatively and quantitatively) however many times it is repeated, the relationship discovered is put forward as a scientific law. What applied scientists do is to look for ways of turning the world's accumulated knowledge of such laws to economic and social advantage in technology, medicine, agriculture, and so on. What makes it possible by these means to deepen and widen our understanding for the purpose of improving our material environment is an interplay between purposefully directed experimental research and efficiently exploitable records of already established data and knowledge. Those two things, original research and the study of existing records, are interdependent necessities for progress. They form a cycle.

Therefore, what makes it necessary for scientists to read books is not so much reading as an end in itself but the possibility of finding data, bits of knowledge, in those books which they in turn can utilise as a basis either for further hypotheses and experiments of their own if they are pure scientists or for practical ends if they are applied scientists. This is quite a different motive from the one that animates the reader of a literary work, for
instance a novel or a biography. He reads the latter as an end in itself—for enjoyment, for education, for culture—because he is interested in the theme of that particular book as a whole and its manifestation of its author's personality, whereas a scientist reading a scientific book is relatively disinterested in the arrangement of character or authorship of the book as such. He regards it rather as a kind of quarry from which he may be able to extract useful raw material for his own work of processing and integrating that material with other material in order to reach a higher stage of development.

This difference in outlook and purpose is, or ought to be, reflected in differences between the types of indexing proper to 'literary' and 'scientific' works. One can arrange varieties of scientific books, somewhat arbitrarily, in a sort of spectrum having those that are most akin to 'literary' on the left and those whose purpose is most exclusively and directly 'scientific' in the sense explained on the right:

<table>
<thead>
<tr>
<th>Science School More Progress Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>fiction text- advanced reviews books,</td>
</tr>
<tr>
<td>and books intro- particular of</td>
</tr>
<tr>
<td>popular science ducing fields, abstracts,</td>
</tr>
<tr>
<td>science students serving and</td>
</tr>
<tr>
<td>to the as frame- compendia</td>
</tr>
<tr>
<td>methods works for of</td>
</tr>
<tr>
<td>of science biblio- established</td>
</tr>
<tr>
<td>and graphical data</td>
</tr>
<tr>
<td>leading references including</td>
</tr>
<tr>
<td>them up to the citations</td>
</tr>
<tr>
<td>to its original of their</td>
</tr>
<tr>
<td>frontiers sources original</td>
</tr>
<tr>
<td>of data sources</td>
</tr>
</tbody>
</table>

The further you go to the right in this spectrum the more essential is the indexing. Indeed, on the extreme right the indexing may be as important as the text. This is because the original sources of scientific data—the 'papers' and articles currently appearing in the world's 50,000 or more scientific periodicals (including notably the proceedings of learned societies and technical institutions) which report the results of new research and describe new applications—run into millions every year, so that nobody has time even to scan, let alone to digest, more than a minute fraction of them for himself. It has to be done vicariously. (Professor J. D. Bernal has estimated that if a chemist were fluent in 30 languages and started on January 1st to read all the papers in his particular field of chemistry, keeping it up for forty hours a week at a rate of four an hour, then by December 31st he would have read not more than one-tenth of the material published during that year, from which the benefit would be nil as he would have had no time to do anything with the knowledge he gained). In order to cope with this situation a vast apparatus of abstracting and bibliographical publications has grown up for the purpose of supplying scientists and technologists with at any rate an indirect awareness of what has been published in the fields that concern them, and of making it possible for the original sources of the items that have the relevance they want to be 'retrieved' (to use the now fashionable word) for them from among all the millions upon millions stored up in the world's libraries. To assist in this process it is important that the indexes to particular books should cover not only items in those books themselves but also bibliographical references cited incidentally in them.

So the answer to our question whether indexes in scientific books need to have different characteristics from those of other books is 'yes', to an extent governed by how closely the book in question comes to the scientific end of the 'spectrum'. Indeed one might go further and argue that
it would be a good thing—though difficult
to carry out—if the systematics of the in-
ternal indexing within particular books
could be directly correlated with the ex-
ternal indexing of all books in the shape
of library cataloguing and bibliographical
controls. (If you like using bastard Greek
words, you could call these two things
‘microindexing’ and ‘macroindexing’,
both ideally forming part of one integrated
system).

All this is rather abstract, so let us now
consider a few practical examples of what
it implies. First, the arrangement of sub-
heads under a main heading in an index
entry. In a narrative literary work the
plan which is easiest for the indexer to
follow, and which also is convenient to
the user of the index, is to put the sub-
headings in the same sequence as they
occur in the text. For instance, in the in-
dex to Sir Winston Churchill's book The
Second World War you will find this ex-
ample:

War Cabinet—Sir Archibald Sinclair and, 11;
composition of, 13; meetings of, 15, 18; on
need to retain air defences, 40; prepared for
evacuation of army, 52-3; determined to fight
on, 80-157; and Italian appeasement, 190-
191; in consultation on fall of France, 175-
176, 180-1; etc., etc.

This is a satisfactory arrangement for
the purpose, because in a narrative work
the order of the page numbers will gener-
ally correspond with the time order of the
events reported and this is as helpful an
order as any for the reader to follow when
searching for whichever items having a
‘wanted relevance’ he requires. But in a
scientific work the historical sequence in
which data were established is of little or
no importance so the sub-heads are usu-
ally, and more conveniently, catchwords
or phrases denoting particularisations of
what the main heads denote, arranged in
alphabetical order under those, like:

<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell division</td>
<td>73</td>
</tr>
<tr>
<td>activation energy</td>
<td>148</td>
</tr>
<tr>
<td>clock models for</td>
<td>470 ff, 477 ff</td>
</tr>
<tr>
<td>doubling time</td>
<td>528</td>
</tr>
<tr>
<td>isochronous</td>
<td>284</td>
</tr>
<tr>
<td>kinetics of</td>
<td>442</td>
</tr>
<tr>
<td>non-clock model for</td>
<td>482</td>
</tr>
<tr>
<td>normalized rate of</td>
<td>498</td>
</tr>
<tr>
<td>partial synchrony of</td>
<td>307</td>
</tr>
<tr>
<td>phased, in marine dinoflagellates</td>
<td>307</td>
</tr>
<tr>
<td>sequence finite number, steps</td>
<td>478</td>
</tr>
</tbody>
</table>

(As explained in a note at the begin-
ing of the subject index to the book from
which this is taken, the letters ff indicate
that the stated subject is mentioned again
on one or both of the next two following
pages’. Personally I see no objection to
this device, which saves quite a lot of
printing.)

Another question that arises in making
the subject index to a book of this kind
is how far you should go, and what prin-
ciple you can follow, in deciding whether
to group many sub-headings under one
main heading; because if you overdo this
you are apt to get what has been called
‘concealed classification’ of subject mat-
ter introduced haphazardly into the middle
of what purports to be a straight alpha-
betical index. That is a bad thing because,
to give a very simple example, somebody
wanting to turn up ‘Electron microscope’
has no means of knowing whether he
should look under ‘E’ or for a sub-item
under ‘Microscope’. It is rather difficult
to arrive at a principle for deciding this
and I should be glad to hear what others
think of it. My own feeling is that the
principle should be not to introduce sub-
headings unless they can be read joined
on to the main headings (the entry
phrases) without involving a reversal of
the ordinary sequence of words in the Eng-
lish language. For instance, in the exam-
ple I gave just now I think it better to put
‘Electron microscope’ under ‘e’ than to
enter it as 'Microscope, electron'. Certainly I prefer 'Mass spectrograph' to 'Spectrograph, mass' as this is a fundamentally different instrument serving a different purpose from an ordinary optical spectrograph. Sometimes the principle I advocate involves using adjectives or adverbs, or present participles of verbs, as entry terms in an index; I see no objection to this if it makes for consistency.

The disadvantages, of course, of generally preferring to make each item a main head of its own rather than enter it as a sub-head under another item in cases of doubt is that if you do that it will inevitably lead to related concepts being spread at random all over the index according to the letters with which their names happen to begin. But you cannot have it both ways. Either you must classify concepts systematically or you must classify their names alphabetically. If you forget which you are doing and sometimes introduce 'concealed classifications' into an alphabetical index, the user of the index is never sure where he is to look and the index is not an efficient one. The proper place to perform systematic classification is not in the index appended to the book but in the list of contents printed at its beginning—or better still in a more detailed list of contents printed at the beginning of each chapter, wherein the subdivisions of that chapter are listed in the order they occur in the text, giving the page numbers where they begin.

These sequential lists of sub-divisions of chapters can be extremely valuable. In effect they constitute a third sort of indexing principle: neither an impersonal 'map of knowledge' like a systematic classification such as the U.D.C., nor, so to speak, an alphabetical gazetteer listing the place names in the map and telling you the map squares in which those places are marked. Instead, they are an itinerary of the route that the author of the book, who presumably is an expert on its subject field, has chosen for guiding the reader through it. If each chapter, or section of the book, is preceded by a list of the sub-section headings or cross headings in the text in this 'itinerary' form, the alphabet indexing at the end can be economised, for if you put a note calling the reader's attention to the fact that the sub-divisions of chapters are listed at the beginnings of the chapters there is no need to include the headings of those sub-divisions over again in the alphabetical index appended to the book as a whole. In that way you can save space and effort, and therefore money, which if desired may be used for indexing additional items which otherwise would have been crowded out. The note can suggest that where a single index entry (such as a geographical name) is followed by a long string of page numbers the reader should consult the list of contents at the beginning of the book in order to differentiate them according to the contents of the successive chapters. This applies to the 'author index' as well as to the 'subject index'. The provision of a separate index to the names of authors whose works are cited in bibliographical references included in the book is another feature typical of scientific books. Commonly each chapter of the text is followed by a bibliography of the papers and articles cited in it. In some subject fields there is need for yet other indexes, as in the case of Chemical Abstracts mentioned later.

In the example here reproduced as Fig. 1, taken from a collective work in which each chapter is by a different author, some typographical distinctions are introduced which may appear complicated but which are useful.

127
The third of the questions suggested for discussion, in the programme for this meeting, is whether it is possible to formulate any general objective criterion for deciding upon the density of indexing, the average number of index entries per page of text. It would be very valuable to authors and still more to publishers, when planning a new publication, if this question could be answered 'Yes', but that, it seems to me, is very difficult. We can assume, I think, that the needed density of indexing increases as you proceed from left to right in the 'spectrum' aforementioned, for even if in popular science books or school textbooks the indexing may be relatively unimportant, in reference books it may be as important as the text. It might be interesting, and perhaps worth while, for this Society to sponsor a survey of successful scientific books for the purpose of establishing what density is typical in each category.

The most densely indexed publication I know is the American Chemical Abstracts, which provides an example of the art of alphabetical indexing brought to its highest perfection in a field where particularly difficult questions of nomenclature continually arise. In that publication there are 650 words of indexing for every 1,000 words of text. The abstracts appear fortnightly and they cover some tens of thousands of articles contained each year in some thousands of journals all over the world, as well as about 12,000 chemical patents every year. Each fortnightly issue of the abstracts has indexes to its contents and on the conclusion of the annual volume these are cumulated in a large separate volume containing nothing but the indexes; also every ten years they are cumulated to make a decennial index. Covering the year 1952, for instance, the
subject index alone totalled 1,586 pages, the author index another 625 pages, the empirical formula index 347 pages, and the patent index 16 pages.

Having now, I hope, been sufficiently provocative in raising moot points of principle in the indexing of scientific books I want to use the remaining few minutes, before inviting discussion, by setting up targets for your criticisms in the field of indexing procedures. In particular it would be helpful to myself to hear the reactions of other indexers if I described the methods I have gradually been developing and improving since I first took up the indexing of scientific books as a sideline to other work.

At first I followed perhaps the most usual method, that of writing each proposed index entry on a separate slip of paper or card and filing these in alphabetical order as I went along. I found however that even with a great many projecting tabs to mark where each sub-section of the alphabet begins much time is wasted fumbling among the slips already inserted in order to make additional entries on them, so I hit on the ‘shingled sheets’ method illustrated in my second article in The Indexer (Vol II, No. 1, Spring 1960). Using this method the entries are written in alphabetical order but leaving gaps between them for later additions, on foolscap sheets stapled together in batches of ten so that the bottom edge of each sheet projects below the previous sheet ¼ inch, making it very much quicker to find the correct alphabetical positions than when each entry is written on a separate slip. The secret for not getting some of the sheets overcrowded (such as those for the unexpectedly prolific letter C) is threefold: allow plenty of space and do not be afraid of wasting some paper; mark the bottom right corners of the successive overlapping sheets with beginnings of words like Bro-Car-Cer-Cey-Cha- . . . copied from the entry words that occur at successive regular intervals in going through an existing large index or dictionary so as to ensure that roughly the same number of entries may be expected between, for instance, ‘Bro’ and ‘Car’ as between ‘Car’ and ‘Cer’; and, to provide for the eventuality that however much foresight you exercise there may be cases where you will need to add another entry between two already written on successive lines, use at first only the right hand half of each sheet and leave the left half free for interpolations.

Except for the practical point that constant lifting of the corners of the sheets tends to make them excessively dog-eared this procedure works well. It does not, however, remove certain disadvantages which are inherent in any system where everything has to be done by one person:

1. It is not realistic to expect that one person copying out many hundreds of technical expressions and personal names of mixed national origins in alphabetical order, followed by page numbers, will never make a mistake.

2. Ability to decide what to include in a subject index requires appropriate technical education and informed judgment, but indexing is not well paid and an indexer having these qualifications is not using his time to economic advantage if most of it is absorbed in the purely clerical part of the task and in checking.

3. Not every indexer's handwriting is disciplined enough to be acceptable by printers, but retyping from a hand written draft involves additional work and risk of error.
mixture of C₆ hydrocarbons, he was apparently the first one to contrive the formula of an allene. In 1875, Henry proposed the trivial name ‘allene’ for propadienes in connexion with his alleged synthesis of tetramethylpropadiene which again seems to have been only a mixture. Curiously enough, Burton and Pechmann synthesized glutinic acid in 1887, and assigned to it the structure, discarding the allenic formula on the grounds that the new compound reacted in an analogous way to acetylenedicarboxylic acid. They even went so far as to claim that this demonstrated the impossibility of a single

$$\text{HOOC} = \text{C} = \text{CH}_2\text{COOH}$$

(2)

$$\text{HOOCCH} = \text{C} = \text{CHCOOH}$$

(3)
carbon atom forming two double bonds. Yet these investigators were the first to have a pure cumulene in their hands! This was revealed 67 years later when Jones and coworkers showed that ‘glutinic acid’ had the allenic structure. Burton and Pechmann’s pessimistic view was proved to be unfounded only one year later when, in 1888, both Russian and American investigators succeeded in synthesizing allene itself and several alkyl-substituted propadienes, whose structures were proved conclusively.

There seems to have been the general belief that compounds with more than two consecutive cumulated double bonds would be unstable, and it was only in 1921 that Brand obtained the next higher homolog of allene, namely a butatriene. Even this appears to have been more of an accidental discovery than the result of a systematic search in this field. It took another 17 years to extend the series of cumulenes still further. This was done by Kuhn and coworkers who synthesized tetraphenylhexapentaene, which turned out to be almost as stable as its forerunner. However, an accumulation of more than five double bonds proved to be difficult. Such compounds can so far only be obtained in solution.

### B. Naturally Occurring Cumulenes

Although the history of cumulenes begins as early as 1864, the right of priority has to be given to nature. In 1906, Semmler had already suggested that carlina oxide, which he had obtained from the essential oil of Carlina acaulis, should have an allenic formula. However, this was later shown to be the isomeric acetylene. In 1924, Staudinger and Ruzicka proposed formula 4 for pyrethrolone, a component of

---

Fig. 2. A page of text marked up by indexer as guide to assistant, showing subject index entries on left and authors’ names index entries on right.

Reproduced by courtesy of John Wiley & Sons Ltd.
Therefore I am now trying out and improving a more radically different procedure, wherein the work and the payment for it are shared between two people—a qualified indexer and an assistant who needs merely to be able to write and type accurately—each of whom automatically checks the previous operation performed by the other. This new method, although simple in practice, would take too much space to describe in detail here but its principle is as follows (Fig. 2): after the indexer has underlined in the text the names to be indexed and made a mark in the margin against each name the assistant writes abbreviations of these (normally the first two letters and the last letter of each name) followed by the page number on very small gummed labels, which, after the indexer has checked them, she sorts into alphabetical order and sticks into place following the full surnames and initials which she has copied from the lists of references at the end of the chapters, as illustrated in Fig. 3. (The printer is instructed to print the underlined numbers, in italics, after the other numbers.) Self-sticking ‘tacky’ labels are more convenient to use (but more expensive) than labels that have to be moistened, and it is easy to type on them without stripping them off the backing sheets on which they are supplied, so they can be kept in page-number order until the indexer has checked them at a glance.

The entries to be made in the subject index are underlined in the text and marked in the other margin, or written in that margin, to be treated in a somewhat similar manner, using larger labels.

Baker, A.K. 53 38 49
Berglander-Wertheim, F.L. 116 73 75 80 92
Bosworth, M.N.W. 116 72 75 83

Fig. 3. Self-sticking labels stuck against index entry headings.

'A hilarious encounter with Mr. Noel Coward is described, in which that maestro is quoted as saying, "It is better that people like us should be friends rather than enemies because we really have so much in common—powers of observation, wit, industriousness and professionalism". Those four qualities are apparent in a diary which, with so much legitimate name-dropping, should have been given an index.—From a review of The years between. Diaries, 1939-44, by Cecil Beaton (Weidenfeld and Nicolson) in The Times, August 12, 1965.