

CHAPTER 3

GENERAL ORGANISATION OF PROGRAMME

There were three main variables to be investigated, namely the system, the indexer, and the indexing time. With four systems, three indexers and five indexing times, the number of permutations is sixty. The documents to be indexed were divided into groups of one hundred (from here on referred to as a "document group"), and therefore 6,000 documents were indexed before the same indexing conditions are repeated.

The procedure was that the first indexer indexed documents 1 - 100 by system A, allowing himself an average time of 16 minutes for each document. Immediately after indexing a document by system A, he allocated the appropriate headings or classification numbers for systems B, C, and D, but this was done without strict time control. Documents 101 - 200 were indexed by system B with the 16 minute allowance for each document, followed by the postings for systems A, C, and D. Items 201 - 300 and 301 - 400 were similarly indexed by systems C and D.

This procedure was repeated for documents 401 - 500, 501 - 600, 601 - 700, and 701 - 800, except that for these document groups the average indexing time was limited to 12 minutes. For documents 801 - 1200, the time was limited to 8 minutes; for 1201 - 1600 it was 4 minutes, and finally for documents 1601 - 2000 the indexing time was limited to 2 minutes per document. Meanwhile the second indexer carried out a similar procedure with documents 2001 - 4000, and the third indexer was doing the same with documents 4001 - 6000. The indexing of documents 6001 - 12,000 repeated the conditions found in the indexing

of 1 - 6000, and the whole stage was repeated for the third time in documents 12,001 - 18,000.

The preparation of these document groups was done in advance; in the earlier stages of the work the procedure was for the documents to be collected together and a list prepared. As described earlier, within each group of 100 documents there were the further sub-groups of having half the documents dealing with high speed aerodynamics while the remainder ranged over the whole subject field. There was the division between periodical articles and research reports, and also between United States publications and those from other countries. To simplify the somewhat complex procedure that this involved, in the later stages of the work we first typed cards giving details of a large number of documents and were able, more effectively, to make our selection of document groups by sorting these cards.

Before being passed to the indexer, each document had a master indexing card inserted. This card, as shown in Fig. 1, contained the project document number, and information concerning the indexer, the main indexing system and the time allowance.

On receiving the group of 100 documents, the indexer had the responsibility of indexing them to the best of his or her ability within the conditions imposed. Presuming that the indexer had received a group of 100 documents to be indexed by the Universal Decimal Classification as the main system with an indexing time of 16 minutes, this meant that the indexer would have to complete the indexing by U.D.C. of this group of documents within a total time of 1600 minutes, and he was given the flexibility to vary the time spent on any individual item. In addition he had to index by the other three systems, and this

question of the subsidiary indexing presented some initial problems. Basically it was not intended that there should be the same strict time controls with the subsidiary indexing as was the case with the main indexing, but it was desirable that the indexing should be approximately the same standard by all systems within any single group.

The difficulty arose in trying to separate the time spent in assessing the subject content of a document as against the time spent in deciding on the most satisfactory notation or headings. The percentage of time spent in these two actions would not remain constant throughout the varying time allowances, and as the time allowance became shorter the ratio of reading time to indexing time decreased sharply. As a result of tests made by the indexers it was decided that the following times should be allowed for the subsidiary systems.

<u>Main system</u>	<u>Subsidiary system</u>
16 mins.	4 mins.
12	3
8	2
4	1½
2	1

Later experience bore out the correctness of these figures, and it was felt at the end of the whole of the indexing work that the assessment had been about right.

The amount of time required for assessing the subject content of a report as against the time taken for assigning suitable headings naturally varied from one report to another. In general, the times listed above made for consistent indexing and the subject matter of a

document could usually be assimilated quickly enough to leave adequate time for assigning class numbers. In fact, with the majority of reports the indexers found themselves assigning numbers very shortly after first examining the document, and further reading merely served to confirm the original classification, except where later in the report mention was made of a subsidiary subject, (e.g. a special method of testing), which demanded separate indexing.

There were cases where all the time available was needed for reading the document, because of the abstruse nature of the subject, but in general this made little difference to the comparative times spent on the subsidiary systems and nothing should be lost on this account.

The greatest difficulty in this respect was caused by the report whose subject matter was easily assimilated, but which required a large number of class numbers or subject headings because of the many different subjects treated. At the longer times even these could usually be catered for adequately in the times permitted for subsidiary systems. In a few exceptional cases at the shorter times, the indexers used some discretion about using rather longer times for subsidiary systems. It would obviously have been wrong to devote almost the full time of, say, 4 minutes for the main system to assigning class numbers for a document which needed little actual reading, and to restrict the subsidiary systems to the stated $1\frac{1}{2}$ minutes. These cases were rare, and it can hardly be doubted that the course taken in these instances was the correct one.

A point which had to be watched carefully, particularly in the early stages, whilst the indexes were building up quickly, was the time spent

in making decisions about the choice of, say, a new U.D.C. number, or a new alphabetical subject heading. Where such a decision was not just the appropriateness of the heading to the document in question but a question of general policy that required a discussion and a joint decision by all the staff, the time was not included in the indexing time, as this would have affected the quality of the indexing of that document adversely.

The original proposal envisaged the completion of the work within two years from the commencement date, at the same time permitting the indexers a generous allowance of leave to compensate for the arduous nature of the work. This estimate was based on a working week of 38 hours, but it rapidly became clear that this estimate was a very poor one, for one of the greatest difficulties in the early stage was that experienced by the indexers in becoming accustomed to the very stringent conditions which continuous indexing with time controlled by a stop watch brought about. The mental strain of working under these conditions was severe, but perseverance eventually brought about an attitude of mind which made the work bearable. It was felt that two years was as much as the average human being could reasonably be expected to index under such controlled conditions, even though the number of hours actually logged as indexing time was a comparatively low proportion of what would be regarded as normal working hours.

Taken as an average in the later stages of the work, it was the experience of the indexers that five hours recorded indexing time represented a good day's work. We endeavoured to find how this might compare with experience in work of a comparable nature, but could only learn of one test that had been done privately in an industrial organisation.

This showed that in an engineering design office an average of 40% of a $7\frac{1}{2}$ hour working day was spent in actual productive work. From this it would appear that the indexers did reasonably well to reach 66%, particularly when it is remembered that some of their unrecorded time would be taken up with discussions directly relevant to their work. The practical implication of this recorded time as against time spent in the office is that our average time of 16 minutes would under normal working conditions be 25 minutes, and the other times would be increased proportionately. The result is that the difference in output for the various timings would be 18, 25, 36, 72 and 144 items indexed for a working day of $7\frac{1}{2}$ hours.

At first it seemed that the quality of the indexing was inconsistent, but greater standardisation through discussion, greater familiarity with the various schemes, and a general working up to peak efficiency led to a more satisfactory situation, and by the time that the first 6,000 documents had been dealt with, the indexers had established a tempo which was maintained to the end of the whole of the indexing. We feel confident that consistency reached a standard that will be more than high enough to ensure that the testing will be done on material of acceptable quality.

It was desirable to have one person on the staff with overall responsibility for each system used in the indexing, and from the start Mr. J. Sharp was in charge of the U. D. C. and Mr. J. Hadlow of the alphabetical subject catalogue. The Director was concerned with the Uniterm index until the appointment of Miss Warburton, who then resumed responsibility for this system.

With the facet schedules, Mr. Opatowski spent a great deal of time in the early stages carrying out technical revisions. (This aspect is

discussed later in Chapter 4 and more particularly in Ref. 7). After his departure Mr. Sharp and Mr. Hadlow prepared an authoritative list of headings to be used for the chain index, but when this had been completed, we "froze" the schedules and made no alteration of any kind. As is discussed later, we had a temporary device which enabled us to introduce subjects which were not included in the original schedules, this being done by giving to all subjects the notation of "Zz". These were comparatively few in number, and as a result there was no need for any further control of this system.

With the U.D.C., for which Mr. Sharp was responsible, if either of the other two indexers wished to generate a new number which had not previously been used, they would use the number and make out an appropriate indexing slip

e.g. Planetary gears 621.833.6

These would be passed daily to Mr. Sharp, who would be responsible for checking that the use of this number did not contravene any previously arrived at policy decisions. He would then edit the slip to conform to the accepted layout, and make any additional entries that were necessary,

e.g. Planetary gears. Power transmission. Mechanical engineering
621.833.6

Gears, planetary. Power transmission. Mechanical engineering
621.833.6

These slips would next be passed to the typist, who would make three copies, one to be filed in each indexer's alphabetic index of the U.D.C.

A similar procedure worked with the alphabetical system, although in this case Mr. Hadlow had the added responsibility of checking that a new heading was not, in fact, a synonym for a heading that was already in the authority list.

With uniterm the procedure was slightly different, in that the indexers entered on their master index cards (see Fig. 1) the code numbers of the uniterms (see Chapter 5). When it was necessary to use a new term, they would not have a number to use, so they wrote the term instead. At the completion of the indexing of each group of documents, the master index cards were passed to Miss Warburton, who looked through the cards to find any new terms that had been used. Having checked that the term was not synonymous with a term already in use, she would allocate to it the appropriate code number and enter this number on the master index card.

In spite of our best endeavours before starting the programme, not only of those closely concerned with the project but of many other individuals who were kind enough to comment on the original proposals, we came up against many difficulties that had not been envisaged. The question of the recorded time as compared to the working time has already been discussed, but probably our main difficulty came in getting started with the indexing. This is illustrated by the fact that whereas we started indexing in April 1958, it was not until the end of January 1959, nearly ten months later, that Mr. Sharp and Mr. Hadlow had completed their first 2,000 documents. By this time most difficulties had been solved and they only took about thirteen months to do a further 4,000 documents each.

It proved, for instance, quite impossible to plunge into indexing by four systems at once, that is to say we could not carry out the indexing by the main and subsidiary systems from the start. The compromise adopted was that the first 100 documents of each indexer were done only by U.D.C. The next hundred were done with alphabetical as the main system, with U.D.C. as the only subsidiary system. Then we did a

hundred documents by facet with U.D.C. and alphabetical as subsidiary systems and finally we did the fourth hundred with uniterm as a main system and the other three systems as subsidiaries.

At this stage it was clear that with the facet classification we were getting into a complete mess and a technical revision of the schedules was necessary, together with some procedural alterations. While this work was being undertaken, indexing proceeded, albeit somewhat slowly, with the other systems but no facet indexing was done of the next 400 documents. During this stage it also became obvious that we had to tighten our control of the form of headings in the alphabetical subject catalogue, and it was at this stage we made up our first set of rules, as discussed later in Chapter 4. As a result of these rules, it was necessary to alter many of the headings already in use. The U.D.C. schedules which we were using suffered from inadequate alphabetical indexes, and as a result there was a continual stream of minor difficulties. In general, however, U.D.C. did not present the major problems which we encountered with the facet and the alphabetical systems, although this, no doubt, was partly due to the several years experience which Mr. Sharp had in starting and maintaining a U.D.C. catalogue in his previous post. The Uniterm system at this stage was perfectly straightforward and it can be said that, as far as the indexers were concerned, they encountered only very minor problems with this system throughout the project.

With the revision of the facet schedules completed, we resumed indexing by this system, but again ran into difficulties, this time with the freedom that had been given to change terms from substantives into adjectives, which caused resultant difficulties in the chain index (as

discussed in Chapter 4). This was a problem that should, I feel, have been foreseen by us but it had never been raised in any discussions or in any papers dealing with facet classification. The alterations which we had to make more or less invalidated all the work that had been done so far, and to be able to use this earlier indexing, we were involved in a considerable amount of revision. However, as previously indicated, from this time onwards, the facet schedules remained unaltered and we had no further major problems in indexing by this method.

As indexing continued, the original rules for the alphabetical subject indexing appeared to be too restrictive and we had to make some slight modifications. This occurred at the end of the first stage of indexing but did not seriously affect the earlier indexing that had been done. Gradually the number of problems associated with the U.D.C. also declined and by the commencement of the second stage of indexing (i.e. documents 6001 - 12,000), we were in the clear with all systems.

It is difficult to know exactly how to assess our experience in setting up these systems, because few people appear to have attempted to review objectively their own experience. In 1959 there was published the results of an investigation in Germany into the costs incurred in indexing by U.D.C. (Ref. 8), and in this case it was found that six months work was required before the system was reasonably well established. From this it would appear that our time of approximately 9 months to establish three systems (it being accepted that Uniterm created no problems) was not unreasonable, even though at the time I was considerably worried at the delay in reaching a stable situation. However, there is no doubt but that it would have been most unwise to attempt to proceed more rapidly with the indexing without first solving the problems that arose.