& Lyle Ltd., and Mr. R.C.Wright, Librarian, Royal Aircraft Establishment. In addition Mr. W.K.Lowry, Manager, Technical Information Libraries, Bell Telephone Laboratories, agreed to serve as a corresponding member, and, following his appointment as Aslib Research Director, Mr. C.W.Hanson became a committee member in January 1959. This committee met at Cranfield on five occasions.

CHAPTER 2

GENERAL OBJECTIVES OF THE INVESTIGATION

"The vastly increased sums of money which have been devoted to research have brought about two major changes which affect bibliographical work. The volume of papers issued is now vastly greater than in the pre-war years, and also the number of individuals or organisations working in the same field of knowledge has increased considerably. There is no longer the situation where a scientist could keep abreast of developments over a fairly broad field of knowledge and be reasonably certain of knowing, either personally or by reputation, most of the other scientists concerned in similar work. Nowadays it is generally agreed that even the specialist is, in many fields of knowledge, unable to give more than a cursory glance at a large proportion of the papers which are of potential value to him. A result of this is that, whenever a new problem is approached, a scientist is far less likely to have a complete knowledge of the work already done, and therefore must make greater use of bibliographical aids.

"With the increasing complexity of research work, experimental test equipment has become very expensive, and its operating costs are usually high. It is therefore more than every necessary to avoid duplication of work previously done. These facts have, to a certain extent, been recognised by government and industry alike, and the result has been a great increase in the number of special libraries and information bureaux serving research organisations. Many new techniques in librarianship have been evolved to suit the changed functions of these organisations, and prominence has always been given to various methods of retrieving information. In particular the factors discussed in the opening paragraph have, during the past years, resulted in an impetus in this work and many non-conventional systems have been suggested in an effort to overcome the problems created by the vastly greater flow of research papers and the increased need to locate information.

"Apart from new systems, there have been rapid developments in the physical aspects of the work. In the same way as a librarian is no longer faced with the plain alternative of using a dictionary catalogue or classified catalogue, but in addition has to consider the merits of Uniterm or Zato-coding, so instead of the practically automatic choice of 5" x 3" cards, there are the possibilities to be considered of hand or machine sorted punched cards, and various mechanical or electronic devices to assist in retrieval.

"In all the controversies that have raged during the past fifty years on the basic points of a book catalogue or card catalogue, with an alphabetical subject arrangement or a classified arrangement, it is interesting to note that no attempt has been made to carry out any

controlled tests that would enable one to make statements based on fact rather than voice theoretical opinions. The changed attitude that must prevail was admirably stated in the editorial of "American Documentation" for April 1955, which, in part, reads,

> "Cautious and searching evaluation of all experimental results is essential in rating the efficiency of documentation systems. May the age old controversies that arose from the conventional concepts of classification not be reborn in the mechanized searching systems of the future. There is hope for the avoidance of such error if we will but regard documentation systems as useful devices, the benefits of which must be determined, not by polemics, but by the intelligent measurement of such benefits in relation to needs and costs. The machines of the future can make us free but only if we are willing to subject them, and ourselves, to the most rigid intellectual discipline".

"The complication in attempting to evaluate the comparative efficiency of any two retrieval systems is due to the number of various factors which have to be considered. These can be summarized as follows :-

1. The documents which are to be indexed
2. The system of indexing
3. The indexer's subject knowledge of the documents being indexed
4. The indexer's familiarity with the indexing system
5. The size of the index

6. The type of question which is to be put to the index
7. The equipment to be used in recording or retrieving data
8. The overall efficiency, which is made up of :-

   a. The time cost in preparing the index
   b. The time cost in locating required information
   c. The cost of equipment used
   d. The probability of producing the required answer
   e. The absence of irrelevant answers ('noise')
   f. The number of searches made. "

The foregoing is taken from the opening pages of the first draft of the proposal which was prepared in 1956 and in general it contained the basic considerations on which the project was planned. This section of the report will deal with the actions which were taken to enable us to investigate the various factors listed above.

The document collection

There are obviously a number of different requirements when indexing collections of documents in different subjects; it may even be the case with different groups of documents in the same subject field and also possibly for the same group of documents when the index is being compiled for different sets of users. The extent to which these differences are important is a matter for conjecture, but it would be rash to suggest that the results which we obtain will necessarily be valid for every other subject field. We think that it will be possible to investigate this matter without it being necessary to do more than a small fraction of the work involved in the present project.

The choice of aeronautics as the subject for the indexing was mainly

determined by the availability of the documents, but it did have a number of other advantages. As it is a subject which embraces a number of disciplines, we were able to investigate the indexing systems over a broad range of subjects including, amongst others, metallurgy, plastics, electrical engineering, production engineering, mechanical engineering, fuels, mathematics and most branches of physics. Approximately half the documents used were in the specialised field of high speed aerodynamics, and this concentration on one detailed subject area, combined with the general coverage of a wide range of subjects, should show the varying capabilities from the viewpoint of special or universal systems.

Reports and papers vary in their presentation and most practising indexers would agree that this is a factor which can definitely influence the time that is taken to index a document and possibly influences the quality of the indexing. Little general consideration has been given to the form of titles, yet it is obvious that there is a considerable difference in this respect between British and United States practice. With the advent of mechanical indexing of the Key-Word-in-Context type, the importance of the terminology of the titles is more likely to be recognised, but for the human indexer also the titles should be a precise and succinct statement which will immediately tell the indexer the basic subject of the paper. From the abstract or summary, the indexer can reasonably expect that the author will have indicated most if not all the points which have to be brought out in the indexing.

It appeared that the project presented an opportunity to investigate any possible differences that might exist in the indexibility of different types of documents. Half the documents indexed were articles in

scientific and technical journals while the other half were research reports. There was another equal division between papers written in the United States and in other English-speaking countries. Papers were not used if they had been written more than ten years previously, and in fact nearly half of the documents were published during the two years of the indexing.

The breakdown between

a. High speed aerodynamics and general subjects
b. Articles in journals and research reports
c. U.S. papers and papers from other countries

were maintained within each "document group" (see page 20) so that the effect of the various conditions could be evaluated at all stages of the work. A representative group of 100 documents is listed in Appendix A.

The system of indexing

Four systems were chosen for comparison, these being

a. The Universal Decimal Classification
b. Alphabetical subject catalogue
c. A faceted classification scheme
d. The Uniterm system of co-ordinate indexing

The basis for this selection was that the schemes differed as fundamentally as is possible and represented the principal types of retrieval systems which have any significance in the present state of the art.

It should be emphasised that the mechanics of the systems, in the

sense of the equipment and methods which are used for recording and retrieval, were regarded as being outside the scope of the project. The basic principles of the systems are not altered in any way by their method of application and as the purpose of the project is to compare the systems as such, the possible use of machines or special apparatus has not been taken into account. Machines can alter the economics of the operation of the respective methods, and can determine the choice of one system as against another for practical purposes, but such a choice can only be made after the relative efficiencies of the systems as such have been assessed. It is felt that it is a serious error to consider whether a given piece of machinery can work as an information retrieval system merely because it exists as a piece of machinery, and particularly, as is so often the case, when it has been designed for quite a different purpose. The correct approach, rather, is to find what system in principle lends itself best to information retrieval under a given set of circumstances and then look for ways of making it as economic a proposition as possible, by the use of machines, or by any other method of application. To have considered, as some have suggested, the comparison of the selected systems with a 'machine retrieval system' would have been to be guilty of the grossest example of 'cross classification'.

Reasons for the choice of the schemes

The Universal Decimal Classification was chosen as the most widely used system of the 'enumerative' type. Though it is to some extent 'synthetic', it is fundamentally enumerative in that the vast majority of its places are fixed within its respective classes, and the synthetic principle applies generally only to those aspect of subjects which might

be regarded as 'auxiliary' to the basic numbers. A more purely enumerative scheme might have been chosen, but the end result would probably have been so little different that the exclusion of U.D.C. would certainly not have been justified.

U.D.C. is not only enumerative, but it is typical of the classical 'tree of knowledge' classification. Moreover, it uses the decimal principle for supposedly unlimited expansion, a feature which has certain advantages, particularly from the point of view of mechanising alphabetical indexing, but which tends to produce numbers of unacceptable length at the level of indexing used for the project. Unfortunately the principle of decimal notation for the purpose of showing hierarchy has been abused to some extent, and this may or may not affect it adversely. However, the principle remains sufficiently well embodied in the scheme to justify its testing opposed to the Facet scheme, which is not based on the ' tree of knowledge' principle, is not enumerative, and does not use a hierarchical notation.

The alphabetical subject catalogue was an obvious choice as one of the systems to be tested, as it is the one system which lists subjects in a known order, regardless of grouping by classification, and without the necessity of any intermediate step in searching such as is necessary with a classified catalogue. Its comparison with the classified catalogues (U.D.C. and Facet) is expected to show whether anything is lost or gained by the absence of the grouping provided by classification both for specific reference and for generic survey. It is sometimes argued that a subject can be classified though it cannot be adequately expressed in words, but others such as Metcalfe (Ref. 6), argue strongly against this. A comparison between the alphabetical and the classified catalogues

should demonstrate the validity or otherwise of such arguments. The idea that classification permits more specific description of a given subject than does the use of alphabetical subject headings is also prevalent in some quarters. That the degree of specificity is unlimited with a system designed for such exhaustive specification as U.D.C. cannot be denied. The crux of the problem of information retrieval is, however, the need to ensure that the description of a subject by the searcher, in terms used by the system, coincides exactly, or more or less exactly with that assigned by the indexer. If such coincidence is difficult or impossible, then the efficiency of the system will be impaired accordingly. It is possible that the shortcomings of classification systems in this respect reach down to that level of coarseness which alphabetical systems possess, and the testing of the systems should show whether or not this is the case.

The alphabetical subject catalogue has in common with the Uniterm system, the use of terms used to describe subjects in ordinary language. What it does not normally do is to provide headings comprising every possible permutation and combination of the words used. The Uniterm system permits such permutation and combination by the very nature of its mechanics.

Finally, without the analysis of principles, the inclusion of the alphabetical subject catalogue would have been justified solely on the grounds of its almost exclusive use in America, as against the popularity of the classified catalogue in British libraries.

It was felt that any attempt to compare systems of information retrieval would serve no purpose if it did not include a scheme based on the most modern principles. A scheme was therefore prepared by

B.C. Vickery and J.L. Farradane specially for the project. Some modifications to this scheme were made before it was brought into operation, but these were mainly of a technical nature (that is, technical from the point of view of the subject matter, not from the point of view of classification principles), but it should be emphasised that care was taken to preserve the structure of the scheme as the compilers were the best judge of what was required from this point of view.

The purpose of including a faceted scheme was to test the effectiveness of subject analysis and synthesis by facet principles as this technique is obviously a very powerful tool. Implicit in this type of scheme is the principle of 'preferred order' and the use of a chain index for the purpose of entering the classified file at a suitable point, and the collection of 'distributed relatives'. The authors of the scheme considered very seriously the question of preferred order, for the subject matter presented a very difficult problem in this respect, mainly on account of the large number of categories involved. Our experience in indexing by this method showed that their choice of preferred order had, in the main, been the correct one. It may be that modifications would have been made for a more satisfactory arrangement, but this is by no means certain, and such modifications could, in fact, have made the scheme worse rather than better. It is extremely difficult to assess the value of one order against another, but the indexers' experience, in general, supported the authors' choice.

What is more serious is the possibility that no single order can cater for all requirements. As preferred order is a fundamental part of a faceted scheme of this kind, the results of the tests should be

significant in that the principle should be shown to be valid or not. Should the testing of the schemes show the faceted scheme to be wanting, it will not necessarily be concluded that facet analysis as such serves no useful purpose. It could be that the categories properly derived by facet analysis have their value jeopardised by being tied to the principle of preferred order. This feature will be investigated by later tests where we shall use the categories without reference to preferred order. This will be done by free co-ordination of terms, one from each category at a time, by the use of punched cards or a computer. This procedure should significantly advance our knowledge in this matter, as it is extremely unlikely that any preferred order fundamentally different from the one adopted would be any more satisfactory.

The interesting feature about the comparison of the faceted scheme with the other classification scheme (U.D.C.) is that the former is built from the outset on scientific principles of classification, whilst the latter is largely empirically derived.

An investigation of this kind would not have been complete without the inclusion of a system based on co-ordinate principles. The main difference which this type of scheme provides is the facility to combine all or any of the relevant terms in the system by way of permutation or combination. This facility is available only to a more or less limited extent in the other schemes and it was imperative that the efficiency of this feature should be tested. It is difficult, in the present state of the art, to assess what is the best level of sophistication in a scheme of this kind, for on the one hand there is the danger of 'false drops' because of ambiguity in relationships, and on the other the risk of the inclusion of irrelevant terms because of difficulty in definition, and the overlapping

of terms. The former has been catered for in some schemes by the use of 'role indicators' and the latter by the use of the 'thesaurus' approach. There seems to be no conclusive evidence of the efficiency of these devices and it was decided, therefore, that no attempt would be made to incorporate such possible refinements, beyond the provision of references from synonyms to the terms selected for use. It may be that 'literary warrant' is a better basis for the selection of terms than any artificial principles and it was therefore left to the indexers to co-operate in generating suitable terms during the course of indexing. The result, therefore, is that we compiled a Uniterm index as originally proposed by Dr. M. Taube.

These decisions created some surprise, and it may be desirable to explain our reasons. We had originally considered using a more complex form of co-ordinate system than simple uniterms. Examples of what might have been done are shown by the following working schemes :-

1. The thesaurus approach, as used by Whelan (Ref. 9)
2. Zato-coding (Ref. 10)
3. The categorizing of terms, as used by Wildhack in the National Bureau of Standards index on instrumentation (Ref. 11)
4. Concept co-ordination (e.g. Wadington) (Ref. 12)
5. The use of role indicators (e.g. Farradane) (Ref. 13)

The first two methods both have a separate dictionary of unit terms, so that the indexer shall know which particular descriptor (or descriptors) to use and the third system appears to be in the main a list of unit terms sorted alphabetically into the categories. In all cases the intention is to help the searcher to devise a programme which is more likely to match

the indexing.

The true value of these different approaches can, we hope, be ascertained in the project if we take our basic uniterms and regroup them in the appropriate manner. This will be a simple problem clerically, but it would obviously not have been so easy to do the same thing in reverse if we had, for instance, used a thesaurus system. Retesting the regrouped terms will indicate the gains and losses of these different forms of co-ordinate indexing.

The concept co-ordination system devised by Wadington is an illustration of what we wished to avoid doing. In this case the organisation (Titanium Division of National Lead Co.) started their search for an improved indexing system by constructing a classification based on facet principles, but found difficulties in its practical application. Switching next to Uniterms, they found it inadequate due to the many incorrect co-ordinations that could be made. Pre-coordinating many of the terms overcomes this difficulty, and it gradually became evident that the unit concepts thus devised were much the same as the terms used to construct the facet classification. So the structure of the classification, without its coding, was used as a retrieval system in a straight co-ordinate manner. If we had developed a co-ordinate system in this manner, we should probably have finished with something similar to the facet system we are using which is, of course, quite capable of being used as a co-ordinate system. It is, in fact, our intention to test it in this way in addition to the basic method of using a classified catalogue and chain index.

As for role indicators, an analysis of our test results will show the number of occasions when false drops would have been avoided by their

use, this being another occasion where the reverse process would not have been so useful.

In every case it appears that having in the first place used Uniterms, we shall be able to investigate many other forms of co-ordinate indexing without having to repeat the indexing process.

The indexer's subject knowledge and familiarity with the systems

Whereas one school of thought insists that technical indexing can be done only by a person with technical qualifications, there are others who argue that this is not necessarily the case, and that anyhow it would be wasteful to employ technical persons on such work. However, no results are known of tests designed to compare the indexing ability of different types of persons. Potential indexers for the project might be described as being within the following broad groups: (A) technical knowledge of the subject plus indexing experience; (B) technical knowledge of the subject but no indexing experience; (C) indexing experience in the subject field; (D) indexing experience in another subject field; (E) theoretical knowledge of indexing; (F) neither knowledge of the subject nor of indexing.

In deciding the types of persons to recruit, it was considered that, of the very few persons with the qualifications outlined in (A), it would be difficult to persuade any of them to join the project. The significance in certain circumstances of using an individual such as outlined in (F) was appreciated, but it was decided not to use this category in the project. The choice finally was made to recruit three indexers who would be representative of groups (B), (C), (E).

As mentioned earlier Mr. Opatowski, the indexer originally appointed as having technical knowledge, left in the early stages of the indexing, and was replaced by Miss Warburton, who did not have technical knowledge but was representative of group (D).

With some, if not all, of the systems, it was to be expected that the indexers would become more efficient as they had more experience in the use of the systems. In order to ascertain the extent of this improvement, the whole programme was divided into three main sections and the indexing conditions were repeated within each section of the whole programme, so that we shall have comparable results for all stages of the indexing.

Size of the index

The size of the intended index is a matter of considerable importance, and whereas one hopes that it should be possible to assess a system without indexing a million documents, it is impractical to expect valid results from very small scale tests.

In a project of this kind it was necessary to index a sufficient number of documents such as would ensure that retrieval is not made too obvious and simple. On the other hand it was undesirable to waste money by continuing the work beyond the stage where useful and valid information could be gained.

The decision as to the number of documents we should index was based on the fact that we had a possible 60 permutations of the three main variables (see page 20) and that 100 documents was a convenient number to use for each of these sixty groups. This would make up a sub-programme of 6,000 documents, and repeating this three times gave us a total programme of 18,000 documents. We are certain that to have concluded the work of indexing at the completion of the second stage of the programme (i.e. at 12,000 documents) would have adversely affected the results and we feel reasonably confident that to have continued the indexing beyond the ultimate stage would not have been justified.

The fact that approximately half the documents fall within the narrow field of high speed aerodynamics (see page 7) means that we shall be searching indexes with a high concentration of entries in this particular area. It is reasonable to suggest that in a normal index of aeronautical documents, such a concentration of references in one specialised field would only be found in a collection which altogether contained possibly 200,000 documents and we therefore feel justified in assuming that results in this area would remain valid for a document collection several times larger than that used in the project.

The overall efficiency

The assumption on which the investigation is based is that the only valid way to measure the efficiency of any system of indexing is by basing measurements on economic costs and in this there are always three matters to be considered, these being :-

a. The cost of indexing
b. The cost of preparing the physical index
c. The cost of searching

The cost of indexing is made up of salary paid to the indexer (with any overhead costs involved) and the average time spent in indexing a document. We attempted to cover the first point by appointing people of different qualifications who would normally expect to receive different salaries. The time spent in indexing is, in any single organisation, a policy decision and will depend on a number of factors. To assess the result of such a decision, the indexers were given fixed times for indexing each group of 100 documents, these times being an average, for each document in the

group, of sixteen minutes, twelve minutes, eight minutes, four minutes and two minutes.

It was deliberately decided that our investigation would not include any comparison of different methods which can be used for recording indexing decisions, of different equipment which can be used for retrieving information or of methods of printing out retrieved data. The reasons for neglecting these matters is that they have no direct bearing on the efficiency of indexing. Unless the indexer has made an entry under a certain heading, it is useless to search for that heading, whether one is using a card catalogue, peek-a-boo equipment, punched cards or a computer. It is true that different equipment might make a system more attractive economically, but it would not affect the end result and its economic advantages will depend on the circumstances. To take an obvious example, the comparison of lists of numbers (as originally proposed for Uniterm) is normally a slow and cumbrous method, and one which might be done more quickly by a computer, albeit the costs for a single search would be higher. However, if one regularly had up to 100 searches to be made at a time, then the cost per search with the computer might well be economic.

These matters are very important, but, particularly three years ago when the project started, the situation regarding equipment for data retrieval was in a state of flux and new developments might well alter the whole picture. Meanwhile a great deal of experience has been gained particularly in the United States, and if no other report on the subject is issued in the meantime, we intend at the completion of the project to do an analysis of the costs of all types of equipment used in information retrieval.