PROPOSALS FOR AN INVESTIGATION INTO THE EFFICIENCY OF VARIOUS RETRIEVAL SYSTEMS.

The Organisation of the Research Programme.

The grant would be administered by the Council of Aslib. A contract would be arranged with the College of Aeronautics at Cranfield, Bedfordshire to carry out the research programme. The Director of the Research would be the Librarian of the College, Mr Cyril Cleverdon, and the Council of Aslib would appoint a small committee which would be responsible for assuring that the programme was proceeding satisfactorily and which would make a report to the Council of Aslib every three months.

The Director of Research would be responsible for the staff, for the organisation and detailed planning of the programme, for the supply of the material to be indexed, for the coordination of all tests and the publication of the results of the programme.

The Staff

In addition to the Director, a staff of three indexers will be required. It is intended that they shall be of varying qualifications. The first person required is the type of person one might expect to recruit into a special Library, that is a person with the necessary professional qualifications but lacking experience in the practice of indexing. The second indexer should be more senior, with, if possible, experience of indexing in aeronautics. The main qualification of the third indexer is that he should be a trained engineer (an aerodynamicist for preference), and experience in indexing would be a secondary consideration.

Such staff would be appointed for the duration of the contract and would be expected to continue in the posts for this period. It is realized that there may be some difficulties in recruiting these staff, but it is felt that locating the research project at the College of Aeronautics at Cranfield will help in this respect. The two librarian indexers would have some activities which would bring them into contact with the College library; as for the technical indexer, the Senate of the College have agreed that such an individual could hold the position of Research Assistant and could, in his spare time, join in the

research work of the Department of Aerodynamics. At least one clerical
assistant would be required; if it should be necessary, extra part time
labour could be recruited without difficulty.

The Material to be Indexed

The decision has been taken to carry out the work in the subject field of
aeronautics. It is intended that approximately one half of the documents
should be in the specialised field of high speed aerodynamics. Although
the documents have not yet been sorted out, it is believed that the majority
of the 20,000 items to be indexed could be provided by the reports and
aeronautical periodicals in the library of the College. If further documents
were required, they could be supplied from the government aeronautical
libraries. In general, all the items indexed would have been published in
the last ten years. The actual selection of the items to be indexed would be
the responsibility of the Director of Research.

The Indexing

The programme calls for the indexing of 20,000 documents by five different
systems in a time of two years, with a staff of three indexers. The two
important matters to control will be the time taken to index a document and
the identity of the indexer of each document.

When indexing a document, the first part of the task is to ascertain the
subject content of the document, and only after this has been done does the
indexer decide the subject headings, classification numbers etc., which the
item requires. It is therefore obvious that a person indexing a document by
five different systems would take less time to allocate the appropriate symbols
once the subject content of the item had been determined. It was first con-
sidered that it might be possible to separate these two aspects of indexing, and
time each part separately. Some practical tests were done, and it was found
that it was very difficult for the indexer to say honestly where one process left
off and the other began. Inevitably one tends to decide on possible subject
headings or classification numbers while still considering the subject content,
and it was therefore decided that this method would be unsatisfactory. To have
had each document indexed by separate indexers for each system would increase

the cost five-fold, and for this purpose was likewise rejected. The most
practical method appeared to be to accept the limitation that a time check
in each particular system should only be kept on 20% of the total documents
indexed. This, however, is considered to be a sample that will be
sufficiently large to bring valid results from the tests. The intention is to
divide the items to be indexed into groups of 500., and every group will be
divided into five sections of 100 each. One indexer will be responsible for
each group of 500, and each section of 100 items will be indexed first by
one of the five different systems. Time recording will only be kept for this
first indexing. To illustrate this, presume that the five systems are
A, B, C, D, and E.

Items 1-100   will be indexed under time controlled conditions by system A,
                 after which, without time control, the indexer will allocate
                 the appropriate headings or symbols for systems B, C, D, E.

Items 101-200 will be indexed under time controlled conditions by system B,
                 after which, without time control the indexer will allocate
                 the appropriate headings and symbols for systems B, C, D, E,
                 and A.

Similarly, items 201-300, 301-400 and 401-500 will be indexed first by
systems C, D and E respectively under time-controlled conditions, with
indexing in the other systems following in sequence.

Meanwhile the second indexer would be doing items 501-1000 in a similar
manner, and the third indexer would be doing items 1001-1500.

This pattern would be repeated throughout the programme, and would cover
the first part of identifying the indexer. The second important matter is
the time taken in indexing. It will be agreed that, other things being equal,
the longer the time taken in indexing a document, the more thoroughly will
it be done. This is not however, the same as saying that it is economically
efficient to take longer in indexing a document, and no one has attempted
categorically to state a reasonable time allowance for indexing. Obviously,
this will vary with circumstances, for no individual organisation, preparing
its own catalogue, could justify the two hours per document taken by the
Nationaalluchtvaartlaboratorium in their Catalogue of Aerodynamic Measure-
ments. Opinions on this point have been obtained from a number of aero-

nautical organisations and it appears that 20 minutes is the limit for indexing the average report. Most organisations take considerably less than this, and an experienced indexer argued that two minutes was adequate for the purpose for which his catalogue was designed, and he further argued that this is probably the most economically efficient indexing time. To work at such a rate implies great familiarity both with the indexing system and the material to be indexed, but it is a viewpoint that demands investigation. It is therefore proposed that the permissible indexing time should be varied in the batch sequences. In addition there will be a time allowance for indexing by the other four systems, and after a preliminary familiarisation period this can be reasonably consistent throughout the tests. The intention is that the following times should be maintained.

Items										
1-1500	20 minutes	for first system		+	15 minutes	for remainder			
1501-3000	10	"	"	"	"	+	10	"	"	"
3001-4500	5	"	"	"	"	+	10	"	"	"
4501-6000	2	"	"	"	"	+	8	"	"	"
6001-7500	20	"	"	"	"	+	8	"	"	"
7501-9000	10	"	"	"	"	+	8	"	"	"
9001-10,500	5	"	"	"	"	+	8	"	"	"
10,501-12,000	2	"	"	"	"	+	8	"	"	"
12001-18,000	Repeat of 6001-12,000									
18001-21,000	Repeat of 7501-10,500									

It should be emphasised that these are intended to be maximum allowances, and would be averaged out over the batches of 100 documents.

Normally it is intended that two of the indexers should be indexing at a time, but for the first groups of documents, it is suggested that all three indexers should be indexing. The timetable, based on a working week of 38 hours, can be worked out as follows.

Items 1-1500	35 minutes a document 65 documents a week 8 weeks for each indexer	Weeks 1-8
Items 1501-3000	20 minutes a document 110 documents a week 5 weeks for each indexer	Weeks 9-16
Items 3001-4500	15 minutes a document 150 documents a week 4 weeks for each indexer.	Weeks 17-22
Items 4501-6000	10 minutes a document 220 documents a week $2\frac{1}{2}$ weeks for each indexer.	Weeks 23-26
Items 6001-7500	28 minutes a document 75 documents a week 7 weeks for each indexer	Weeks 27-36
Items 7501-9000	18 minutes a document 125 documents a week 4 weeks for each indexer	Weeks 37-42
Items 9001-10,500	13 minutes a document 170 documents a week 3 weeks for each indexer	Weeks 43-47
Items 10,501-12,000	10 minutes a document 220 documents a week $2\frac{1}{2}$ weeks for each indexer.	Weeks 48-51
Items 12,000-18,000	Repeat of 6001-12,000 Total of 25 weeks	Weeks 52-76
Items 18,001-21,000	Repeat of 7501-10,500 Total of 11 weeks	Weeks 77-87

This allows a potential 17 weeks leave for each indexer over a period of two years, and each indexer would have to spend only 63 weeks of the whole period in actual indexing. As the time allowances for each section have been generously estimated, it is clear that there is no reason to doubt that the indexing can be completed with the period of two years.

The time checks could be kept by each indexer having a clock which would be started and stopped at the beginning and end of each period of time indexing. Every single document would not have to have the exact specified time spent on it, but the specified time would be an average over a batch. For instance, where the time per document was 10 minutes, then an indexer would be expected to index 20 documents in three hours and twenty minutes.

The Clerical Work

This aspect will also have to be time checked so that the necessary calculations can be made to cover the various systems tested. It is, however, clearly unnecessary to make these checks so detailed or comprehensive as for the indexing. With the conventional card catalogues, the average time for typing each card, and for inserting it in its correct place in a catalogue can be readily found, and will remain reasonably constant throughout the whole programme. With the non-conventional systems which require different posting techniques, there is now a considerable amount of recorded experience on various ways of doing this. The opportunity might be taken to carry out some controlled tests of the various possible methods, but clearly it would only be fair to debit the system with the time factor for the most efficient method.

In practice, the simplest possible methods would be used in order to eliminate as much clerical work as possible. For instance, as three systems would require conventional catalogue cards, it would be simpler to make duplimat masters and run off sufficient copies to meet the requirements of the indexers. Without suggesting that this aspect has been worked out in such detail as the indexing, it is possible to gauge the time of some of the various processes. To consider the work involved with the three conventional systems, this would involve the typing of 21,000 master cards. Allowing 3 minutes for each card (author, title and identification) this equals 1050 hours work over two years or about 11 hours a week. The time for reproduction of the cards is estimated to average about five hours a week. Allowing 6 cards for each document in each of three systems this equals 378,000 cards to be filed. Our tests have shown that 6 cards a minute can be averaged, so here again the time is approximately 1,050 hours. This shows that the clerical processes connected with three systems would average about 27 hours of a working week of 38 hours, so there is a reasonable possibility of one clerical assistant being able to do the whole job on their own. However, it must be remembered that after the first 8 weeks, there would always be one of the indexers having a rest from indexing and this person would be able to help in some of the clerical work.

The Systems

As originally proposed, the programme called for the testing of five different systems. This remains the intention, but there remains a certain flexibility concerning the details.

The Universal Decimal Classification is in a state of constant change, and there is no definite edition which can be used. The subject content of the documents to be indexed will concern mainly 533.6 Aerodynamics and 629.13 Aeronautical Engineering. Revision of these two sections has been discussed for some time, and satisfactory schedules have been evolved. These have not yet passed through all the necessary stages to receive international approval, but it is considered to be reasonable to use these schedules for the programme. Standard detailed schedules in English are available for the remaining sections likely to be used.

The alphabetical subject catalogue raises the question of whether the headings should be decided while the indexing is to be done. There are unlikely to be any available lists which could be used without amendment or addition, but it is considered desirable to select the most suitable list of headings and add all relevant headings. It should, in fact, be the case that the subject heading list would be as complete and up-to-date as the U. D. C. Schedules.

Again, with the schedules to be prepared by the Classification Research Group, it is hoped that these schedules would be satisfactory for the duration of the test. A group of four people, two experts in faceted classification and two aerodynamicists, have agreed to prepare these schedules, and with one of the classifiers devoting a large proportion of his time to the project, they are confident in being able to prepare schedules within three months.

The fourth system proposed is Uniterm. Here it would be desirable to have a prepared list of uniterms ready before the indexing commences, although it is certain that many other terms will have to be added as the work proceeds.

The fifth system was originally proposed as being one suitable for use with punched cards. For some years, leading theorists in documentation retrieval, both in England and America, have been saying that it was unreasonable to condemn mechanical retrieval methods on the basis of their failure to produce satisfactory results, unless one designed systems which they could handle. This problem was being considered and discussed in 1954, at the time when the research programme was first proposed, and at that time it appeared possible that a satisfactory system could be devised for use with machines. There has been some progress since that date, both with regard to methods and to the equipment, but it appears unlikely that either the system or the machine would be ready in time for the research programme. Since it was never intended that machine methods should be tested unless a satisfactory solution had been reached on the other point, it is now considered advisable to forego the testing of any mechanical system.

This leaves the alternatives of only testing in the four previously described systems, or of adding another system. It is tentatively suggested that Zato-coding might be tested, and comments on this point would be welcome.

Testing

It is important to bear in mind that the completed indexes will not be perfect. Indeed, the whole object of the programme is to produce indexes which vary in their efficiency. This variation may be due to the ability of the indexers, to the time which they have been allowed for indexing, to the system used, or to the documents being indexed. Among further factors influencing true efficiency of the catalogue will be the type of question that is put to it, the individuals who try to find the answers to the questions (their basic familiarity with the system, their growth familiarity with the system, their technical knowledge, etc), the number of questions asked of an index over a given period, and the number of questions which can be put to an index simultaneously.

It is therefore obvious that the testing will not be simple, but will require equally close control as with the indexing. In that the problem is at least two years away, it has not seemed so necessary to specify in detail the individuals or the methods to be used in testing. More important, however, is the fact that once the indexes have been compiled, there will be available a tool which can be used in any way required, and which, within obvious practical limitations, can be used by any individual who feels that he may be able to extract further valid information by doing more tests.

While, as already said, the testing has not been worked out in detail, there are quite clearly two parts to this work. The first is the selection of questions; the second is the interpretation of the questions and searching of the catalogues. Although there are two aspects of the second part, yet they would normally be done by the same person. It is, of course, true that the same person might ask the questions and try to find the answer, but this would not necessarily be the case.

Obviously test questions will have to be formed so as to cover every section of the indexing programme. Each section consists of 100 documents, so if one question is to be asked of each section, then 210 questions will be required. This is too low a figure to give valid results, and it is considered likely that we shall have to ask at least ten questions of each section before our results are valid. This figure is based on experience in testing some existing indexes, but it may be too low although at present one can make no better guess. The correct figure will have to be discovered by making thorough tests in a few sections. If, after making ten searches in one section, the results are still fluctuating, then it will be necessary to continue to make more searches in the section until an acceptable degree of stability has been reached. By repeating the process in other sections, one should be able to arrive at a standard figure for the whole test programme.

Presuming that it was found that a 10% test coverage was required, then it would be necessary to have ten people each compiling 210 questions.

Some sample questions would be devised by the Director varying from very specific down to general, and would be coded 1-5. Those making up the questions would be asked to spread them equally over the range of specificity and to code them accordingly by using those devised by the Director as samples. All these questions would be based on documents which were known to be in the collection indexed, but the question compilers would also include 40 other questions within the general subject field, but which were not definitely based on actual documents. The compilers would be either librarians or persons with technical knowledge of the subject field. A sufficient number of individuals have already offered their voluntary assistance to prepare their questions, but clearly there is no reason why individuals in the United States might not also prepare lists of questions.

Many possible ways of using the test questions can be devised. The indexers themselves can carry out the searches, or the compilers of the questions can put them to the indexes. Others, whether librarians or scientists, with no knowledge of the indexes, could make tests, and in all cases the basic set of questions would be used to provide comparison between the suitability of the systems for different types of people. Time control would have to be maintained in the retrieval processes, but it does not seem necessary to impose any series of time limits.

ANALYSIS AND INTERPRETATION OF TEST RESULTS

Presuming that all the records concerning the indexing and retrieval processes have been faithfully made, it will be possible to analyse the results of the programme in such ways as will reveal a number of factors. Basically there should be the answers to two main points. Firstly, that a certain system will give the highest retrieval efficiency if all economic factors are ignored, and secondly that if a stated retrieval efficiency is considered satisfactory, then economically one particular system is better than the others. For instance, all systems (A, B, C, D, and E) might attain 70% retrieval efficiency, but possibly only system B might give this efficiency with the shortest indexing time of 2 minutes and the lowest paid indexer. Yet it is conceivable that system D might increase its retrieval

efficiency to 90% under the most expensive conditions, whereas system B might never get beyond 80%.

Within these two basic answers, there will be a large number of variations. It has to be remembered that each index is, in itself two different kinds of index. Half of the documents will range over a fairly broad subject field, while half will be in the narrow field of high speed aerodynamics. It is possible that different systems will prove most efficient for these two different aspects.

While the final tests will be the most important, it is desirable that some intermediate tests should be carried out. These would be done after the completion of 6,000 documents and 12,000 documents and in the final analysis might help to show the effect of the size of the collection on the comparative efficiency of the systems.

One of the more important and difficult problems in interpreting the test results will be to measure the importance of relevant information and the time cost of irrelevant information produced by the indexes. The important difference between the tests proposed in this programme and those attempted in the ASTIA-Uniterm programme is that in this case there will be a single known and definite answer to each question Failure to find the required specific reference will indicate a failure which would not be mitigated by finding other relevant information. This might seem a harsh decision, but appears reasonable when one remembers that each question is designed to test a particular indexer and time allowance, in addition to a particular system. It would result in intolerable complications and invalid conclusions to take account of other relevant answers.

There will, however, be many occasions when the index fails to reveal the correct answer as a single item, but will, for example, leave the searcher with five references, one of which may give him the correct answer. To carry the work through to its final stage, the relevant document must be put into the hands of the intended user of the information. The time involved would obviously vary with different organisations, depending on size of collection, location of stacks, etc. Some standard time will, however, have to be decided so that it can be built into the results.

In the case mentioned at the start of this paragraph, a charge will
have to be made for the estimated time spent in finding the relevant
documents. For instance, if the index produces five references, but
fails to distinguish the single correct reference, then it would be
necessary to consult each document if two documents are irrelevant,
two relevant but not the correct answer, and the fifth document is
the correct one, then 2x minutes would be added to the search time
(x being the agreed standard time for locating a document.)

It will be appreciated that there are many possible grounds for argu-
ment when deciding how to assess the results produced by the
proposed programme. So many possible variations are to be con-
trolled and tested that there is a danger that the proponents of each
system will be able to claim success in certain circumstances.
This is not necessarily a bad thing, as it has always been realised
that there is unlikely to be an overall "winner". What is important
is that individual organisations should be able to judge the results
in the light of their own requirements, which vary considerably
between different organisations, and then be able to decide which
system is most likely to suit their requirements.