1961 b

APPLICATION FOR GRANT TO

THE NATIONAL SCIENCE FOUNDATION

WASHINGTON

An investigation into the performance characteristics of descriptor languages

Director

Cyril W. Cleverdon

Cranfield

November 1961

Introduction

In 1956 the National Science Foundation made a grant to Aslib for the first stage of an investigation into the comparative efficiency of four indexing systems. This stage of the work continued until March, 1959, and a further grant was made to cover the test programme. It is from the results of this work that the present proposals have evolved.

PART I

1. Whereas the present project has provided a considerable amount of information concerning indexing techniques, it had definite limitations, in particular the environment in which we worked and also the subject field of the collection that was indexed. The environment was intended to be that which might be found in an engineering research organisation, and comparison with the indexing which was done by many such organisations showed that we did simulate the average level. However, it is obvious that there can be wide variations, and these require to be investigated to ascertain whether the test results remain valid. Equally so, it remains a matter of doubt whether the results which were obtained for a collection of documents on aeronautical engineering are applicable in other subject fields.

2. The testing techniques which were developed for the present project have been adapted so that they may be used on any form of index and be capable of giving, at a relatively low cost in time and money, performance figures which are more precise and reliable than has previously been the case. Such a test method is therefore in itself worth developing and refining. This can best be done by carrying out more tests on existing systems; in the process of doing such tests it will be possible to extend the validity of our results into other subject fields and environments.

In addition, and possibly most important, would be the additional basic data which is obtained from such tests and which would be of great value in the main programme.

3. The technique of testing other existing indexes has already been tried at the Nuclear Power division of English Electric Co. Ltd. (Ref. 1), and is now being used for tests at Western Reserve University. The method is to have 200 questions prepared, preferably by engineers or scientists familiar with the work of the organisation. These questions, which would be intended to be representative of questions normally put to the index, are all based on documents known to be in the collection. It is also necessary for certain information to be available concerning the indexing decisions for these documents and desirable that we should also know the time taken for indexing and who did the indexing. Searches based on these questions are then made by project staff and by both information staff and technical staff of the organisation concerned.

Full records are kept of the results of the searches and in each case the searcher will be endeavouring to locate the source document. From the number of source documents retrieved it is possible to give the efficiency in recall. From the other documents retrieved in the course of the searches a statistically valid sample is taken and passed to the technical staff who originated the appropriate questions. These people are asked to place each document in one of four groups:-

0. Of no interest
1. Of minor relevance
2. As useful as the source document
3. More useful than the source document

From these assessments it is possible then to estimate the relevance figure, that is, the percentage ratio of relevant documents retrieved (see Part II, Para. 6).

From an analysis of the results of any such test, it is possible to make recommendations regarding methods of improving the efficiency and economic operation of the system tested.

4. Basically this is a technique of testing which could be tried here and now by any organisation interested in finding the efficiency of their system, but from a single test or a number of unrelated tests, it would be more difficult to make any firm conclusions. In addition, the analysis of the test results could probably be done most effectively by those who have been engaged on the present project.

5. A further four or five tests would probably be sufficient for the validity of this method of testing to be accepted and allow the most economic and efficient techniques to be evolved. For establishing the validity of the results of the present work in other subject fields, a larger number of such tests would be desirable. A third and most important reason for doing such tests will be to check, in a real life situation, tentative conclusions reached as a result of our work in the main programme and also to obtain supplementary data of a nature that cannot be obtained in "laboratory" work. In the present project, for instance, we reached a tentative conclusion concerning chain indexing which was quite contrary to generally accepted views. The English Electric Co. test done on a facet catalogue, completely confirmed our conclusions, and has convinced people who might otherwise have been sceptical. The test on the Western Reserve University, although uncompleted, has provided information concerning the characteristics of the system which will influence the approach made in the main programme.

In short, it has been shown that such tests can provide information both of a confirmatory and an exploratory nature, and can be used in this way as an integral part of the main programme.

6. It is therefore proposed that ten further tests should be made for the purpose of establishing the testing technique and extending the subject range of the results of the present project. Any other tests which were made would be planned so as to supplement data obtained in the main test programme. A series of ten tests might cover the following indexes:

	General Subject field	Category of organisation	Indexing System	Size of collection	Degree of detail*
1	Aluminium	Industrial	Alphabetical	60,000	3
2	Aeronautical	Research	Co-ordinate	12,000	1
3	Rubber	Industrial	U.D.C.	60,000	3
4	Electrical engineering	Research	Alphabetical	80,000	3
5	Physics (general)	Abstract journals	Alphabetical	100,000	2
6	Production engineering	Research	Alphabetical	100,000	3
7	Shipping	Industrial	Co-ordinate	60,000	2
8	Botany	Research	Alphabetical	50,000	3
9	Fuels(Patents)	Industrial	Co-ordinate	100,000	1
10	Medicine	Research	U.D.C.	100,000	2

* 1. Very detailed indexing
2. More detailed than normal
3. Normal indexing (as used in Cranfield project)

It is estimated that for each test project staff are likely to be engaged for ten days in the preparation and actual testing, with a further twenty-five days for analysis of results. Allowing a working year of 230 days, approximately $1\frac{1}{2}$ man-years would be required for the completion of ten tests.

PART II

1. The programme which is now nearing completion was limited in scope, and the main activity in the present proposals envisages a more fundamental study over a wider field of descriptor languages.* It may first be desirable to consider the conclusions emerging from the present work.

2. Originally the programme proposed the indexing of 18,000 documents by four descriptor languages, namely, Universal Decimal Classification, alphabetical subject indexing, facet and uniterm. Certain controls were built into the indexing covering such matters as indexing time, qualifications of the indexer and type of document indexed. Essentially the intention was to produce indexes which were reasonably similar to those found in normal working situations and the figures given in Table 10 of the report on the first stage (Ref. 2), show that this was in fact done.

3. The major test programme was considerably larger than had been attempted in other investigations of this nature, a total of 1,200 searches being made in all four systems. This testing was done by using questions which were based on documents in the collection and was heavily biased towards the ability of the indexes to retrieve a known relevant document. The searches were considered as complete when the source document had been retrieved or, alternatively, were abandoned when, the source document not having been retrieved, it was not possible to devise further search programmes which could meet certain specified conditions. Further tests

---

* "Descriptor language" has been adopted as a more precise term for the indexing method, e.g. U.D.C. classification, Uniterm, alphabetical subject heading, etc., so that the term "system" may refer to the whole; that is to the descriptor language plus the physical methods which are used (e.g. U.D.C. with card catalogue, Uniterm with Peek-a-boo, etc.).

have been made using different methods in an attempt to assess the relevance or non-relevance of other documents retrieved in the searches. The final report on the testing is now being prepared, but this section will, without producing supporting arguments, mention some of the conclusions which affect the proposals for future work.

4. First, it appears that conventional standards of indexing have an efficiency of around 80% in regard to recall of relevant documents. If the human error which occurs either in the indexing or in the searching were completely eliminated, this figure would rise to around 90%. This human error is not concerned with such matters as failure to appreciate the content of the documents or the significance of the question, but consists of mistakes which might reasonably have been avoided. An analysis of 4,000 documents indexed by many other people (see Chapter 7 of Reference 2), as well as tests of the kind mentioned in Part I of this proposal, show that this level of error is not unusual.

5. Most significantly, in the analysis of the test results we found that the descriptor language was responsible for only 5% of the failures, or, to put it another way, with only one document in 100 was there some factor in the descriptor language which mitigated against the successful retrieval of the document.

6. As stated earlier, the main test programme was concerned with retrieval of relevant documents and the preceding discussion of results does not take into account the question of the relevancy of the other documents which were retrieved in the course of the searches. The further tests to cover this matter have been difficult, mainly because of the inconclusiveness of attempting to decide and obtain agreement on varying degrees of relevance, but these tests have shown a definite trend which has influenced the views on future work. To discuss this it is

necessary to define the meaning given to two terms, namely "recall ratio" and "relevance ratio". "Recall ratio" equals $100 \frac{R}{C}$ where C equals the total number of documents in the collection which have an agreed standard of relevance to a given question, while R equals the number of those relevant documents retrieved in a single search. On the other hand, "relevance ratio" equals $100 \frac{R}{L}$ where L equals the total number of documents retrieved in a single search. As an illustration, presume that in a given collection of documents, ten are known to have an agreed satisfactory standard of relevance. In a single search, six of these documents are retrieved, plus another twelve documents which were irrelevant. In this situation recall ratio equals $100 \times \frac{6}{10} = 60\%$ while the relevance ratio would be $100 \times \frac{6}{18} = 33\%$.

7. In communication engineering terms, recall ratio is "message", while relevance ratio refers to "noise". In the same way as the endeavour is to obtain the maximum of message with the minimum of noise, so in a retrieval system, the endeavour must be to obtain the greatest number of relevant documents with the minimum number of irrelevant documents. In the present work, an analysis of some 2,000 of the documents retrieved by the four systems in 100 complete searches has shown quite clearly that these two requirements are, in the various descriptor languages which we used, incompatible. A typical situation was that recall was 80%, with relevancy at 13%; relevancy could be raised to 40% but only by reducing recall to 50%.

8. 100% recall can, of course, be obtained if every document in the collection is searched for each question (not, it should be noted, the index entries for every document, but the actual documents). This will, depending on circumstances to be discussed later, probably involve a low relevancy ratio, since in essence every document in the collection is

retrieved in every search. While it is obvious that no descriptor language, however sophisticated it might be, or however detailed the indexing, could possibly improve on 100% recall ratio, it might reasonably expect to improve the relevancy ratio. The basic aim of indexing is therefore shown to be to make this improvement in relevance ratio without lowering too significantly the recall ratio.

9. To return to the analogy of communication engineering, a descriptor language can be considered as the equivalent of a filter. Sufficient is known of their performance so that a filter can be designed to meet any given requirement, but there are practical limitations which cannot be evaded.

10. In his recent book "On Retrieval System Theory", B. C. Vickery lists 17 kinds of descriptor languages as follows:-

	Means of control	Field of use
1.	No control.	Some amateur alphabetical indexes.
2.	Rigid control - fixed vocabulary of descriptors.	Some mechanized systems with limited coding capacity.
3.	Confounding of variant word forms.	Professional alphabetical indexes, including Uniterm, and most other systems.
4.	Confounding of true synonyms.	Ditto.
5.	Confounding of near synonyms.	Some subject heading lists, some classifications, and systems based on thesauri.
6.	Generic descriptors.	Many mechanized systems.
7.	Specific and generic descriptors linked hierarchically.	Classifications, thesauri, some subject heading lists, some mechanized systems.
8.	Multiple generic links for each specific descriptor.	Some classifications, subject heading lists, and thesauri, a few mechanized systems.

9.	Categories of descriptor, forming facets.	Faceted classifications, some mechanized systems.
10.	Semantic factors to represent subject terms.	To some extent in faceted classification, the W.R.U. system, mechanized patent office systems.
11.	Correlation of descriptors.	Many alphabetical indexes, some classified catalogues, all mechanized systems.
12.	Weighted descriptors.	Some experimental computer systems.
13.	Interlocking sets of descriptors.	Alphabetical indexes, classified catalogues, computer systems.
14.	Regulated sequence of descriptors.	Alphabetical indexes, faceted classifications, fixed-field punched cards, some computer systems.
15.	Interfixing descriptors.	Mechanized patent office systems.
16.	Role indicators.	Some faceted classifications, some mechanized systems.
17.	Relational terms.	Alphabetical indexes, some faceted classifications, some mechanized systems.

There are many working examples of the different types of descriptor languages, but the results of the present investigation tend to show that any variation in different examples of the same type of descriptor language will only be marginal.

11. The main proposal is that we should investigate the efficiency of each type of descriptor language as listed above, the purpose being to produce a series of performance curves for each method. The necessity for such an investigation is that at present it is not possible to make a realistic assessment of which type of descriptor language can be most efficiently used to meet a given requirement, either as regards efficiency or economy. Such requirements will vary in different organisations; there will be those cases

where the demand is for the best possible recall ratio, as for instance in a patent search (the "fail-safe" system postulated by Mr. S. Newman (Ref. 3)), and a satisfactory system might have the performance shown in Fig. 1.

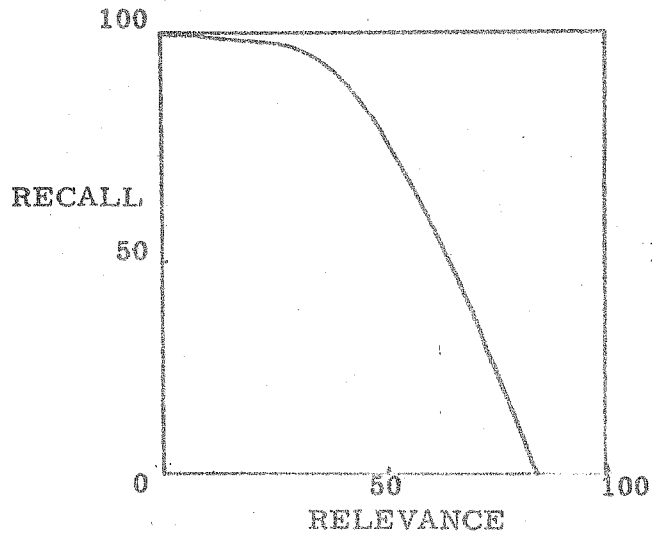


FIG. 1.

There are other situations where relevance ratio is of greater importance, and a more suitable performance curve would be as in Fig. 2.

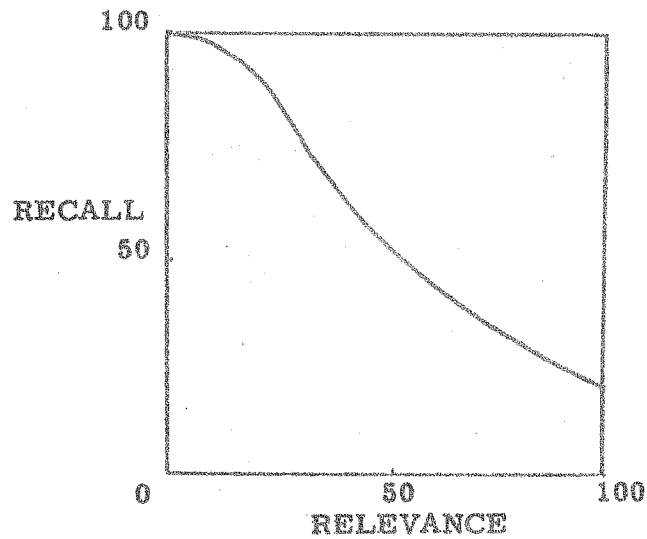


FIG. 2.

12. As has been shown, any system can be made to give 100% recall ratio (albeit with a minimum relevance ratio) by ignoring the indexing and retrieving every document. The various descriptor languages as set out by B. C. Vickery introduce alternative types of controls which lay emphasis in different ways on two opposing points. On the one hand there is the use of classifications and cross references which tend to ensure the possibility of obtaining the maximum recall of relevant documents. On the other hand operators and roll indicators in such schemes as proposed by Farradane and Western Reserve University are, by emphasising so strongly the relationship between terms, more likely to improve the relevance ratio. With some forms of descriptor languages (and more particularly, with some conventional methods of using certain descriptor languages), there may be an artificial restriction which will limit the improvement that can be obtained in the relevance ratio; however, with such systems as well as with those having tight controls, it is always possible to relax or broaden the search programme and thereby improve recall ratio at the cost of relevance ratio. It should therefore be possible with each descriptor language to plot a series of points which will enable performance curves of the type shown in Figs. 1 and 2 to be obtained.

13. Such performance curves have little meaning unless the environment in which the system operates is also defined. The main factors here are specificity of request and required degree of relevance. Commonly one finds questions being defined as "specific" or "general", but it is difficult to give any agreed assessment to these terms; a question that might be considered specific in a general reference library would be a general question for a library specialising in the particular subject area of the question. Reductio ad absurdum, it would be possible to obtain the goal of 100% recall ratio with 100% relevance ratio if a question were so general in relation to the collection, that every document in the collection was relevant.

14. It is therefore intended that this matter should be controlled by giving specificity a definite measure which will be obtained by relating the size of the total collection to the average number of relevant documents which can be retrieved in a search. This would assume that the subject content of the collection was restricted to a single discipline, and the following table is suggested as giving a measure of specificity based on a collection of 100,000 documents.

Number of Relevant Documents	Specificity Figure
0 - 4	1
5 - 10	2
11 - 20	3
21 - 40	4
41 - 100	5
101 - 500	6
501 - 2000	7
2001 - 10000	8
10001 - 20000	9
More than 20000	10

Some suggestions have been proposed for defining specificity by relating it to the number of concepts contained in a question, and we have considered this method. While it may be reasonable to do this in comparing different questions in relation to each other in a given situation, it appears to be a doubtful use when attempting to obtain such precise values as in this future proposed work.

15. It is possible that the performance curves of descriptor languages may also be altered by the degree of relevance which is demanded. That is to say, if it is specified that only documents of a high degree of relevance are required, then the resultant performance curve might be different from

that produced when the requirement was that, in addition, documents of peripheral interest should also be retrieved.

Automatic Indexing

16. It can be shown that there is, at each stage of the processes involved in information retrieval, dilution of that 100% efficiency in recall which can be regularly obtained only by consulting the whole collection in relation to each question. A set of index entries is necessarily something less than the complete document, for the only set of index entries which can match the document is the document itself. The resulting loss may be due to several factors, such as the analysis of the document, the interpretation of that analysis into concepts, the decision as to which concepts are relevant to the purpose of the index and the translation of the concepts into the descriptor language. In retrieval, the same situation comes into play with the question, although here it is often the case that a set of search programmes must be something more than the question as framed.

17. An attempt has been made to measure these factors in the present investigation and, as mentioned earlier, human error would appear to be the dominant reason for loss. The failure has been to work consistently to an unwritten set of rules. The rules were unwritten because, in order to cover all eventualiities, such rules would have to be so multitudinous and complex as to be impracticable to attempt to use. However, a computer can consistently remember and apply all the rules that can be given to it, and from the small investigations which we have made, it would appear quite possible that with machine indexing the elimination of human error would more than compensate for the ability of the human intelligence to make sensible decisions concerning index entries in a situation for which it would be impracticable to formulate any rules. It is therefore possible that machine

indexing could reach a level of efficiency, both for recall and relevance, which would compare favourably with human indexing. While at present it could hardly be considered to be of equal economic efficiency, the increased availability of machine readable text or the availability of print reading machines could materially alter the present situation.

18. A number of investigations on the practice of machine indexing have been published (e.g. Refs. 4 and 5) and no doubt many others are at present under way.* The work thus far has been on a comparatively small scale and it is difficult to measure it against any accepted present day methods, or to compare one technique against another. Work on automatic indexing appears to be concerned with two aspects, firstly the means whereby the computer can select keywords and secondly the means of arranging such key words to ensure retrieval. We would argue that this second point is not directly relevant to the problem, since exactly the same considerations are involved as in a human-operated system. To put it briefly, if for example, some form of thesauri-type headings is shown to be most satisfactory in a certain situation with a manual system, then a similar method should be equally satisfactory in a computer. The selection of the key words (which may then be combined in some way to form thesauri-type terms) is with human indexers an intellectual process; if the investigations showed that this was a satisfactory method for machine

---

*We would not for our purposes include in "automatic indexing" such devices as KWIC indexes, for in such cases the machine has been presented with a list of key words (normally the words which make up the title), and prepares for print-out a permutation of such key words. Our concern in this proposal is with the methods that can be used when the machine is presented with complete text.

indexing, then the investigation must be to find what instructions can be given so that, by an analysis of the text, the computer will select key words which approximate as closely as possible to the terms selected by the human indexer.

19. The earlier work in this particular aspect of automatic indexing was concerned with first establishing a method whereby key words could be selected, and then endeavouring to find whether or not this method produced useful results. Recent work (e.g. Ref. 5) is showing a more logical trend by first hypothesising good indexing and then attempting to formulate a set of instructions for the computer which will give reasonable correlation with the human indexing. It is because, in the main part of this proposal, we shall be analysing in great detail a number of methods of indexing a set of documents, that it would be so valuable to make this approach for automatic indexing.

Methods to be used in the investigation

20. The first requirement is to assemble a basic collection of documents that can remain constant throughout the first stage of the project. The number of documents required for this collection must be large enough to obtain results of a reasonable statistical significance; it is desirable that it should not be larger than necessary to do this, since this will raise the cost of the programme. The most practical size of a collection would appear to be in the range of 1,000 to 1,200 documents.

21. To assemble this collection, we will take some 400 recent research papers in the field of theoretical and applied high speed aerodynamics. Such papers contain, on an average, about five references to other papers. Letters will be sent to the author or an author of each of the 400 papers, requesting their co-operation in the project. They will be asked to do three things, the first two of which will be as follows :-

(i) The author will be asked to frame the original question on which the research was based, that is to say, the question which at the time could not have been satisfactorily answered or otherwise the work need not have been done. In addition, he will be asked to give, if possible, one or two other subsidiary questions which arose in the course of the research.

(ii) The author will then be requested to indicate the relevance, within a certain range, of each of the references (excluding books) included in the bibliography to each of the questions that he has framed.

22. It will be these references only which will form the basic collection, and the original 400 reports will not be included. If half of those to whom we make our request are willing to co-operate, sufficient references will be obtained; otherwise it will be necessary to ask more people. Ultimately it is presumed that there will be a collection of about 1200 documents and also we will have a file of approximately 400 questions. For many questions there will be known to be one or more documents of a certain degree of relevance, namely those which were contained in the list of references of the report written by the compiler of the question. However, it is only by inference that one can say that the other documents in the collection are not relevant and it will therefore be necessary for every document to be analysed in relation to every question.

23. This analysis will be done by a number of post-graduate students working in the subject field, and they will be required to select any documents which appear to be relevant to any particular question. Such documents will then be sent to the appropriate authors and for their third task, the authors will be requested to make a decision as to the relevancy of such papers to the questions they have formulated. Eventually there will be a collection in which every document has been assessed as against

every question, and in which every relevant document has had its relevance decided by a person who has recently been closely concerned with the subject.

24. The next stage will be for the project staff to study each document and prepare a conceptual analysis of it. This analysis could consist of a combination of key words and brief sentences, in some ways not unlike a telegraphic abstract of Western Reserve University, except that it will have no allegiance to any descriptor language. This basic conceptual analysis will contain everything considered worthy of indexing, and will be used as the basis for all further work. Because of its importance to the rest of the work, it will be necessary to make a close check to ensure that it has been done satisfactorily. This preliminary stage having been completed, the objective will be to transcribe this basic conceptual analysis into examples of the 17 different types of descriptor languages previously listed. While at this stage it is difficult to be precise, it is not contemplated that this will involve seventeen different operations, since many of the variations listed are of a degree of refinement of controls rather than being fundamentally different. The result is that in many cases the simpler is included in the more complex and a change in test procedure would provide information concerning both methods.

25. In each case, the 400 questions will be put to the completed indexes and all desired records kept of the results of the searches. In order to obtain the required data to establish the relationships between recall and relevancy, it will be necessary to adopt somewhat different techniques to those used in the present work. This is due to the fact that the relevance of every document in the collection will be known for every question but there will not be a single "source" document. Each search programme will be taken and recorded separately, and the finding of a single relevant document would not complete

the search, which must continue until every document of relevance has been retrieved. The reason for this can best be explained by taking a hypothetical case.

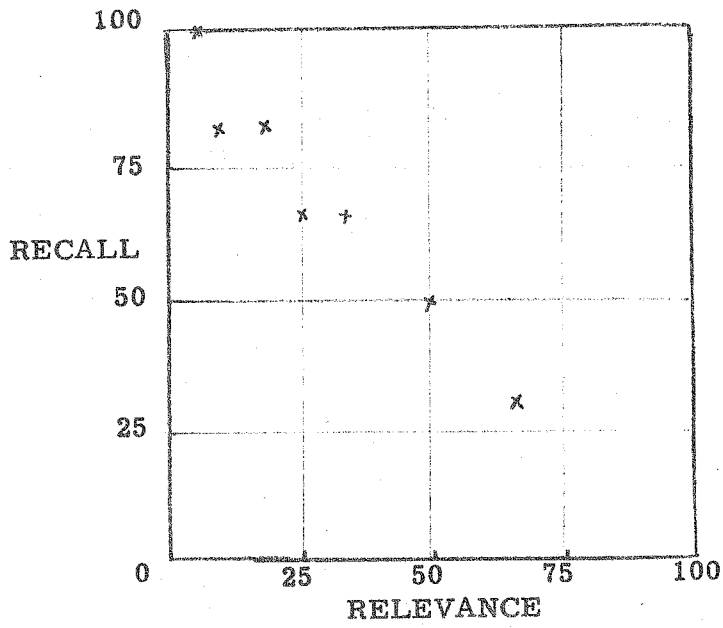
Assume that for a certain question, there are six known relevant documents. Seven different programmes are used in the search, in a particular descriptor language, the results of these being as follows:-

1st	Programme	2	Relevant	+ 1	Non-relevant
2nd	"	1	"	+ 2	"
3rd	"	1	"	+ 5	"
4th	"	0	"	+ 4	"
5th	"	1	"	+ 10	"
6th	"	0	"	+ 18	"
7th	"	1	"	+ 20	"

These results would give the following figures at the end of each separate sub-search :-

Recall	Relevance
33%	67%
50%	50%
67%	33%
67%	25%
83%	18%
83%	11%
100%	9%

This would enable the following points to be plotted:-



With a total of 400 question searched this should result in possibly 1,000 points which can be plotted and from which reliable curves can be obtained.

26. When these tests have been made on all the descriptor languages it may be said that the first objective of this part of the programme will have been completed, but as in the present project it is certain that there will be a considerable amount of data which will justify analysis. Not only might such analysis be expected to develop basic knowledge on the characteristics of descriptor languages, but also to make a contribution to the art of indexing per se, in particular the question of redundant indexing. Some data has been obtained on this point in the present work, but it has not been possible to ascertain with any real degree of precision what might be termed the profit and loss account of redundant indexing.

27. At this stage also it should be possible to decide which techniques appear to have the most satisfactory characteristics for adaptation to

automatic indexing. Dr. J. O'Connor has already explained (Ref. 6) the techniques which can be used to investigate methods of automatic indexing without actually using computors. Our approach would be partly to investigate new techniques but might as usefully be concerned with testing methods proposed by others working in the field and assessing those against the results from human indexing.

28. As in the present work, this proposal is concerned mainly with the descriptor language rather than with the complete system, in even a more specialised sense than at present, where we have been simulating a real life situation, and undertaking a form of experimental engineering. In this proposal, where we shall be operating in artifical conditions, it might be said that we are going back to laboratory testing. It is reasonable for this to be done because it is believed that many of the problems arising from the contiguous aspects of the work have been solved or alternatively because we now know of simpler methods for obtaining the required data.

29. To summarise the proposal, the present investigation has resulted in a definite advance in knowledge concerning the recall efficiency of indexing systems and has shown the relative importance of the various factors which prevent 100% efficiency; it has also given some indications of what can be expected with regard to relevance ratio. We now propose to investigate this latter point in great detail, at a number of different levels of complexity of indexing techniques, conjoin the two factors of recall and relevance, tie the results in with practical matters investigated in the present work and checked in the supplementary testing described in Part I, and produce data which will enable the performance and cost of any system in any environment to be calculated or conversely to allow a system to be designed to meet any given performance or cost requirements.

FINANCIAL REQUIREMENTS

It is proposed that the full-time professional staff should include the following:

(a) Deputy-Director, who would be a senior person with considerable experience in the field, and would be mainly responsible for the day-to-day activities of the project.

(b) Senior Assistant, who would also be expected to have considerable theoretical and practical experience of indexing.

(c) Junior Assistant,

In addition, provision should be made for some part-time staff to help with particular phases of the work. Such staff might include members of the Aslib Research Staff who could be seconded to the project work for short periods and who would be particularly useful for their experience in other disciplines.

The work of the project can be roughly divided into the usual three stages of preparation testing and analysis, although in this programme these stages are less closely defined than in the present work of the Cranfield group. Experience suggests that analysis will take considerably larger than anticipated, but it is felt reasonable to suggest that two years should be sufficient for the programme. Application is therefore made for a grant for $60,952 to cover costs as detailed in Appendix I.

REFERENCES

1. Warburton, B., and Cleverdon, C.W. — Report on the first stage of a test on the Library catalogue of The English Electric Co., Ltd., Whetstone. Cranfield, June, 1961.

2. Cleverdon, C.W. — Report on the first stage of an investigation into the comparative efficiency of indexing systems. Cranfield, September 1960.

3. Newman, S. — Towards searching with assured exhaustive retrieval. Revue de la Documentation. Vol. 28, pp. 5-7.

4. Maron, M.E., and Kuhns, J.L. — On relevance, probabilistic indexing and information retrieval. Jnl. Assoc. Computing Machinery, Vol. 7, 1960, pp. 216-244.

5. Maron, M.E. — Automatic indexing; an experimental enquiry. Jnl. Assoc. Computing Machinery, Vol. 8, 1961, pp. 404-17.

6. O'Connor, J. — Some suggested mechanized indexing investigations which require no machines. American Documentation, Vol. 12, 1961, pp. 198-203.

7. Cleverdon, C.W. — Final report of the test programme of an investigation into the comparative efficiency of indexing systems. (to be published).

APPENDIX I

BUDGET PROPOSALS

	1st Year	2nd Year
Staff (Full time)		
Assistant Director	$5800	$6000
Senior Assistant	4500	4700
Junior Assistant	3500	3600
Superannuation & Insurance 12½%	1725	1787
Staff (Part time)		
Director	4500	-
Technical Assistant	2000	2000
Technical Assistant	2000	2000
Clerical	2000	2000
Consultants	1500	1500
Equipment (permanent)	400	200
Supplies and materials	400	400
Travel	800	800
Publication costs	300	500
Total direct costs	29425	25487
Allowance for indirect cost 11% (Rent, postage, overheads etc.)	3236	2804
TOTAL	$32,661	$28,291

Total Programme $60,952

APPENDIX II

PROJECT STAFF

The Director of the project would be Cyril W. Cleverdon. Mr. Cleverdon was at Bristol Public Libraries from 1932-1938, and then Technical Librarian at Bristol Aeroplane Co., Ltd., from 1938-1946, before taking his present appointment as Librarian of The College of Aeronautics, Cranfield. From his work in 1953 with Thorne on testing of information retrieval systems grew the proposals for the original project which is now nearing completion and for which Mr. Cleverdon has acted as Director.

The Deputy-Director would be Mr. J. Mills, who is at present Senior Lecturer in the Department of Librarianship at North-West Polytechnic College, London. From 1948-1952, Mr. Mills was Librarian of the City of London College, where he was one of the earliest users of the Bliss Classification. A founder member of the Classification Research Group, Mr. Mills has been recognised for many years as a leading figure in the field of classification and indexing, and is the author of numerous papers on the subject.

Among the assistants would be Mrs. B. Murdoch (formerly Miss B. Warburton), who has been engaged on the present project for three years, and Mrs. J. Aitchison (formerly Miss J. Binns) who is also well known for her work on indexing and who has been engaged in the present project for the last year.

No definite decisions have been taken regarding the other posts, but it is known that persons with suitable qualifications would be interested in joining the project group.

On the consultative committee would be Mr. Brian Vickery, who is internationally known for his work in the field and who has been closely connected with the present work since its inception.