CLM - LM 2/71

CLM - LM 2/71

United Kingdom Atomic Energy Authority
RESEARCH GROUP
Library Memorandum

TOWARDS AN EFFECTIVE ON-LINE REFERENCE RETRIEVAL SYSTEM

A. E. NEGUS
J. L. HALL

Culham Laboratory
Abingdon Berkshire

1971

CLM - LM 2/71

TOWARDS AN EFFECTIVE ON-LINE REFERENCE RETRIEVAL SYSTEM

by

A.E. Negus
J.L. Hall

(Paper presented at the Third International Conference on Mechanized Information Storage and Retrieval Systems, Cranfield, 20 - 23 July 1971)

ABSTRACT

Culham Library has been carrying out experimental on-line retrieval studies since 1968. The aim has been twofold: to develop a cost-effective and robust system, and to use this to assess the potential benefits that interactive retrieval could bring to the user community.

This paper describes the Culham RIOT (Retrieval of Information by On-line Terminal) system and the experiments carried out using a KDF9 computer and two widely different data bases, one in computational physics and one specially constructed for the POLLS (Parliamentary On-Line Library Study) experiment. Timing tests on the POLLS data base led to an examination of the economics of on-line retrieval; in the Culham environment the three most important factors in assessing overall economic viability are the cost of storage, the cost of computer time per search and the frequency of use. In RIOT, as in many other systems, the cost of storage rather than the cost of search is likely to be the most critical economic factor. In order to economise on storage costs a technique has been developed in which finding tags, extracted from each bibliographic record, are stored in a linear keyword file complementary to the main data base. Serial search of this concentrated file has given promising results.

U.K.A.E.A. Research Group,
Culham Laboratory,
Abingdon,
Berks.

July, 1971.

CONTENTS

1. INTRODUCTION

The prospect of machine-aided information systems has long been attractive [1], as indeed has been the use of an electronic machine for the storage and retrieval of information [2]. But these early visions of the reference library of the future were clearly ahead of the technology then available and it was only the subsequent development of the electronic computer that allowed practical advances to be made. These at first centred on computer "batch" search, but the advent of time-sharing systems, making it possible for a user to make direct use of a computer without monopolizing it, provided the opportunity to construct information systems with two desirable characteristics. The first of these is speed of access, which will be very important in certain circumstances and is in any case a desirable service aim. An even greater benefit in some cases will be the control the user has over the course of his search as it proceeds, a facility not available to him with an off-line search.

Much pioneering development has been carried out in the United States, particularly noteworthy examples being the Technical Information Project at MIT [3], the Lockheed RECON system used by NASA [4] and the Information Transfer Experiment INTREX [5]. In the social sciences an operational system has been set up in the ILO library in Geneva [6]. Experimental work is also in hand in the UK; in particular, work is proceeding with the support of the Office for Scientific and Technical Information (OSTI), at Culham, at Newcastle University where an investigation is in hand into the use of on-line techniques in testing the formulation of queries to be put to the MEDLARS medical literature computer data base, and at Queen's University Belfast where on-line techniques for retrieval of both data and bibliographic material are being studied.

In this paper we report on the development at Culham of an experimental on-line reference retrieval system, RIOT (Retrieval of Information by On-line Terminal), operating on a KDF9. Section 2 outlines the overall concept of RIOT. In Section 3 we report the stages of development from 1968-1971, dealing particularly with experiments on two data bases, the CPG data base containing computational physics literature references, and the POLLS data base containing references in the Parliamentary sphere of interest. Some aspects of the economics of on-line retrieval are examined in Section 4, while some general considerations are discussed in Section 5. The principal conclusions are noted in Section 6. Details of the Culham KDF9 used in these experiments are given in Appendix A. Extracts from the POLLS User's Manual are contained in Appendix B.

2. RIOT - OVERALL CONCEPT

The aim of the RIOT project is to investigate the value at Culham of on-line retrieval as a routine library tool. The intention is to capitalize on the existing asset of the machine-readable records resulting from the computerized plasma physics SDI system which has been in regular operation since 1966 [7]. These records in the main are based on "titles only" since in the field of plasma physics titles are generally very explicit and contain a high proportion of "key" words; titles can, however, be enriched whenever this seems desirable.

Early consideration in 1967 [8,9] and in 1968 [10], suggested that

(i) in some special libraries, particularly when library bulletin production is via tape typewriter, on-line retrieval could be useful as a technique for searching the "recent" literature, thereby providing adequate immediate in-house access to this much-used but often inadequately indexed portion of stock.

(ii) an on-line retrospective retrieval system offering a feedback facility could be a useful library service.

(iii) the development of video techniques would be particularly important.

The RIOT video terminal and linked teletypewriter (Fig.1(a)) are located

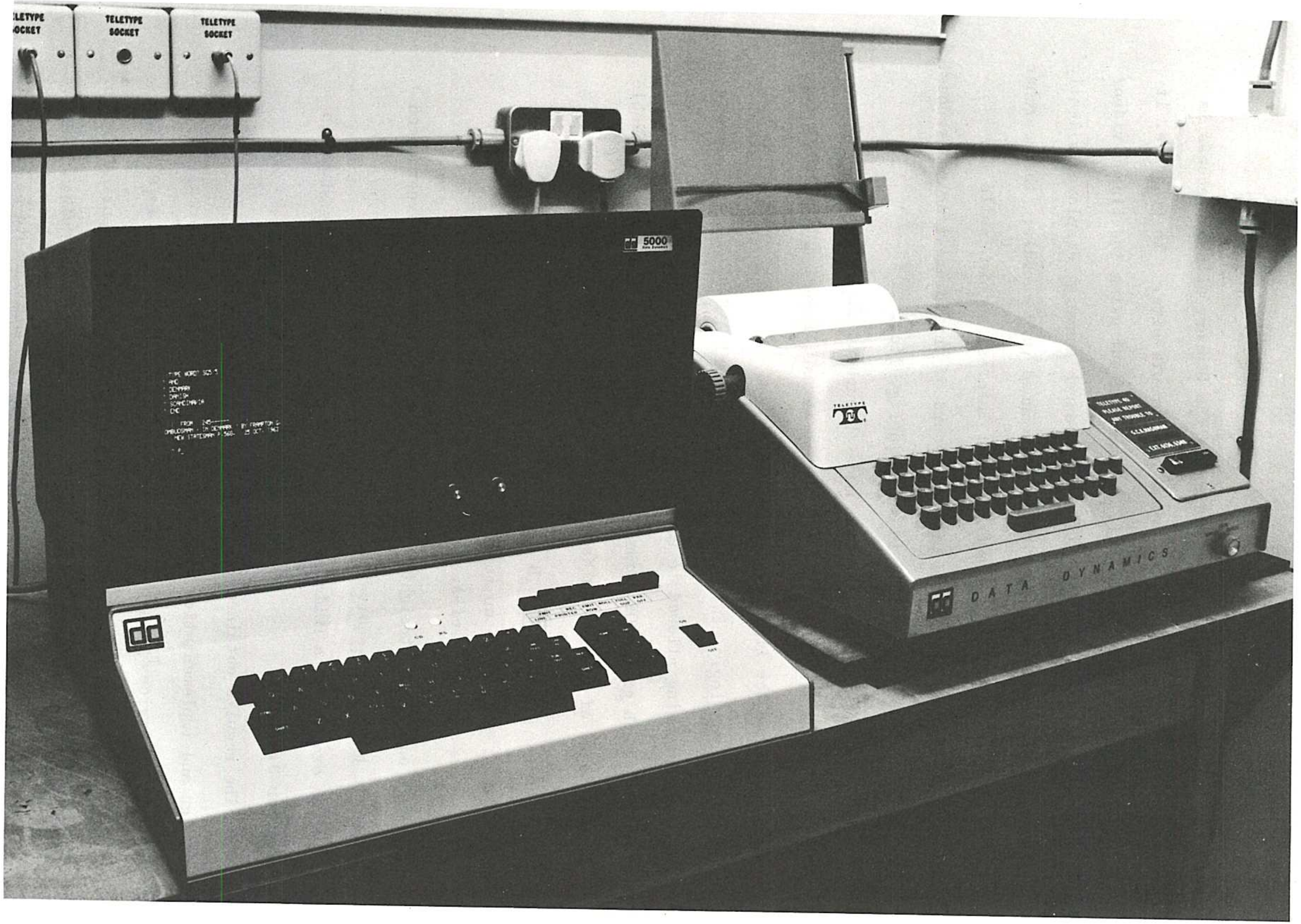


Fig.1(a). RIOT: Video Terminal and Linked Teletypewriter. CLM-LM2/71 near the Scientific Information Office in the Library although access is possible from any other terminal either on or off-site. In practice, it is expected that most searches will be carried out using the linked video/teletypewriter station in the Library. Search answers, displayed silently on the video screen (Fig. 1(b)), can be scanned very rapidly; the searcher can immediately discard references not of interest to him, while the linked teletypewriter keeps a printed record of the interesting references. The schematic outline of RIOT is shown in Fig. 1(c).

A deliberate attempt has been made throughout the development of RIOT to build a system that would be easy to use, robust and amenable to modification in the light of user-feedback.

The search logic now used is based on normal AND/OR/NOT techniques, the searcher being allowed to use up to 4 AND groups together with 1 NOT group. In each group, OR terms (synonyms and alternatives) are allowable up to a total of 9 terms in each group. Thus the general logic employed is:-

A [AND] B [AND] C [AND] D [NOT] X

where up to 8 [OR] terms are allowed for each of A, B, C, D, X if desired.

Although outwardly straightforward, it is possible for the experienced searcher to use the equivalent of a "nested" statement.

An example of a POLLS search carried out in mid-1971 is shown in Section 3.5.1.

The results obtained from the RIOT studies have been used in the design and implementation of an on-line plasma physics information retrieval service [11] to operate on the Culham ICL System 4-70. Over 20,000 references will be available on-line in 1972 representing much of the literature core which is central to the interests of Culham Library users, accepting that other search methods will continue to be used for information searches outside this literature core.

3. STAGES IN THE DEVELOPMENT OF RIOT

Most of the RIOT developments have been carried out on two widely

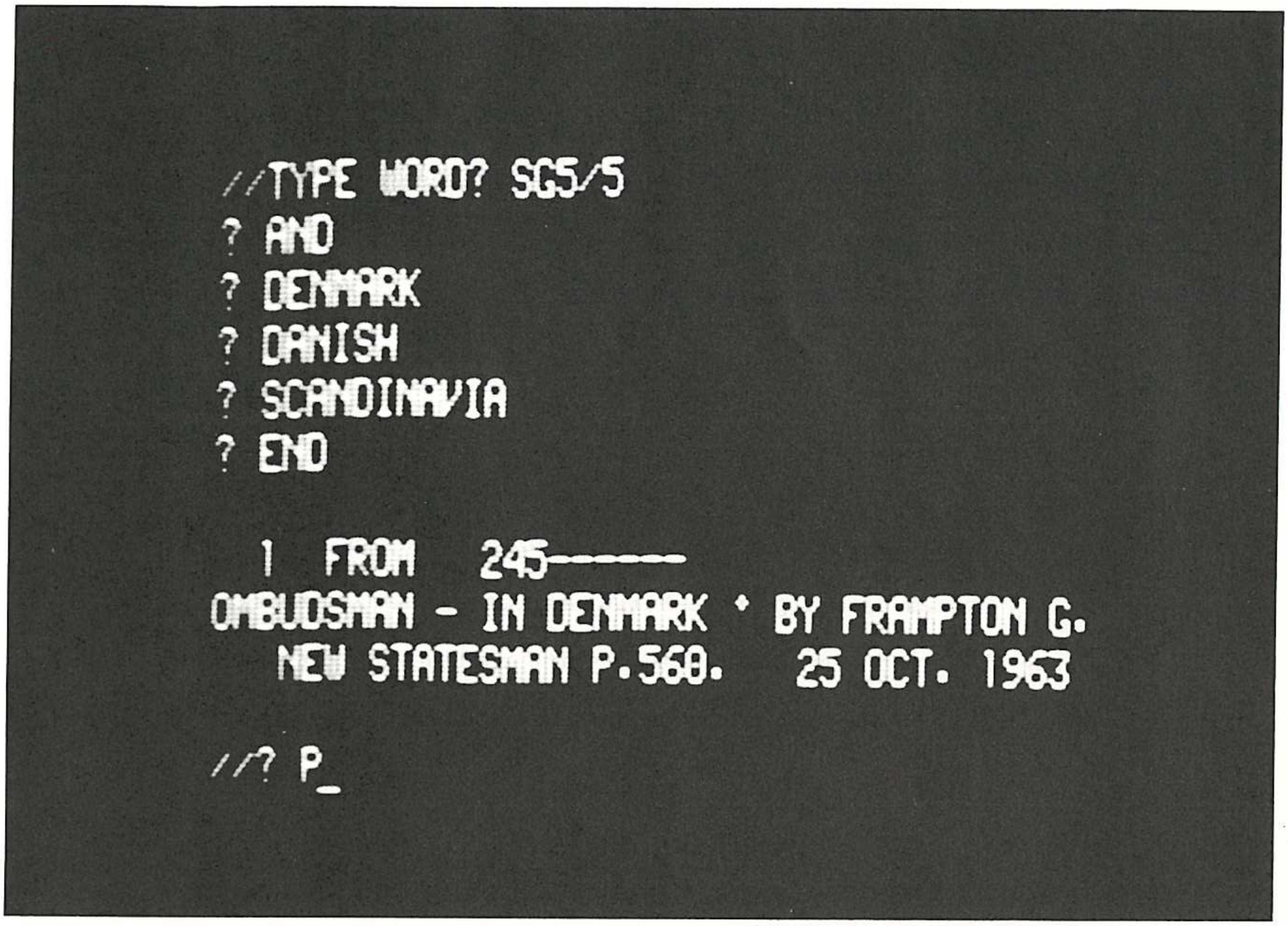


Fig.1(b). RIOT: Close-up of video screen display

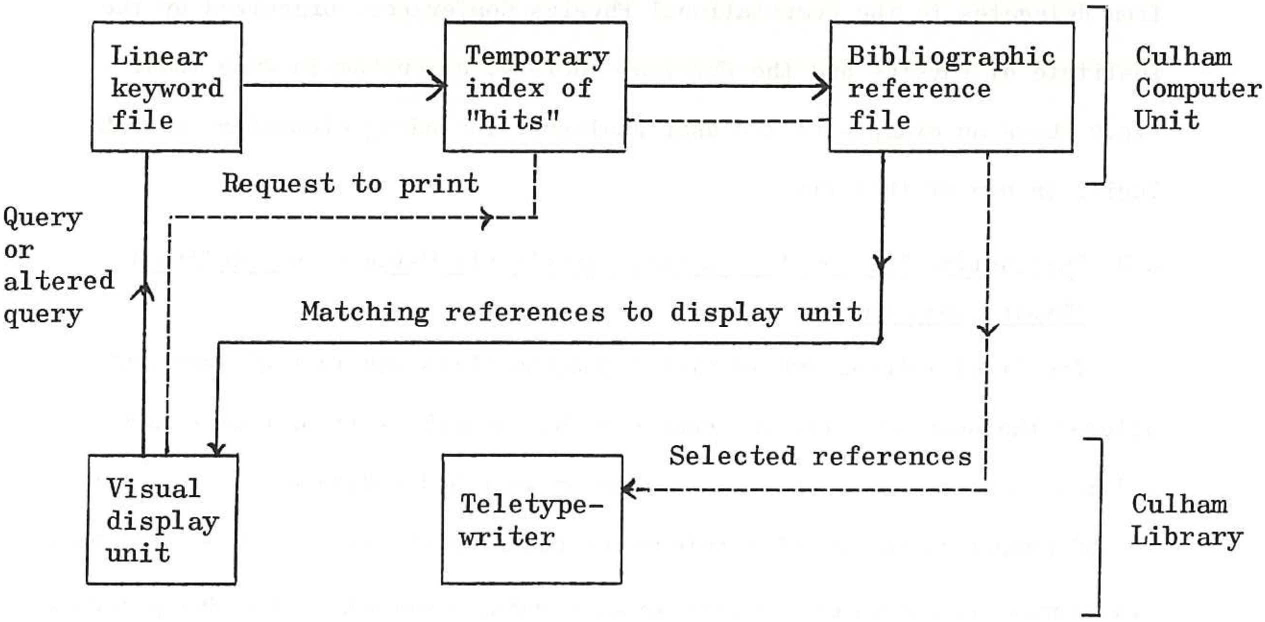


Fig. 1(c). Schematic outline of RIOT as used in tests on the POLLS data base after revision of file structure

CLM-LM2/71

different data bases - the first containing computational physics references, the second covering certain social science literature areas of interest to the House of Commons Library. These data bases, although not large, were large enough to enable a prototype system to be constructed and initial user-reaction tested.

3.1 Remote Entry Batch Search (1968)

The first program in the RIOT development was a remote entry high priority batch job, searching a small file of 100 report literature references drawn from the Culham Library Bulletin during November 1968. This early system used a purpose-built program to search the reports data base which had been constructed by the routine SDI program. Query formulation was handled by the file editor facility available in the Culham COTAN system. Cross-country use of this system was successfully demonstrated at a UKAEA meeting in May 1969.

The first system providing true retrospective retrieval was a development of the remote entry program, searching some 200 references in the field of computational physics [12]. Useful user reaction was obtained from delegates to the Computational Physics Conference, organized by the Institute of Physics and the Physical Society, at Culham in July 1969. Fig.2 shows an example of the user dialogue, including elementary search logic, in use at that time.

3.2 Interactive Retrieval - Initial Experiments Using a Computational Physics Data Base (1969)

The first interactive retrieval program which ran in real time and allowed the user to alter the course of his search as it proceeded was written in September 1969. This program searched a data base consisting of 750 computational physics references covering the period 1946-1968. These references were directly written to disc using a subset of the SDI program.

In December 1969 the program was modified to allow the use of a CRT display for showing references. The user was able to request

```
?** LOGIN
PASS?
USER? LIBDEM
CODE? CPBB
FNS? 0
17.59.38
WAIT


CULHAM ONLINE SYSTEM - COTAN 3/1    18/08/69


SPACE ALLOCATED -     68 SECT.
SPACE RESERVED  -     64 SECT.
FNS 0008 UNITS
?** -CPGBIBLIA
18.00.07


LOGIC (A .OR. B) AND (C .OR. D) BUTNOT (E .OR. F)
FORMAT  AAAAAAA BBBBBBB CCCCCCC DDDDDDD EEEEEEE FFFFFFF

EDIT? R/MONTE   ELASTIC CARLO   SCATTER DIFFUSI ZXZXZXZ
? STOP
18.01.29
WAIT


SAVED

REMAINING RUNNING TIME ON BACKGROUND JOB 1 MINUTES.
YOUR MAXIMUM TIME FOR THIS QUEUE IS 5 MINUTES.

0 JOBS IN THE QUEUE--TOTAL TIME 0 MINUTES.


?** -REFS
18.12.13


SAVED

----------- THE CALCULATION OF NONLINEAR RADIATION TRANSPORT  BY A
 MONTE CARLO  METHOD.  BY J.A. FLECK.   IN  METHODS IN COMP
UTATIONAL PHYSICS.  VOL.1.  1963.  PP.43-65.
```

Fig.2 Sample dialogue used on the computational physics literature data base, August 1969 an immediate print-out of selected references on an adjacent linked teletypewriter.

The resulting system has been fully described [13]. Like its predecessor it was shown to be capable of cross-country use via normal GPO lines [13]. At this stage the RIOT developments were aimed at demonstrating the technical feasibility of interactive retrieval using the facilities available at Culham. In addition, particular attention was paid to the user interface since it was clear that users, while intrigued by the novelty of the on-line approach, could rapidly become disenchanted by difficulties of query formulation and by cryptic dialogue. On the other hand it was also clear that the continual use of lengthy prompts could soon become a major irritant and so it was decided to provide two versions for most prompts - short reminders for the experienced user, but with longer explicit instructions available for the novice. The longer prompts were available, if an "experienced" user needed them, simply by typing HELP at any stage.

3.3 The POLLS (Parliamentary On-Line Library Study) Experiment (1970)

During 1970 the RIOT program was used to search a social sciences data base for the POLLS (Parliamentary On-Line Library Study) experiment. This study, carried out at Culham, was supported by the House of Commons Library, OSTI and Aslib. The data base consisted of 406 references selected, in the three areas of Nuclear Energy, Consumer Protection and the Parliamentary Commissioner, from those references used in the current awareness and retrospective index experiments of 1968 [14]. For this experiment minor modifications were made to the SDI program used to write references to disc, notably by increasing the 'trivial word list' to include words such as Mr, Sir, Hon, which occur frequently in a Parliamentary data base although rarely in the literature of the natural sciences. The on-line data base also included the "enrichment terms" assigned to each reference by House of Commons Library staff during the original experiment. The RIOT program can use these enrichment terms for retrieval as well as the words contained in the reference itself.

The retrieval program was enhanced by adding a number of instruction frames which could be displayed by simply answering YES at the appropriate point in the dialogue. This meant that the system could be used successfully without necessarily reading or rereading the "POLLS User's Manual" [15] which was provided. (A later version of the "POLLS User's Manual" is given in Appendix B.)

3.4 Data Base Structure and Method of Search

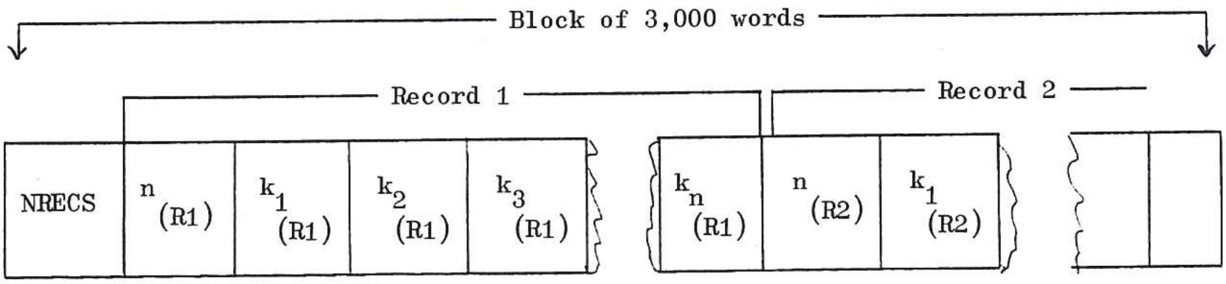
In the spring of 1971 search times on a larger data base such as would be used in an operational service at Culham were investigated. A data base of 12,000 references was constructed by multiple copying of the POLLS data base. Tests on this larger data base constructed by the SDI program confirmed that speed of search was unacceptably slow, the timing for search of 12,000 references being 9 minutes. This clearly indicated that a reorganisation of the data base was essential before a routine, economic service could be provided and an investigation was made of the ways in which the search speed could be increased without affecting the user image of the program and without requiring a complicated, expensive procedure for creating and updating the data bank.

This investigation necessarily included an examination of some of the factors involved in the overall economics of on-line retrieval. To preserve the continuity of this Section the examination of these economic factors is given separately in Section 4. It is sufficient to note here that, while cost of computer time used in actual search is important, the major factor influencing the overall economics of RIOT was the high cost of on-line storage. For this reason, further development has been based on a linear file approach rather than on the more space consuming inverted file approach.

The system finally devised, while retaining the linear approach none-the-less avoids the manifest inefficiencies of straightforward serial search. This is achieved by maintaining two complementary linear files, one being a "keyword file" containing only retrieval tags, the other being the main bibliographic reference file. Information searches are made using only the relatively short "keyword file". The search program can readily calculate the disc address of a reference whose keywords match the query and the actual title and bibliographic details can be displayed. Records in the keyword file are packed as variable length blocks of KDF9 words, each word containing a keyword of up to 7 characters. Blank characters are not removed, as, being a word machine in which the records would have to be unpacked in core, there is little advantage to be gained, the potential saving in disc accesses and overall transfer time being offset by the cost of unpacking.

3.4.1. <u>Keyword File Structure</u>

The keyword file consists of blocks of 3,000 KDF9 words, each containing the variable length records for 200-250 references. The structure of the keyfile is indicated in Fig.3.



NRECS = number of references in this block

n = number of keywords to reference

R1 = reference 1 in this block

$k_{1(R1)}$ = keyword 1 of reference 1

Fig.3. Contents of keyword file block

These 3,000 word blocks are read into core using double buffering, so that after the first block has been transferred to core ($\frac{2}{3}$ second on KDF9) the time taken in searching the files can be within the total transfer time. The overall speed of transfer is therefore 200 - 250 references in $\frac{2}{3}$ second, in effect a rate of 300 - 400 records per second. There seems to be little prospect of any significant further improvement in overall transfer rate on the Culham KDF9 so the problem is to keep the search speed at or above the transfer speed. The original search program could easily achieve a searching speed of this order with a simple query; in fact a figure of over 900 references per second has been recorded when searching the POLLS data base for occurrences of the word "act" and over 1,000 references per second (i.e. a rate of 60,000 references per minute) when making a deliberate null search (e.g. for AAAA) - much faster than the transfer rate which limits overall search speed to 300–400 references per second.

3.4.2. Matching Procedure

A search with two terms, both of which must be present to cause a match is naturally slower than the examples given above but not appreciably so in most cases, as the majority of references will be rejected at the first term and will consequently never be checked against the second term. Even with four 'AND' terms, it is doubtful if the cpu time in searching would be greater than the total transfer time.

The problem is much more pronounced where the question is formulated in such a way that several alternatives are specified rather than a single search term. However, the keywords associated with each reference are already put into alphabetical order by the SDI program so maximum speed can be obtained by putting search terms linked by 'OR' logic into alphabetical order and searching one list against the other.

A single term is generally faster than a group containing several terms, so the search speed can often be increased by looking first for those groups containing the lowest number of alternatives. This does not, however, apply where the single term is high in the alphabet and a group where the highest term begins with, say, K is also present. Equally it does not apply when frequently occurring terms are specified in the smallest group of search terms. A ranking order for groups of terms which allows for these cases can be calculated, with weightings given to frequently occurring terms. Although not necessary at present in RIOT, the use of such a calculated preferred order of search could be used to give considerable increases in speed if this was necessary to keep the overall time taken dependent on the total disc transfer time.

3.5 POLLS Using Linear Keyword File (1971)

Following the studies outlined in Section 3.4, a linear keyword file was created for the POLLS data base, and the search program altered accordingly. The terminal equipment remained unchanged, i.e. a visual display unit with associated teletypewriter providing hard-copy of those references the searcher judges useful. A User's Manual for POLLS is given as Appendix B.

3.5.1 Dialogue

A sample dialogue is shown in Fig.4. The searcher is looking for documents dealing with the work of the Ombudsman in Denmark and Sweden. The various stages of the dialogue are shown with marginal markings, explained below:-

A. The user identifies himself to the computer.

B. The user calls for the search program by typing -RIOT.
The computer then asks him to specify which data base he wishes to search and what sort of terminal he is using. To save the user's time the program is constructed to recognise replies by their first character only.

C. He asks for all those documents indexed with the code SG 5/5 (the code for Ombudsman/Parliamentary Commissioner) which contain at least one of the natural language words likely to be associated with "Denmark" and "Sweden". (This is a relatively simple enquiry using only a part of the full search logic described in Section 2.). The first reference is displayed almost immediately, preceded by a statement of the number of references already searched.

```
A ?** LOGIN
  PASS?
  USER? NEGUSAE
  CODE? ENCP
  FNS? 0

  CULHAM ONLINE SYSTEM - COTAN 3/1    16/07/71

B ?** -RIOT
  *JOBFOREGRND/ENCP  //NEGUSAE /20  *BPRETNEW    0020

  //DO YOU WISH TO SEARCH
  PLASMA PHYSICS
  SOCIAL SCIENCES, OR
  COMPUTATIONAL PHYSICS
  ? S

  //ARE YOU USING 338, TELETYPE OR VDU? V

  //WOULD YOU LIKE AN EXPLANATION OF THE USE OF THIS PROGRAM? N

C //TYPE WORD? SG5/5
  ? AND
  ? DENMARK
  ? DANISH
  ? SWEDEN
  ? SWEDISH
  ? SCANDINAVIA
  ? END


    1  FROM    245-------
  OMBUDSMAN - IN DENMARK * BY FRAMPTON G.
     NEW STATESMAN P.560.   25 OCT. 1963
D //? HELP

  /PRESS RETURN TO SEE NEXT REF ON SCREEN

  /TO PRINT THE REF ON THE SCREEN TYPE  P  BEFORE RETURN

  /TO AMEND THE QUERY TYPE  A  BEFORE RETURN

  /TO STOP THE SEARCH TYPE  S  BEFORE RETURN
  //?

    2  FROM    257-------
  THE OMBUDSMAN * BY BRIAN CHAPMAN * ARGUES THE
  DIFFICULTIES OF IMPORTING A SCANDINAVIAN OMBUDSMAN
  INTO THIS COUNTRY.
     PUBL. ADM. P.303.   WINTER 1960
E //? P

    3  FROM    264-------
  THE OMBUDSMAN * DENMARKS PARLIAMENTARY COMMISSIONER FOR
  CIVIL AND MILITARY ADMINISTRATION * BY HURWITZ S * THE
  ACT AND DIRECTIVES WHICH GOVERN HIS WORK ARE INCLUDED.
     COPENHAGEN.   1962
  //?
```

```
   4  FROM    275------
THE DANISH PARLIAMENTARY COMMISSIONER IN ACTION * BY
PEDERSEN I M * A COMMENTARY ON THE FIRST YEARS OF THE
DANISH PARLIAMENTARY COMMISSIONERS WORK - BY AN
OFFICIAL OF THE MINISTRY OF JUSTICE COPENHAGEN.
   PUBL. LAW P.115.    SUMMER 1959
F//? S

/       4  REFS FOUND
/     275  REFS SEARCHED
/YOU HAVE SELECTED   1  REFS FOR PRINTING
/NOW DEPRESS THE REC/PRINTER BUTTON
/AND PRESS RETURN
//?

-------
THE OMBUDSMAN * BY BRIAN CHAPMAN * ARGUES THE
DIFFICULTIES OF IMPORTING A SCANDINAVIAN OMBUDSMAN
INTO THIS COUNTRY.
   PUBL. ADM. P.303.    WINTER 1960

/NOW RELEASE THE REC/PRINTER BUTTON
/AND PRESS RETURN
//?

G//DO YOU WISH TO SEE THEM AGAIN? Y

/       4  REFS FOUND
/     275  REFS SEARCHED

-------
H OMBUDSMAN - IN DENMARK * BY FRAMPTON G.
//? P

-------
THE OMBUDSMAN * BY BRIAN CHAPMAN * ARGUES THE
//? S

/YOU HAVE SELECTED   1  REFS FOR PRINTING
/NOW DEPRESS THE REC/PRINTER BUTTON
/AND PRESS RETURN
//?

-------
OMBUDSMAN - IN DENMARK * BY FRAMPTON G.
   NEW STATESMAN P.560.    25 OCT. 1963

/NOW RELEASE THE REC/PRINTER BUTTON
/AND PRESS RETURN
//?

//DO YOU WISH TO SEE THEM AGAIN? N

/SEARCH ENDED
?**
```

Fig.4 Sample dialogue used on POLLS data base, July 1971. This is a complete teletypewriter print-out of a search carried out on a video terminal.

D. Various options are available to the user. In this instance he is not sure what to do next so he types HELP and is given a list of these options. He decides to move to the next reference and does so simply by pressing the "return" key.

E. The user asks for a print-out of the previous reference.

F. He has seen enough references so instructs the machine to stop searching. He is given a report on the search and is told how to obtain the print-out.

G. He is given the opportunity of reinspecting the captured references. This could be important if a larger number had been found.

H. As the first reference is being shown the user presses a special key to suppress all but the first line, thereby saving time. He then requests a print-out.

After the search has ended, the computer indicates (by typing ?**) that is is ready for a further instruction. The user can then make another search or tell the computer that he has finished using the terminal.

It should be noted that this is a complete teletypewriter print-out of a search carried out on a video terminal. In the real-search situation the user can if he wishes have a hard-copy print of only those references of interest to him rather than a lengthy and possibly inconvenient full record of the search.

3.5.2 Controlling the direction of search

One of the most important advantages of on-line searching is the opportunity the user has to alter the course of his search. Once the search has started, those references which match the query-formulation are displayed singly, the first usually appearing within 5 or 10 seconds. After each reference has been displayed the user can (if he wishes) amend his query formulation (Fig.4 section D). He may then refine the query by adding additional word groups, or adding synonyms to an existing word group, within the overall logic described in Section 2. Alternatively he may start again with a simpler or different query formulation.

To prevent abuse, or misuse, of the system an automatic cut-off operates if the user seems to be selecting few of the displayed items for printing; the assumption here is that if he does not want a hard-copy of a reference then that reference is not relevant to the enquiry. After this automatic pause (arbitrarily pre-set by the information unit) the user is of course given the choice of refining his query formulation, starting the search again with a differently phrased query, or, if he is indeed satisfied with a small percentage of relevant items amongst those displayed, continuing the search. Once this cut-off has been reached, the percentage of references shown that must be printed to prevent further cut-off is lowered.

When a teletype is being used there is no option to print, so, to minimise core swaps, the program only pauses after printing ten references and then gives the user the options to continue searching, alter the query, start again with a different query or stop searching altogether. It will then print 20 references before pausing again to give the same options.

3.5.3. Speed of response and speed of search

The overall time taken for a search once the query has been formulated depends of course on the number of references displayed and on the length of time taken in showing each reference. Although the present system writes at only 10 characters per second, meaning that it will take perhaps 20 seconds to display a reference, plans are in hand to upgrade the output speed to 120 characters per second, so that the time taken will be virtually dependent on the time it takes the searcher to read a reference and decide whether or not he requires a print-out.

With the KDF9, which has a relatively slow disc backing store, the time to reload the program after an interrupt is slow so there is a delay of 5 - 10 seconds before the next reference is shown. On a computer using a drum or fixed head disc the time lag would be hardly noticeable.

Computer timings on this system have been carried out for 10,000 references and indicate that a search of 25,000 references - which is the number on the plasma physics archive tapes from 1966 - could take 2 minutes transfer time and 30 seconds cpu time.

4. FACTORS TO BE CONSIDERED IN THE OVERALL ECONOMICS OF AN IN-HOUSE ON-LINE SYSTEM

It is well known that computing costs have fallen dramatically with succeeding "generations" of computers e.g. in the case of IBM the stated [16] cost to perform 100,000 calculations is given in Table I.

Year	IBM Model	Cost in new pence
1954	704	50
1959	7090	11
1966	360/75	$1\frac{1}{2}$
1970	370/165	$< \frac{1}{2}$

Table I. Falling cost of computing
(Cost of performing 100,000 multiplications). [16]

Clearly, however, the cost of time used in actual computing is only a part of the total cost and in this Section we attempt to identify the high-cost factors affecting the overall economics of an in-house on-line computer information retrieval system.

The principal factors considered are:-

1. the cost of obtaining the references in machine-readable form
2. the cost of converting these records to the form used by the on-line system

3. the cost of on-line storage

4. the computer time used in searching

5. the charging system in use on the computer

6. the frequency of use of the system.

The relative effects of these factors will of course vary depending on the overall library-user environment, and indeed in some environments there will be other factors which may be very significant, e.g. cost of telecommunication facilities etc.

4.1 Cost of Obtaining the References in Machine-readable Form

Cost will vary depending on individual circumstances. Libraries, such as Culham, already creating machine-readable records will in essence be maximizing the use of an existing asset, although of course it is only reasonable to attribute a portion (e.g. 40%) of the cost of creating that asset to the information retrieval function. Other libraries may be able to build a reference base at reasonable cost by subscribing to a centralized magnetic tape service offering references at rates which can be very low indeed, e.g. 50 references per £1. Many subscribers, of course, will find that tape services sometimes do not capture material of interest to them and often include much material not relevant; it may still be reasonable to use such tapes to provide an SDI service since redundant material is in effect searched only once - but the material "saved" for the information retrieval data base should be free from obviously irrelevant references, either by buying sub-field tapes or by carrying out an in-house "weeding" operation. In general, the cost of references added to a selected "mission" data base will vary widely depending on the route chosen (references prepared in-house or purchased), depth of indexing (if any), computer facilities available, staff overheads etc; rates of 5 - 25 references per £1 could be achieved in many units, the exact figure depending on local circumstances. In these cases a data base of reasonable dimensions could probably be obtained at a cost which is not unreasonable when compared with other components of the library budget.

4.2 Cost of Converting to the Form Used by the On-line System

This will vary with (i) the amount of restructuring necessary, (ii) with the frequency of updating and (iii) with the relative size of the data base and of the added material. With serial files this third factor will have no effect, and, providing an appropriate program is used, neither will the second.

The cost at Culham for converting 1,000 records from the existing cumulative SDI tapes to the form of the file structure held on disc for the retrieval program is < £2. The cost of updating and rearranging an inverted file would naturally be higher.

4.3 Storage Costs

Although the true cost of direct access storage is difficult to estimate due to variations in accounting practice, some idea of its magnitude can be obtained by considering current rental charges.

The following table (Table II) assumes that the information retrieval system is well used and that the reference data base is always on-line. Where the data base is not permanently on-line, storage costs can be considerably lower as the major proportion of these costs is attributed to the hire of drive and control units.

Device	Cost per month	
	1 megabyte	50,000 references at 200 bytes each
RDS Replaceable Disc Store (4 x 7 mbyte)	£46	£460
EDS Exchangeable Disc Store (9 x 30 mbyte)	£12	£120
LFD Large Fixed Disc Store (350 mbyte) (700 mbyte)	£ 7 £ 5	£ 75 £ 52

Table II. Estimated cost of direct access storage

True costs when allowance is made for overheads will naturally be much higher. Clearly, a crucial cost factor in most environments will be the kind of storage device available.

4.4 Computer Time Used in Searching

In machine terms the time used will depend on the size of the data base, the type of search (e.g. serial, inverted etc.) and the number of references found.

In practical terms, however, it is worth pointing out that as systems become more "sophisticated" they often become more difficult to use. The implication, which is often overlooked, is that inexpert use of sophisticated search capabilities can easily nullify any savings arising from the sophisticated logic and programming techniques employed. But whether the logic is simple or complex the skill of the user in formulating his query will still be vital, as will be the provision of early feed-back alerting him to possible errors or short-comings in the query formulation.

Figures will of course vary considerably depending on the approach adopted but taking the specific case of Culham, the timing tests mentioned in Section 3.5.3 indicate that the cost of computer time used in searching 25,000 references could be as low as £3. In the case of a complex query, or one producing many references, computing costs will be greater. These figures are based on computer charges taken at £95 per hour on a second generation machine. With a third generation time-sharing system where charges are based on cpu time, computing cost would probably be less than £1.

One point to note here is that the time taken in search with an inverted file system increases in almost direct proportion to the number of terms in the query, whereas the affect of each additional term on speed of search in a serial system is normally less than the effect of the preceding term.

It is also worth noting that in many real-life units most searches are for some literature on a topic rather than all. Unlike many inverted file systems, the RIOT program stops searching when the user stops looking at captured references thus economising on computer search time.

4.5 The Charging System in Use

Most time-sharing systems make a charge based on cpu time, as of course the machine is running another program while transfer is taking place. The user will probably be unable to alter the rates at which he is charged. An interactive job will almost certainly run at the highest cost rating on a machine. Some idea of the range of costs that might be involved is given above in Section 4.4.

4.6 Frequency of Use of the System

In the costs listed above, some are fixed costs and others are costs incurred each time a search is undertaken. The more searches there are on a system, the cheaper the cost per search will be. (This may itself cause a change in user demand and this is discussed further in Section 5.3.)

4.7 Effects of These Factors on the Economics of an In-house On-line System

4.7.1 Data base permanently on-line

It is apparent from the preceding examination that assuming machine-readable references can be purchased or prepared at a reasonable annual cost then the most significant factors affecting the overall economic viability of an on-line reference retrieval system are (i) the cost of storage, (ii) the cost of computer time per search, and (iii) the frequency of use. These three factors are examined in Fig.5; the curve for each storage device represents costs which are essentially the total of two components - the storage cost without overheads (derived from Table II) and the computer search cost (based on the figures given in Section 4.4). By ignoring overheads, and assessing computing costs at a notionally low level, on-line retrieval can be made to look very attractive indeed.

However, "real" costs are much higher. The major extra component of these real costs is the allowance that must be made for overheads, particularly the heavy overheads attracted by the storage units. The "shaded" region (in Fig.5) shows possible "real" costs taking overheads into account; the major portion of these costs in all cases is the standing storage charge arising from having references available on-line, the cost of search being a relatively small proportion of the total. (In fact as the "real cost" region in Fig.5 is extrapolated the "search cost" as a proportion of the total cost assumes increasing importance until with a very high number of searches per month it can become dominant.)

It follows from the above that the most crucial factor for many units, particularly small/medium size units, will be the cost of storage and not the cost of search; where a system is expected to operate with maximum economy the best approach will be that which uses the least on-line storage, providing always that the system operates at a speed which is acceptable to the user. In effect, it must be borne in mind when considering whether to adopt an inverted file approach or a linear file approach, that although inverted file search systems generally use less computer time in actual search, this may be more than offset by the cost penalty imposed by the increased amount of computer storage required for the larger files. Comparative studies on the relative sizes of linear files and inverted files [17] are valuable in giving some idea of the storage penalties which an inverted file system can incur.

As mentioned above, the cost of computer time, as a proportion of

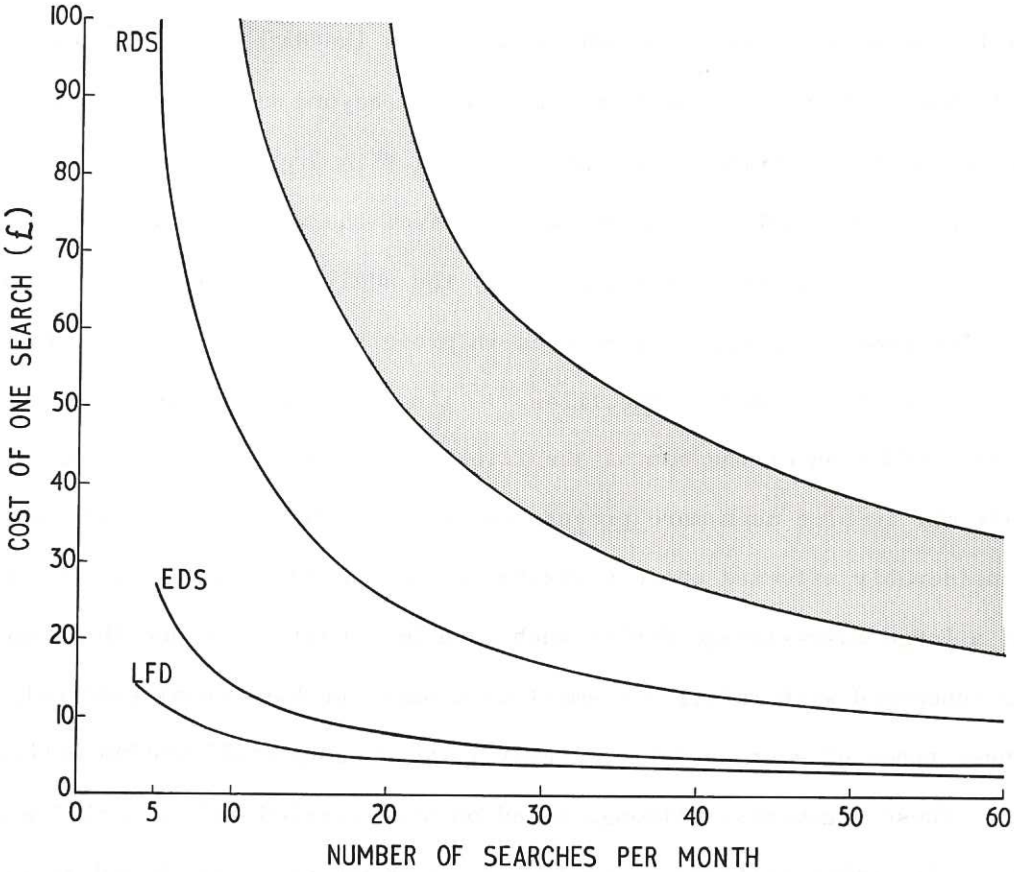


Fig. 5 Estimated costs per search as a function of the number of searches made per month, for a data base of 50,000 references held continuously on-line. The "line" curves represent search costs based on rental charges for individual types of storage device (RDS = Replaceable Disc Store; EDS = Exchangeable Disc Store; LFD = Large Fixed Disc.)
"Real" costs when overheads are taken into consideration will fall within the "shaded" region. All costs include an estimated charge of £2 for the computer time used in searching (even the cost of a complex, more expensive search will still be a relatively small proportion of the total).

CLM-LM2/71

the total cost of the average search, can become a significant factor in a relatively heavily-used system. This may happen, for instance, in the case of a large-scale system covering a discipline and available on a national or international basis. If there are no natural sub-field compartments which in effect partition the data base into small data bases, then a threshold may be reached beyond which decreased search costs resulting from, say, inverted file organization may more than offset (i) the additional costs arising from the extra storage requirements of the larger inverted files, and (ii) the additional significant costs arising from the computer time used in processing inverted files when new material is added to the files. (In a linear file new material is simply added on at the end of the file, whereas up-dating an inverted file may involve much more processing time.) The economics will be considerably affected where a system is large enough to justify the use of a large mass-storage device such as a Data Cell. (Since this paper is concerned with relatively small data bases we have considered only those types of storage device likely to be used by small/medium units.)

These arguments, although based on work carried out on a KDF9 are valid for other machines. In fact, the storage costs are based on 1970/71 equipment whereas the search costs are for a second generation machine This means that the relative importance of storage as a very significant high-cost factor is perhaps even greater than indicated above, the System 4/70 for example being approximately three times faster than the KDF9.

4.7.2 Data base not permanently on-line

It is clear, considering the case of the small/medium data base examined in Fig.5, that if the references are held continuously on-line then with a traffic of, say, 3 searches per day the cost per search would be about £30. This is the assessed "real" cost but it is likely that in many units the expenditure (in terms of money actually spent) would be less. Nevertheless, costs of £15 - £30 would seem high enough to cause library managements to ask if there was a sufficient need for the reference store to be on-line for the whole of the computer's working day.

In many cases (particularly in units using high-cost storage) it would be judged adequate, initially at any rate, to provide on-line service for only part of each day. Many factors have to be evaluated - for what proportion of the computer's working day is the retrieval system to be available? Is this at a peak demand period attracting a high tariff? What kind of storage devices are available and is there spare capacity or must extra drive/control units be obtained?

Clearly, when the data base is held on-line for only part of each day the search cost will depend very much on local conditions. However, Fig. 5 may be used to indicate upper limits from which an appropriate deduction may be made e.g. it is not unreasonable to suggest a cost of £5 - £15 per search when the traffic is, say, 3 searches per day on a data base of 50,000 references available for 4 hours per day.

If the computer searches are performed with reasonable efficiency, then these costs (£5-£15) are not high when compared with the cost of manual searching (when staff overheads etc. are taken fully into account). For instance, one attempt [6] to assess the cost of a manual search derived an estimated cost per search of $65.

4.7.3 Data base partitioned

In some cases there may be a distinct advantage in partitioning the data base, either into naturally occurring sub-fields (e.g. an industrial library may have considerable collections covering technology and a number of quite separate fields such as marketing), or in some other way. If on-line storage is restricted but a limited amount can be made available throughout the day then "core" literature can be on-line continuously but with peripheral literature available only for short periods. Partitioning can also increase the speed of search and indeed simplify query-formulation. An example of simple partitioning can be seen in the dialogue given in Fig.4.

5. SOME GENERAL CONSIDERATIONS

5.1 Amount of Computer Storage Required

The studies reported in this paper are particularly centred on the literature core of practical interest to the Culham Laboratory - about 25,000 - 50,000 references. This is not a large data base but nevertheless it is typical of many special libraries. It is useful at this point to make an estimate of what these figures, e.g. 50,000 references, mean in computer storage terms. This has been done in Table III which gives an estimate of the number of "characters" involved together with the the number of magnetic tapes or disc units required to store this information. It is clear that the conventional book or bound journal is a remarkably compact storage device comparing well in volumetric terms, and indeed in cost terms, with machine storage. It is also clear that 50,000 references stored in machine-readable form, e.g. on 1 magnetic tape, occupy considerably less storage space than would the equivalent conventional card index. But while the bulk factor is important what is more important is that in a card index or book index the entry points are predetermined. In contrast, given appropriate search and programming techniques almost any part of a machine-readable bibliographic record can be used for retrieval in a computer search.

	Approx. no. of characters	Approx. no. of magnetic tape reels	Approx. no. of magnetic disc units
1 reference	2×10^2	not calculated	not calculated
25,000 references	5×10^6	$\frac{1}{2}$	$\frac{1}{4}$
50,000 references	10^7	1	$\frac{1}{2}$
500,000 references	10^8	10	5
1 journal article	10^4	not calculated	not calculated
1 volume of Encyclopaedia Britannica	7×10^6	$\frac{2}{3}$	$\frac{1}{3}$

Table III. Amount of computer storage required for various sizes of data base. These are based on 10^7 characters for a magnetic tape and 2×10^7 characters for a disc unit; clearly they are order-of-magnitude estimates since many widely different conditions occur in practice e.g. packing density, type of disc unit available etc.

5.2 Time-scale for Possible Implementation

It is clear that on-line retrieval is technically feasible and has certain advantages which make it a potentially powerful information finding tool. It is also obvious that few libraries are yet in a position to implement a working on-line system giving access to their own particularly valuable and relevant in-house literature collection. However, the development of on-line retrieval as a routine library tool must be seen in its proper time-perspective as likely to be spread over the next decade, or even longer. Development will speed up as the library community's computer expertise increases, as centralized data bases are established, and as equipment and storage costs fall and the computer "terminal" becomes just a commonplace item of office equipment.

On-line retrieval will be particularly attractive to special libraries where the literature core is of a size such that it is quite realistic to consider storing references in an on-line file. Ideally, of course, a data base would be fully retrospective but for practical reasons it is possible that there may have to be an arbitrary time cut-off, e.g. references would be held on-line for, say, 5 years and with each up-dating of the total file the oldest material would be passed automatically to an archive off-line store. If a selective filter of some sort exists, or can be set up, then the amount of material to be stored on-line could be relatively modest and the overall cost of maintaining an on-line file could be in reasonable harmony with other components of normal library budgets.

5.3 Possible Effect on Pattern of User Demands

The possibility has already been mentioned, in Section 4.6, that the provision of a rapid search service might cause a real change in the pattern of library "user enquiries", and indeed might have an effect on user "satisfaction" in general. It is well-known that many enquiries are best answered using standard works, e.g. studies [18, 19] at AERE Harwell indicated that perhaps only 20 - 35% of the enquiries put to the Scientific Information Office might be suitable for computer search. The prevailing use-patterns seem to indicate that an on-line service might expect only a moderate traffic flow. But this does not necessarily follow; if an on-line service met valid needs then as increasing use was made of the system individual search costs would decrease, thereby encouraging more intensive use (but only, of course, up to some as yet unascertained level).

6. CONCLUSIONS

For many information units the most crucial factor in considering the economics of in-house on-line retrieval will be the cost of the computer storage facilities available to them rather than the cost of computer search. The cost of storage can be minimized by having the reference store on-line for only part of each day, and/or by having the data base organized in such a way that it occupies minimum storage space. At the Culham Laboratory the storage required for the library reference files is minimized using a simple, robust technique in which finding tags, extracted from each bibliographic record, are stored in a linear keyword file complementary to the main data base. Serial search of this concentrated file has been shown to give promising results. In the case of the specialized literature collection on which the RIOT system has been tested the keyword file occupies a minimum of expensive on-line storage while achieving acceptable search rates.

It is also clear that video retrieval techniques, particularly when associated with a printer device to give hard-copy on request, are much more advantageous than a straightforward teletypewriter approach - interaction is silent, "found" references can be displayed at a much faster rate than is feasible with a teletypewriter, while the provision of a linked printing device enables the user to obtain immediately a printed list of only those references judged to be of interest.

Acknowledgements

We would like to thank OSTI for encouragement throughout the whole of the period during which RIOT has been developed, and for part-sponsorship of some parts of the work. Thanks are also due to the Librarian of the House of Commons and his staff for cooperation in the POLLS experiment which was part-sponsored by the House of Commons Library. We are also grateful to members of the Aslib research staff for the part they played in the POLLS experiment.

At Culham, we wish to acknowledge the many helpful discussions we have had with Mr A. G. Cheney in the early years of the development of RIOT, and more recently with Mr D.J. Dancy who is currently working with us on the implementation of an operational plasma physics reference retrieval system based on RIOT. We are also indebted to the many members of the staff of the Culham Laboratory Computing and Applied Mathematics Group for much useful advice and help given to us from 1967 onwards. Our thanks are due to Mr R.S. Peckover, of the Computational Physics Group at Culham, who collaborated with us in the early stages of the RIOT development.

REFERENCES

[1] V. Bush: As we may think. Atlantic Monthly, 1945, 176, 101.

[2] L. Mountbatten: Presidential Address to the Institution of Electronic and Radio Engineers, London, October 1946. J. Brit. I.R.E., 1946, 6, 221.

[3] M. Kessler: The MIT technical information project. Physics Today, 1965, 18, 28.

[4] D. Meister and D.J. Sullivan. Evaluation of user reactions to a prototype on-line information retrieval system. Report; NASA CR-918, October 1967. (N67-40083).

[5] C.F.J. Overhage and R.J. Harman: INTREX. M.I.T. Press (1965).

[6] G.K. Thompson: Some cost estimates for bibliographic searching in a large-scale social sciences information system. Inform. Stor. Retr., 1970, 6, 179.

[7] L.J. Anthony, A.G. Cheney and E.K. Whelan: Some experiments in the selective dissemination of information in the field of plasma physics. Inform. Stor. Retr., 1968, 4, 187.

[8] L.J. Anthony, D.H. Carpenter and A.G. Cheney: Selective dissemination of information using a KDF9 computer. Aslib Proc., 1968, 20, 40 (in particular, page 49).

[9] J.L. Hall and R.D. Pullen. UKAEA, AERE Harwell; unpublished work, 1967.

[10] A.G. Cheney: Development of an on-line retrieval experiment at Culham Laboratory. UKAEA,Culham Laboratory; Library Memorandum; LM 1/68, January 1968.

[11] J.L. Hall, A.E. Negus and D.J. Dancy: Towards instant information. New Scientist, 1971, 51, 210.

[12] A.E. Negus and R.S. Peckover: CPG BIBLIA. UKAEA,Culham Laboratory; CPN No. 3/69, August 1969.

[13] A.E. Negus: A real time interactive reference retrieval system. Paper, including live demonstration, presented at Institute of Information Scientists 4th Conference, Reading, April 1970. Inf. Scientist, 1971, 5, 29.

[14] J.L. Hall, J. Palmer and J.B. Poole: An experimental current awareness service in the social sciences; the House of Commons Library/Culham Laboratory Project. J. Doc., 1970, 26, 1.

[15] J.L. Hall and A.E. Negus: POLLS User's Manual. UKAEA, Culham Laboratory; Library Memorandum; CLM-LM 4/70, July 1970.

[16] Minutes of Evidence taken before the House of Commons Select Committee on Science and Technology, Sub-Committee D, 25 February 1970, 137-iii, page 68 (IBM Supplementary Memorandum). The figure in respect of Model 370/165 was provided later by IBM.

[17] L. Schultz: in Library Use of Computers (editors: G.L. Smith and R.S. Meyer), Special Libraries Association, New York, 1969, pages 78-81.

[18] R.M. Fishenden: Information Use Studies; Part I - Past results and future needs. J. Doc., 1965, 21, 163.

[19] R.C.M. Barnes: The present state of information retrieval by computer. UKAEA, AERE Harwell; Report; AERE-R4514, London, HMSO, January 1964.

Details of Culham KDF9 Computer

Main store:	32,768 words
Word length:	48 bits (= 8 characters)
Disc:	3,932,160 words

Maximum transfer rate 10,660 words/second

Typical effective transfer rate 5,000 words/second

Front end machine:	PDP8 - 8K core
Operating system:	EGDON 3
On-line multi-access system:	COTAN 3
Programming languages used in retrieval program:	EGTRAN (FORTRAN IV) and USERCODE (UCA3)
No. of remote console channels:	19
Graphical display:	DEC 338
Terminals:	Teletype model 33 and Data Dynamics VCT 5000 video communications terminal

POLLS: USER'S MANUAL

(The following is based on Culham Laboratory Library Memorandum, CLM-LM 4/70, by J.L. Hall and A.E. Negus [15]. It should be noted that this User's Manual was primarily intended for users with virtually no experience of computer processing of information.)

P O L L S

USER'S MANUAL for the

Parliamentary On-Line Library Study

1. This experiment uses the Culham Library's Video RIOT (Retrieval of Information by On-line Terminal) facilities, in which "captured" references can be displayed on a video screen. It should be noted that Video RIOT is a prototype and is capable of much further modification and development.

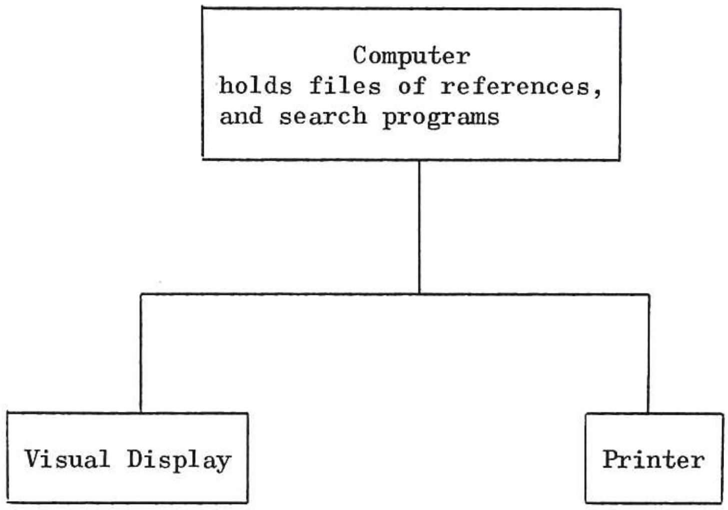
2. The on-line store for this experiment contains 406 references, as follows:

   Nuclear Power (196 references)

   The Parliamentary Commissioner (105 references)

   Consumer Protection (105 references)

3. The following diagram shows the basic arrangement of Video RIOT:



The Video unit allows the user to communicate with the computer and to scan references.

The teletypewriter enables the user to have immediately a permanent list of those references he wishes to keep.

HOW TO USE THE SYSTEM

4. Information can be retrieved by combining terms from any or all of the following groups of finding "keys"

Enrichment terms added by Library

Subject Group (SG) numbers*	e.g.	SG17/13 is the "key" for the subject area "Nuclear Power - Safety and Siting".
Document Material Type (MT)*	e.g.	MTO is the "key" for "Legislation".
Natural language enrichment	e.g.	CAPITAL INVESTMENT IN GAS INDUSTRY

Natural language terms occurring in title

Surnames & place names	e.g.	HINTON, CULHAM
Words in title of document	e.g.	FUSION, PLASMA, BLOOD, PRIME MINISTER

---

*All references in the computer store are enriched with SG and MT keys. Full details of the range of SG and MT keys used were given in the original Manual but are not repeated in this extract.

5. Questions can be framed by combining up to 4 search "words" (and if desired up to 8 synonyms or alternatives can be used for each "word"). One exclusion word (with up to 8 synonyms or alternatives) can also be used. Thus the general logic employed is

A [AND] B [AND] C [AND] D [NOT] X

where up to 8 [OR] terms are allowed for each of A, B, C, D, X if desired.

6. Only a complicated search would require the use of the full logical string; many "real-life" enquiries may be framed using only a few "words".

SOME EXAMPLES OF QUESTION FORMULATION

USING SUBJECT "KEY" AND DOCUMENT-TYPE APPROACH

7. Consider a search of the subject area "Nuclear Power - Safety and Siting" (which is Subject Group SG17/13) for "Legislation" documents (Material Type MTO). This is clearly equivalent to an

A [AND] B search

in which the computer is being instructed to retrieve only those documents which somewhere within their bibliographic description or indexing contain both the groupings

SG17/13 [AND] MTO

(Note that

MTO [AND] SG17/13

will capture exactly the same references). The appropriate computer-user dialogue for the first of these two formulations is as follows (user response follows the computer-generated ? which indicates that the machine is waiting for data or for an instruction):-

//TYPE WORD? SG17/13 ?AND	) To combine "finding keys" link) them by AND as shown
?MTO	
?END	) On typing END then pressing the) "return" key the computer will) carry out the search

FINDING PLACE NAMES (OR SURNAMES)

8. Questions relating to place names (or indeed other names) may be formulated very simply; thus references containing the word Culham will be obtained from the following formulation:

```
//TYPE WORD? CULHAM
?END
```

FINDING ANY (non-trivial) WORD(S)

9. More complicated "natural language" searches can be made by combining the finding "keys" in any combination allowed by the logical string given in para. 5. Consider a search for references dealing with the PLASMA PHYSICS or NUCLEAR FUSION RESEARCH carried out at CULHAM LABORATORY, but excluding references on STELLAR PLASMAS. The main useful finding keys (remembering that it is necessary in this prototype program to enter plurals where appropriate) are:-

PLASMA PLASMAS FUSION	These can be considered as alternatives: at least one of them must be present
CULHAM	This is a separate element that must be present
STELLAR ASTROPHYSICAL	These are exclusion terms and only one of them need be present to prevent capture of a reference. The searcher has thought it wise to put ASTROPHYSICAL as an alternative to STELLAR.

The searcher therefore wishes to construct an A AND B NOT X search

A	AND	B	NOT	X
PLASMA (or PLASMAS or FUSION)	AND	CULHAM	NOT	STELLAR (or ASTROPHYSICAL)

The appropriate dialogue with the computer is as follows:-

//TYPE WORD? PLASMA ?PLASMAS ?FUSION	Synonyms and alternatives are typed on succeeding lines - there is no need to type OR.
?AND ?CULHAM	To combine "finding keys" (or groups of alternatives) link them by AND as shown.
?NOT ?STELLAR ?ASTROPHYSICAL	The exclusion term group is linked by using NOT as shown
?END	On typing END then pressing the "return" key the computer will carry out the search

SOME SYSTEM OPTIONS AVAILABLE

PROCEDURE FOR SCANNING REFERENCES

10. As each reference is displayed the computer reports the number of references so far searched. The following options are then available to the user:

(i) he may cause display of the next reference by pressing the "return" key

(ii) he may request a print-out of the reference by typing the letter P

(iii) he may stop the search completely by typing S

(iv) by typing A he will be given the opportunity to alter his query; the computer will then show no more captured references but will respond with the standard prompt:

//CONTINUE, REFINE OR START?

and this is explained in full in the following paragraph.

N.B. selective use of the RETURN and P options means that the searcher can scan a large number of captured references and build up a printed list of those references, and only those references, in which he is interested.

PROCEDURE FOR CONTINUING FULL SEARCH, AMENDING QUERY OR STARTING AGAIN

11. The options available here are CONTINUE, REFINE OR START?

(a) type C User satisfied with search so far and instructs computer to continue with search of rest of store of references.

(b) type R User slightly dissatisfied with references displayed and wants to refine the query by adding new terms. (To remove terms use the next option i.e. start again.)

(c) type S User completely dissatisfied and wants to start again with a different query formulation.

AUTOMATIC CUT-OFF

12. A poorly formulated query may use an excessive amount of computer time. To guard against this an automatic cut-off is incorporated in the search program. If 10 references have been displayed and none selected for printing it is assumed that the references found are not relevant and so the user is asked to decide whether the search should continue; the computer reports the number of references searched and asks the user if he wishes to

//CONTINUE, REFINE OR START?

HELP

13. Suitable prompts and comments built in to the computer dialogue allow the user to obtain help and guidance from the system. The "HELP" facility can be used immediately after any computer response ending with a question mark, thus the dialogue

//CONTINUE, REFINE OR START? HELP

will result in a more detailed explanation of the prompt preceding the ?

TO TERMINATE INTERACTION

14. The program may be terminated at any stage (e.g. if in severe difficulty) by holding down the "control" key and typing R, release "control" then press "return".

TO CORRECT A TYPING ERROR

15. To correct a typing error, "cancel" the immediately preceding incorrect character(s) by pressing the "rub out" key (which prints out as ? on the paper), e.g. COMIS??MISSIONER will be interpreted by the computer as COMMISSIONER.

NOW PRACTISE USING THE SYSTEM

16. In general, it is better to begin by constructing simple query formulations. The computer will always report how many references it has found and then give the searcher an opportunity to refine the enquiry thus avoiding waste of computer time.

17. A sample dialogue is provided (Fig.4). Note that the computer will type a question mark ? when it expects a reply or response from the user. The user types his response on the same line then presses the "return" key (i.e. the initiative for action is being returned to the computer). In the case of the actual query formulation, however, where the searcher wishes to retain the initiative he simply presses the "line feed" key to obtain a fresh line for his next search term.

18. Using the sample dialogue provided (Fig. 4) as a guide the enquiries given in the paragraphs above may be put to the POLLS file. The following answers should be obtained:

Para. above	Query	No. of Refs.
7	SG17/13 AND MTO	2
8	CULHAM	3
9	PLASMA (or PLASMAS or FUSION) AND CULHAM NOT STELLAR (or ASTROPHYSICAL)	2

19. Users may wish to construct their own "test" queries based on the material known to be in the data base. A full list of the material, arranged in broad subject categories, is available. This demonstration file is deliberately small so that users can cross-check, in respect of each query, that the query formulation was valid, i.e. that it retrieved all the material sought.