3 SYSTEM DESCRIPTION

3.1 Introduction

This chapter describes the retrieval systems which were used for the evaluation. In a previous report [WALK87b] we separated "Design and implementation" from "System description". One reviewer [HANC88] suggested that it would have been easier to read if the two had been combined. In the present report we have done this, although this makes rather a long chapter.

Three systems were used in the evaluation experiments. These are referred to as "dumb" or "D", "qe" or "Q" (for "query expansion") or "intermediate", and "full" or "F".

The dumb system is similar to the "exp" Okapi system described in [WALK87b, Ch 7], performing a "best match" search on words, phrases and stems extracted and derived from users' input. It has the added capability of allowing the user to compile a list of selected books to be printed out.

The qe system has all the features of the dumb system together with a query expansion option in which it offers to "look for books similar to" those already chosen.

The full system has all the features of the qe system and an additional "shelf browsing" option. This enables the user to browse through books in Dewey class number order - roughly the order in which they would appear on the shelves, using the book most recently selected as a starting point.

The three experimental systems are very similar to each other in the sense that it is not always immediately obvious which system one is using. They differ mainly in the browsing facilities provided. We also occasionally refer to a hypothetical "production" system - one that might be used with real users in a real library. All the systems are embedded in one program which can be run with different parameters to produce various search systems.

3.2 The bibliographic database

Each of the systems searches an index of keys derived from title-like and subject heading (Library of Congress and PRECIS verbal feature heading) fields of PCL's MARC monograph catalogue. This index also contains corporate and conference names from author fields. There is an index of Dewey numbers which is used by the qe and full systems.

Okapi bibliographic files are derived from UK MARC tapes. Much information is discarded during this process, including physical description and statements of responsibility. However, because they are intended for subject searching they contain far more information than the minimum recommended by the Centre for Catalogue Research report [SEAL82]. In particular they contain all names, most of the title

information and all subject headings. Some punctuation is added to enhance display, and there are a few special characters which indicate the type of indexing required or the presence of non-ASCII characters which may not be displayable.

The bibliographic and index files are described in more detail in Appendix 1.

3.3 Hardware and software

The search and indexing programs are written in C, and run on Sun 3/50 processors under Sun Microsystems' Release 3.4 of their version of the 4.2 BSD UNIX operating system. Many of the functions and data structures are similar to previous versions of Okapi (which were written in Z80 assembly language and ran on 8-bit machines). Each Sun processor would probably handle upwards of eight simultaneous users, although no serious testing has been done. The terminals are of a monochrome VT220-compatible type using 25 lines of 80 character spaces, with some use of a line-drawing character set to make windows and boxes.

3.4 Dialogue style

All commands are single keystrokes. The user only has to type the initial letter of the appropriate option and this is always highlighted and underlined for emphasis. An effort has been made to avoid side effects resulting from typing more than the initial letter, but this does not always work very well when the terminal is connected over a network. A large proportion of computer users are used to screen editors and other programs where commands do not have to be terminated, and we have found that most other people quickly become accustomed to the single keystroke style. The system is actually looking for a minimum unambiguous abbreviation, but the command names have all been chosen so that they differ in the first character whenever they are valid at the same time. "D" is used both for "display" and "down", but these are never simultaneously valid, and anyway have closely related functions, the "down" command resulting in a new display screen containing the "next" records. If an invalid command is entered it is echoed and remains on the screen for a few seconds or until another key is pressed. Occasionally an invalid command produces a warning message: a "down" command can produce a little box or "snippet" containing "You're already at the end - try something else".

3.5 Obtaining and processing user input

3.5.1 Obtaining input

As the three systems are designed for subject searching only, there is no choice of search types such as author, title, subject. The first screen the user sees is the search input screen (Fig 3.1). This is clearly headed "SUBJECT SEARCH" to deter people from trying to look for specific items by author and/or title. A single sentence tries to suggest what the system does. This description is simplified but, in stating that the system looks for books including "all (or most) of your words in their titles or subject descriptions", it should give enough information to be useful in deciding how to phrase a search. The term "subject description" is doubtless unfamiliar to most users but it has a degree of transparency. "Subject heading" might be preferable in the United States. It is important to convey that the system searches on

information other than the title alone.

The catalogue will accept a search statement consisting of a single word, a list of related words, a "natural language" phrase or even a Dewey number. However, it simply asks for "a word or a phrase which describes the books you want" as most people seem to find it easier and more natural to type a fairly coherent phrase than a string of unconnected terms. The input area is delineated by a rectangular box which will take up to 76 characters of input; experience has shown that this is almost always enough. To start the search process it is necessary to press the Return key. People familiar with multi-user computer systems may do this without being reminded, so there is no permanent prompt. If after a few seconds the user appears to have finished typing and has not pressed return, a message (or "snippet") appears just beneath the input box (Fig 3.1). Messages which appear suddenly and unpredictably are more likely to be noticed than ones which are part of the scenery.

3.5.2 Input preprocessing

The search is now put into a form in which it can be parsed into constituent words and subphrases to be looked up in the indexes. Punctuation and spacing are tidied up, letters are put into lower case and most non-alphanumeric characters removed. This produces a pre-processed search statement which is comprehensible but not always exactly what the user typed. For example "Post-war theatre & cinema actors organisations in the U.S.A" becomes "postwar theatre and cinema actors organisations in the usa", "statistics,regression,correlation" becomes "statistics regression correlation". The pre-processed search is then disassembled into words. Regular English plurals and "ing" endings are removed and spellings are adjusted to remove many of the differences between American and British English, and the search is reassembled into a phrase. The example now becomes "postwar theater and cinema actor organization in the usa". This is purely an internal representation of the search. It is not shown to the user, who would not normally need to know that these minor adjustments have been made. There is a fuller description of this pre-processing in Appendix 2. (Note that a similar process is applied to the text of the bibliographic records during indexing.)

3.5.3 Parsing and lookup

Then the look-up screen appears (top of Figs 3.2 - 3.4). Constituents of the pre-processed search statement are looked up in a database which contains the systems' rather meagre linguistic knowledge. This database is known as the "go/see list", or gsl. The gsl contains five classes of entries. These are (1) stop words like "the" and "a", (2) common prefixes, (3) classes of terms which are to be treated as synonymous ("usa", "united states", "united states of america"; "child", "children"), (4) phrases which behave like single words ("soap opera") and (5) "dubious" words and phrases which although they cannot be stopped are not to be given much weight in comparison with other terms. Examples of terms in class 5 include "introduction", "lecture notes", "system", "theory". There are fuller details of the construction and use of such a gsl in an earlier version of Okapi in [WALK87b, Chapter 6].

All the components of the pre-processed search statement are then looked up in the gsl to identify stop words, words and phrases which have synonyms, phrases to be treated as words and "dubious" terms. For

example, the search "TV commercials and soap opera" will have been pre-processed to "tv commercial and soap opera". "Tv" finds a match in the gsl (it is equivalenced to "television"), so the first constituent term is "television" and its synonyms. "Commercial" is not in the gsl, so this is the second term. "And" is in the gsl as a stop word. Finally "soap opera" is matched as a phrase.

As soon as a term has been identified and looked up in the indexes the result is shown to the user. If the term is found the system displays a line like '343 books under "tv"' (Figs 3.2 - 3.4). If the term cannot be found in either the gsl or the indexes there is a message like 'can't find "televsion"' (Figs 3.2 and 3.3). If it is in the gsl but not in the indexes (the system can "know" terms which do not appear in any bibliographic record) there is a message like "No books under "haitian"'. In the two last cases the user is forced either to replace the word or to instruct the system to ignore it; when the user has taken the appropriate action the search restarts with the message at the top of the screen showing the revised query.

Few other reference retrieval systems force the user to make a decision about terms which the system does not know. In the case of keyword systems with implicit AND this usually leads to "zero hits" (although there are systems which recover to the last "non-zero" result). In other "best match" systems such as LIBERTAS there may be a non-zero result but it will often be a misleading one. Users often do not notice that a term has not been found. Misspellings and miskeyings are frequent. Our way of dealing with this problem is simple to implement and makes a significant contribution to efficiency and usability. (A previous Okapi system would sometimes try to suggest a replacement. We concluded that this was worth doing, even if it did suggest "teacher" for "Thacher". The rather tight matching procedure was less profligate with absurdities than are some of the commercially available spelling checkers. It managed to suggest a replacement rather more than half the time, and more than three-quarters of the suggestions were correct. We did not have time to incorporate this feature in the systems described here.)

3.6 Search processing

3.6.1 Term weighting and the merge

Each term is assigned a weight which depends inversely on the number of records which contain it - a rare term is worth more than a common one. Following Croft and Harper (2.2) and many others we use one of the formulas

term weight = $\log(N / n)$

or

term weight = $\log((N - n) / n)$

where n is the number of postings for the term and N is a constant which must be larger than the number of postings for the most highly posted term in the search. For the experiments we used $N = 32768$ for all searches. The most highly posted term (the class containing "great britain") had about 12,000 postings. Terms of "dubious" subject content (see 3.5 above) then have their weights (arbitrarily) reduced by 50%.

The weight of a retrieved record is the sum of the weights of those of the query terms by which it is indexed. A "maximum possible" weight is calculated. This is the weight of a record containing all the terms of the query, a record which would be retrieved by a system performing an implicit boolean AND on the search terms. Two threshold weights are then assigned - a weight above which a record will be deemed to match "quite well", and a weight below which a record will not be retrieved at all. In general these thresholds are set at two-thirds and half the maximum possible weight. These values are arbitrary but well tried and seem to be reasonable. They could be lowered somewhat for searches containing more than three or four terms. Searches containing only one or two terms are treated differently. Previous experience has shown that up to two-thirds of all searches contain one or two terms.

In a single term search the weight of the term is of course irrelevant: a record either matches or not. Two term searches are treated as follows. Terms are either "rare" or "common". A term is rare if it would not be too tedious for the user to look at all records containing the term. For the experiments the rarity threshold was set to 64. Any other term is common. For all two term searches both terms have to be present for a record to match "well". The setting of the lower threshold depends on whether the terms are (1) both rare, (2) one rare and one common or (3) both common.

TWO RARE TERMS

Example: "vebe consistometer". All records containing either term are retrieved.

ONE RARE TERM AND ONE COMMON

Example: "the psychology of biscuits". All records containing the rare term are retrieved.

TWO COMMON TERMS

Example: "economic history". Records must contain both terms.

When counting the number of terms a "dubious" term is not supposed to be counted. There are some errors in the program which is supposed to deal with this, so some searches containing dubious terms do not quite behave as expected.

When term weights have been assigned the postings lists for all the terms are merged. The output from the merge consists of all the postings for records which have a weight of at least the lower threshold (minimum acceptable weight). These are counted and sorted in order of decreasing weight. This is the order in which records will be shown, if the user chooses to display records.

3.6.2 Initial search results and options

On completion of the merge, the system summarises the results of the search (Fig 3.4, bottom half). The messages give some indication of how well the records may be expected to match the search. There are never more than two lines in the results message, which can indicate how many match the search "well" or "fairly well" and how many have been found altogether. If there are no records of "good" weight the message may say "n books found, but they don't match your search very well".

Options are displayed beneath the results message (Fig 3.4). "Display" is always at the head of the list (unless no items have been found). It is set apart from the other three options, encouraging the user to look at the books the system has found. Display is not discouraged even when the system has found a large number of records of good weight. With the dumb system it might be sensible to permute the order of the options according to results, and make suggestive prompts about trying "a more specific search" (many records) or "a different way of expressing your search" (very few records). The qe and the full systems are intended to help the user alter the breadth and focus of a search as it progresses. Even so, there was no significant difference between the systems in the number of searches per session.

The other options are presented in a separate group. "New" goes to the screen of Fig 3.16, which displays up to three previous search statements in a session, to remind users of what they have already tried. "Edit" goes to a screen which is the same as Fig 3.16 except that the input box contains the previous search with the cursor positioned two spaces after the end of it, so that the search can conveniently be repeated, added to or truncated back. "Quit" takes the user to an empty initial input screen (Fig 3.1).

3.7 Record displays

3.7.1 Brief record display

When given the "display" command the system shows a screenful of brief bibliographic records, each one taking up a single line on the screen (Fig 3.5). The display is headed "LIST OF BOOKS". Up to nine records are displayed on the screen. Each record contains title, author, Dewey number and publication date, displayed in fixed length fields under column headings. Any field except date may be truncated. The title field is 39 characters long so most titles are truncated. In some cases the truncated title could mean that a user fails to identify a possibly useful item. However, a truncated field is always transparently indicated by the two dots immediately following it. This type of fixed field display probably has more advantages than disadvantages. It is almost certainly quicker to scan than the two-line layout used by one or two of the commercial systems, and it carries more information per screen without being difficult to read. Single-line displays with variable length fields separated by slashes have been used in some systems. They carry more information, but are unappealing and difficult to read.

RECORD NUMBERING

Brief records are numbered with single digits 1 to 9. Ten or eleven records would fit quite comfortably on a 25 line screen, but this is not compatible with the single keystroke command style. (A few systems have used letters for line identification instead of numbers. Most people are much quicker at finding numbers than letters on the keyboard. This problem can be alleviated - as can many others - by using a mouse or tracker ball for command selection.)

If a set of records extends over more than one screen the numbering starts again from 1. There are arguments for continuous numbering, but this again is not compatible with single keystroke commands. In any

case, unbounded, browsing displays such as the shelf browsing display in the full system (3.10, Fig 3.13) cannot be numbered continuously. Whenever the user is looking at a "finite" set of records the systems display a message of the form "Books 10 to 18 of 33" in the top right hand corner of the screen.

ORDER OF DISPLAY

Records are displayed in descending weight order. Records with equal weights are displayed in reverse order of publication date (most recent first) and within this by author and title. There is no sorting of records (apart from by weight) when the system is operating. The reverse date order is achieved by keeping the bibliographic file in what is in effect pre-sorted order.

When the weight of displayed records crosses one of the threshold weights there are messages "the rest of the books may match your search less well" or "the rest of the books may not match your search very well". One of these messages can be seen in Fig 3.5, where the first three records contain both the words of the search but all the rest contain only the word "insecticide" (records under "environment" alone have not been retrieved because it is a common word, with more than 1200 postings).

CONTEXT INFORMATION AND OPTIONS

Other important information is displayed at the top and bottom of the screen (Fig 3.5). Context information is at the top. In the top lefthand corner is the heading which tells the user that this is a "LIST OF BOOKS". The search statement is displayed immediately underneath this. In the top righthand corner is the message indicating the current position in this list of books.

The information at the bottom of the screen tells the user what can be done next. Immediately beneath the record display window is the message which encourages users to look at further details for particular items: "Type its number to see if a book is relevant". The word "number" is highlighted and refers to the reference number from 1 to 9 which is assigned to each record. In some previous systems we have used "Type its number for fuller details of one book", and in others there was no brief display at all. For the experiments described here it was essential that users should indicate which records they considered relevant. Brief records are unreliable indicators of relevance. We did consider using full records only, as in the systems described in [WALK87b].

The list of valid commands is given at the foot of the screen. If there are more than nine records the user can go "down" to see the other books which have been retrieved. Having gone down a screen the user can come back up again and can move freely up and down the list. Other options, including the umbrella "Restart/new search/quit", are discussed in later sections.

3.7.2 Full record display

On responding to the prompt to "Type its number to see if a book is relevant", the user obtains a "full" bibliographic record for that book (Figs 3.6, 3.7).

While the record is being formatted for display it is "indexed" - that

is, terms are extracted from subject-rich fields, including Dewey number, just as they were when the file was being indexed. These terms are, to start with, associated only with the current record. They are used to enable the systems to highlight those terms in the record which caused it to be retrieved. (In systems like Okapi where records may be retrieved by terms which are not literally in the search this highlighting is not very straightforward. It cannot be done, as it is in some systems, by stringsearching the text of the record. Again, there are some faults in the program, and highlighting does not work very reliably in the experimental systems.)

In the dumb system these extracted terms have no further use, but in the qe and full systems they may be used to expand the current search. If the user answers "yes" to the relevance question (see below) the extracted terms are added to a list which contains the terms extracted from all the chosen records. This list is used as the source of search terms if the user asks the system to look for "More books like the ones you have chosen".

RECORD CONTENT

The display shows the author and other associated names, main, series and part titles, publisher, date of publication, and subject headings. Each field or field group is on a separate line and is clearly labelled. Information regarding number and location of the copies held, along with the shelfmark, is given at the foot of the record. This location information was redundant (except perhaps for the Dewey number) during the experiments, because subjects were not asked to consider availability, merely to choose a list of books. It would be essential in a production system.

As in the brief display, context information is given at the head of the screen. This comprises the display heading, search statement and the position occupied in the list of retrieved items by this particular book as "Book 1 of 26" (Fig 3.6). Terms in the record which are included in the user's search statement are highlighted so that the user may see why the item has been retrieved.

3.8 Obtaining relevance judgments

At the foot of the full record in all three systems users are presented with a question which they must answer before proceeding. In the dumb system the user is asked "Would you like the computer to remember this book in case you want to look at it again?". After two or three records have been seen this rather ponderous question is shortened to "Remember this book?". If the user answers "y" the system adds the item to a list of chosen books which can be reviewed (3.11) and which is printed out at the end of the search.

In the qe and full systems the question is different: "Is this at all the sort of book you are looking for?" (wording suggested by Professor S E Robertson of City University). As with the dumb system the item is added to a list of chosen books if the user's reponse to the question is "y". However, the qe and full systems also use the subject-rich fields from chosen records in their query expansion facilities and the full system can use a chosen record as a starting point for shelf-order browsing. The phrasing of the question in these two systems reflects this attempt to build up a profile of the user's area of interest.

In all three systems, if the user answers "no" to the question at the foot of the full record the system returns to the original display of brief records. A rejected record is marked by a single star in this and all subsequent displays. A record chosen as relevant is marked with two stars. The stars are not explained, and some of the experimental subjects did not grasp their meaning.

3.9 Query expansion

The primary object of this research was to take some steps towards testing the efficacy of query expansion. The general idea is that terms extracted from records chosen by the user as relevant are added to the terms derived directly from the user's query, and a new search carried out.

3.9.1 Term extraction and weighting

Figs 3.6 and 3.7 show the user judging two books relevant to the search for "insecticides and the environment". The system adds the terms it has extracted from title, subject and class number fields to the list of terms which it will use if the user requests query expansion. This list already contains the query terms "insecticide" and "environment". It now also contains potentially useful terms such as "insect", "control", "organochlorine", "persistent", "pollutant", "pesticide", "environmental" and two Dewey numbers, as well as a few unhelpful terms like "economy", "natural" and "selection". Against each term is the number of relevant records in which it has occurred. This number is referred to as r, and is used in re-calculating term weights before an expanded search is done.

As soon as at least one record has been chosen as relevant on the Q or F systems an additional prompt, "Type More to look for books similar to the one(s) you chose", appears on the brief record display screens (Fig 3.8). If the user selects this option, weights are assigned to the terms in the term list and the "best" terms are selected and used for a new search. The formula used for weighting is based on the F4' formula of Robertson and Sparck Jones given in 2.2. The formula is

$$\log((r + 0.5 / R - r + 0.5)/(n - r + 0.5 / N - n - R + r + 0.5))$$

where N and n are as given above in 3.6.1 for the original search,
R = number of records chosen as relevant,
and r = number of chosen records which contain the term

When R and r are zero - that is, before any records have been judged relevant - the formula gives substantially the same weights as the initial weighting rule given above. Its general effect is to increase the weight of terms which occur in most of the relevant records and decrease the weight of those which tend not to occur in relevant records.

Example:

Original search term weights (N = 32768):

Term	postings (n)	weight
insecticide	47	103
environment	1204	26

The most highly weighted terms after the records shown in Figs 3.6 and
3.7 have been judged relevant (N = 32768, R = 2) are:

Term	postings (n)	rel recs with term (r)	weight
organochlorine	2	1	144
668.651	2	1	144
chlorine	3	1	136
insecticide	26	2	127
persistent	7	1	123
632.95042	11	1	116
pollutant	28	1	102
pesticide	63	1	90
biochemical	67	1	89
insect	118	1	81
compound	138	1	78
organic	204	1	73
environment	1204	2	70
selection	331	1	66

After further records have been retrieved and chosen, the Dewey number
632.95042 eventually reaches the top of the weight table, followed by
'pesticide'. This Dewey number represents undesired effects of pesticides and
their control, and it is used to classify nine of the thirteen books
eventually chosen in this search (Fig 3.15).

The system selects the most highly weighted terms from the term list and
performs a new search. A parameter determines the maximum number (up to
32) of terms used for query expansion. Some informal experiments were
done to determine a reasonable range of values for this parameter. If
set too low, at 4 for example, expansion searches tended to retrieve
just a handful of records indexed by a few rare terms (one of the
criticisms of the Robertson/Sparck Jones formula is that it gives
comparatively high weight to rare terms even if they have appeared in
only one relevant record - see 2.2). At the other extreme, using a large
number of terms seemed to give generally good results, but with
decreasing returns, and searches using dozens of terms need computing
resources which might be unacceptably high in a live situation. There
was a noticeable but not very significant difference between the results
obtained with 16 and with 24 terms. The parameter was (fairly
arbitrarily) set to 16 for the experiment. The term selection process is
described in the following subsection.

3.9.2 Selection of terms for query expansion

The term list is sorted by descending term weight. Starting at the top
of the list terms are selected provided (1) the user has not already

seen all the records indexed by this term, (2) the term is not in a list of "dubious" words ("introductory", "fiction", etc) (see 3.6), (3) the term weight is positive and (4) fewer than 16 terms have been selected.

Each of the selected terms is then looked up, and a merge performed just as in the original search. This process may take some seconds. As in the original search, a threshold weight is applied below which a record will not enter the output set. This threshold is based on the weights which would be achieved by the records already chosen by the user, if they were retrieved now. In other words, the new records must have a substantial amount of indexing in common with those previously judged relevant. Normally, some or all of these records will come out at or near the top of the list in the new set of records. Since they have already been chosen they are removed from the new set. Any records which have been previously rejected by the user are also removed. However, the new set often contains records which have already been seen in brief, but neither chosen nor rejected. Sometimes there will now be no records remaining in the new set, particularly when the expansion has been based on just one richly indexed record.

3.9.3 Query expansion search - screen display and options

A window opens overlaying the record display from which "More" was selected (Fig 3.9). If there are no records of high enough weight in the new set there is a message "No more books found". Unlike in the original search the system cannot realistically guess how close any of the new records are to what the user wants, so it merely reports that the most similar books should appear first (unless there are very few, in which case it is silent). If there are some new records but the user chooses not to see them the "More" option will still be available, and simply retrieves the same set again.

Fig 3.10 shows the two records found by query expansion. The display is similar to other brief record screens, but is headed "LIST OF BOOKS similar to the one(s) you chose", beneath which is a reminder of the original search. The first book is on the analysis of waters for insecticides. Although it is certainly relevant it is perhaps rather specialized. This record has no less than eleven subject headings. Its retrieval, at the head of the list, illustrates the tendency which almost all keyword systems have to retrieve the most extensively indexed records. Sparsely described records, which may be recognizable as relevant to a knowledgeable user, are rarely retrieved. Browsing titles in classification order should help in the retrieval of some of these records (3.10).

Fig 3.11 shows the user selecting the second record. The terms by which this record have been retrieved are highlighted (apart from the Dewey number). Selection of this record will add the useful term "pollution" to the list, and will increase the weight of a number of other good terms.

3.10 Shelf order browsing

In the full system, as soon as a user has judged a record relevant (from the original set of records or from a set resulting from query expansion) the system asks a second question: "Would you like to see books shelved near this one?". If the user replies "n" the system returns to the screen of brief records from which the full record was

chosen. If the user replies "y" the system displays a screenful of brief
records in classmark order. An example is shown in Fig 3.13. The book
from which the display was chosen is in the centre of the display window
(record 5) and is of course marked with two stars to indicate that it
has already been chosen as relevant. Using the "Down" and "Up" options
the user can browse indefinitely forwards and backwards in the classmark
list. Records already seen in full are marked with one star if the
record was rejected or two stars if the record has been chosen. Records
may be selected from this shelf-order display in exactly the same way as
from the original brief display.

Records in the classmark display contain classmark, title, author and
publication date in that order. As with other brief displays, the first
three fields may be truncated. Apart from the fact that the classmark
is the leading element, the structure of the displayed record is the
same as in the initial display of brief records. Records are numbered 1
to 9 on each screen. The options below the display window include "Type
Back to return to your previous list of books". "Back" returns
to the display from which the classmark option was chosen. Context
information is given at the head of the screen. As the number of records
the user can look at is not limited it is not possible to give an exact
position in the list as with other brief record displays. The display is
headed "LIST OF BOOKS classified near" the classmark of the book from
which the display was entered, with the original search statement shown
below this message, as in the display of the results of a query
expansion search.

3.11 Other options

Other facilities which are always available from a brief display are
"quit", "new search" and "edit search". There are two further options
which are only available if at least one record has been judged
relevant. These are "Print" and "View". "Print" does what it says -
sends the chosen records to the system printer. It did not appear on the
experimental systems as they were set up to print automatically on
completion of a search. The "view" option was introduced at the
suggestion of some of our advisers. They felt that users would forget
what they had chosen and hence there should be an easy way of reviewing
choices (Fig 3.15).

It was originally intended that these options would be displayed on and
directly available from the bottom line of the brief record display
screens along with "Up" and "Down". This was tried, but the designers
felt that the options line was becoming cluttered and difficult to read.
It was therefore decided that these important but less frequently used
commands should be offered from a subsidiary menu. This is shown in Fig
3.14. We were unable to think of a good name for this choice. "Other
commands" commits the sin of not making it obvious to the user that
there is a way of finishing or starting again, so the option was
labelled "Restart/new search/quit". Unfortunately this does not make
evident the existence of the "View" option. During the evaluation
experiment subjects were shown the "Restart" menu during the
experimenter's introductory spiel. This went against our principle that
retrieval systems should be operable at sight without training. Some
redesign of the interaction and options layouts would be necessary in a
production system.

Fig 3.16 shows the second and subsequent input screen which appears if

the user selects "New" from the subsidiary menu. The functions of "New",
"Edit" and "Quit" have already been described (3.6.2).

Fig 3.1 Subject search input screen - new search

```
===============================================================================
                           SUBJECT SEARCH                      ** OKAPI


The computer will look for books which include all (or most) of your words
in their titles or subject descriptions


Type a word or a phrase which describes the books you want :

  ------------------------------------------------------------------------
: insecticides and the environment█                                       :
  ------------------------------------------------------------------------
              -----------------------------
              :Press Return when you have   :
              :finished typing your search  :
              -----------------------------


===============================================================================
```

Fig 3.2 Display while looking up - word not found

```
===============================================================================
SUBJECT SEARCH for 'insecticides and the envirnment'

   26 books under 'insecticides'
CAN'T FIND 'envirnment'


Press the Return key if you want to change this word

Press the space bar to continue without this word

===============================================================================
```

Fig 3.3 User replacing a misspelt word

```
===============================================================================
SUBJECT SEARCH for 'insecticides and the envirnment'

   26 books under 'insecticides'
CAN'T FIND 'envirnment'


Type your replacement : environment█......................................


===============================================================================
```

Fig 3.4 Lookup and search results

```
================================================================================
SUBJECT SEARCH for 'insecticides and the environment'

    26 books under 'insecticides'
  1204 books under 'environment'


3 books match your search well
(26 books found altogether)

Type Display to look at the books found


Type New if you want to do a different search
Type Edit to change or add to your search
Type Quit if you have finished


================================================================================
```

Fig 3.5 Brief record display - original search

```
================================================================================
LIST OF BOOKS                                           Books 1 to 9 of 26
Search: 'insecticides and the environment'


--------------------------------------------------------------------------------
No. Title                                 Author         Classmark  Date
--------------------------------------------------------------------------------
1   Biochemical insect control : its impact.. QURAISHI M S   668.651    1977
2   Advances in environmental science and ..  (METCALF R L)  628.5      1976
3   Organochlorine insecticides : persisten.. (MORIARTY F)   632.95042  1975
THE REST OF THE BOOKS MAY NOT MATCH YOUR SEARCH VERY WELL
4   Insects, experts and the insecticide cr.. PERKINS J H    632.7      1982
5   Proceedings of the 1979 British Crop Pr.. British Crop .. 668.65    1980
6   Proceedings of the 1979 British Crop Pr.. British Crop .. 668.65    1980
7   Proceedings of the 1979 British Crop Pr.. British Crop .. 668.65    1980
8   The insecticide, herbicide, fungicide ..  PAGE B G       632.95     1979
9   Organochlorine insecticides and polychl.. Standing Commi.. 628.168  1979
--------------------------------------------------------------------------------

Type its number to see if a book is relevant 1


or type Down (next), Restart/new search/quit
================================================================================
```

Fig 3.6 Full record display

```
==============================================================================
FULL DISPLAY                                                    Book 1 of 26
Search: 'insecticides and the environment'
------------------------------------------------------------------------------
  AUTHOR(S): QURAISHI M S
    TITLE(S): Biochemical insect control : its impact on economy, environment,
              and natural selection.
PUBLICATION: Wiley, 1977.


  SUBJECT(S): Insecticides.




Not in this branch
No. of copies in other PCL libraries : E&S (1)
                        Shelved at : 668.651 QUR
------------------------------------------------------------------------------
Is this at all the sort of book you are looking for? (y/n) YES


==============================================================================
```

Fig 3.7 Full record display

```
==============================================================================
FULL DISPLAY                                                    Book 3 of 26
Search: 'insecticides and the environment'
------------------------------------------------------------------------------
  AUTHOR(S): MORIARTY F
    TITLE(S): Organochlorine insecticides : persistent organic pollutants.
PUBLICATION: Academic Press, 1975.

  SUBJECT(S): Pesticides - Environmental aspects. Chlorine organic compounds.
              Environment. Pollution by pesticides: Organic chlorine
              compounds.


Not in this branch
No. of copies in other PCL libraries : E&S (1)
                        Shelved at : 632.95042 ORG
------------------------------------------------------------------------------
Is this at all the sort of book you are looking for? (y/n) YES


==============================================================================
```

Fig 3.8 Brief display after some records have been chosen

```
================================================================================
LIST OF BOOKS                                                Books 1 to 9 of 26
Search: 'insecticides and the environment'


--------------------------------------------------------------------------------
No. Title                                Author          Classmark  Date
--------------------------------------------------------------------------------
1** Biochemical insect control : its impact.. QURAISHI M S    668.651    1977
2*  Advances in environmental science and .. (METCALF R L)   628.5      1976
3** Organochlorine insecticides : persisten.. (MORIARTY F)   632.95042  1975
THE REST OF THE BOOKS MAY NOT MATCH YOUR SEARCH VERY WELL
4   Insects, experts and the insecticide cr.. PERKINS J H    632.7      1982
5   Proceedings of the 1979 British Crop Pr.. British Crop .. 668.65     1980
6   Proceedings of the 1979 British Crop Pr.. British Crop .. 668.65     1980
7   Proceedings of the 1979 British Crop Pr.. British Crop .. 668.65     1980
8   The insecticide, herbicide, fungicide .. PAGE B G       632.95     1979
9   Organochlorine insecticides and polychl.. Standing Commi.. 628.168    1979
--------------------------------------------------------------------------------
Type its number to see if a book is relevant

Type More to look for books similar to the ones you chose


or type Down (next), Restart/new search/quit
================================================================================
```

Fig 3.9 Display during a query expansion search

```
================================================================================
LIST OF BOOKS                                                Books 1 to 9 of 26
Search: 'insecticides and the environment'


--------------------------------------------------------------------------------
No. Title                                Author          Classmark  Date
--------------------------------------------------------------------------------
1** Biochemical insect control : its impact.. QURAISHI M S    668.651    1977
2   Advances in environmental science and .. (METCALF R L)   628.5      1976
3** Organochlorine insectici ----------------------------------------------- 75
THE REST OF THE BOOKS MAY NO:Looking for more books similar to the       :
4   Insects, experts and the:2 you have chosen...                        :82
5   Proceedings of the 1979 :                                            :80
6   Proceedings of the 1979 :Found 2 more books                          :80
7   Proceedings of the 1979 :                                            :80
8   The insecticide, herbici:Type Disp to see the books                  :79
9   Organochlorine insectici:                                            :79
---------------------------:                                            :--
                           :or type Back to go back where you were before :
Type its number to see if a :                                           :
                           :                                            :
Type More to look for books :                                           :
                            ---------------------------------------------

or type Down (next), Restart/new search/quit MORE
================================================================================
```

Fig 3.10   Brief display of records found by query expansion

```
================================================================================
LIST OF BOOKS similar to the ones you chose              Books 1 to 2 of 2
(Original search: 'insecticides and the environment')

--------------------------------------------------------------------------------
No. Title                                    Author        Classmark  Date
--------------------------------------------------------------------------------
1   Organochlorine insecticides and polychl.. Standing Commi.. 628.168  1979
2   Environmental pollution by pesticides.    EDWARDS C A      632.95042 1973
    ** END OF LIST **




--------------------------------------------------------------------------------
Type its number to see if a book is relevant

Type Back to return to the books you originally found




or type Restart/new search/quit
================================================================================
```

Fig 3.11   Full display from query expansion

```
================================================================================
FULL DISPLAY of books similar to the ones you chose             Book 2 of 2
(Original search: 'insecticides and the environment')
--------------------------------------------------------------------------------
  AUTHOR(S): EDWARDS C A
   TITLE(S): Environmental pollution by pesticides.
             Environmental science research.
PUBLICATION: Plenum Press, 1973.

  SUBJECT(S): Pesticides - Environmental aspects. Chlorine organic
              compounds. Environment. Pollution by pesticides: Organic
              chlorine compounds.




Not in this branch
No. of copies in other PCL libraries : E&S (1)
                        Shelved at : 632.95042 EDW
--------------------------------------------------------------------------------
Is this at all the sort of book you are looking for? (y/n) YES


================================================================================
```

Fig 3.12 Choosing classification browsing

```
================================================================================
FULL DISPLAY                                                      Book 1 of 26
Search: 'insecticides and the environment'
--------------------------------------------------------------------------------
  AUTHOR(S): QURAISHI M S
   TITLE(S): Biochemical insect control : its impact on economy, environment,
             and natural selection.
PUBLICATION: Wiley, 1977.


 SUBJECT(S): Insecticides.



Not in this branch
No. of copies in other PCL libraries : E&S (1)
                        Shelved at : 668.651 QUR
--------------------------------------------------------------------------------
[Chosen]
                              ------------------------------------
                              :Would you like to see books shelved :
                              :near this one? (y/n) YES             :
                              ------------------------------------



================================================================================
```

Fig 3.13 Brief display in classified sequence

```
================================================================================
LIST OF BOOKS classified near 668.651
(Original search: 'insecticides and the environment')


--------------------------------------------------------------------------------
No. Classmark  Title                            Author        Date
--------------------------------------------------------------------------------
1   668.65     Pesticide chemistry : proceedings of ..  International ..  1971
2   668.65077  Pesticides : an auto-tutorial approach.  WARE G W         1975
3   668.650941 The non-agricultural use of pesticides.. Great Britain... 1974
4   668.650973 Manual of chemical methods for pestici.. United States .. 1976
5** 668.651    Biochemical insect control : its impac.. QURAISHI M S     1977
6   668.651    A critical review of the techniques fo.. BUSVINE J R      1971
7   668.9      Engineering with polymers.               POWELL P C       1983
8   669        Metals in the service of man.            STREET A         1979
9   669        An introduction to metallurgy.           COTTRELL A H     1975
--------------------------------------------------------------------------------
Type its number to see if a book is relevant

Type More to look for books similar to the one you chose
Type Back to return to your previous list of books

or type Down (next), Up (prev), Restart/new search/quit
================================================================================
```

Fig 3.14 Other options menu

```
================================================================================
LIST OF BOOKS similar to the ones you chose              Books 1 to 5 of 5
(Original search: 'insecticides and the environment')


---------------------------------  ------------------------------------------ --
No. Title                          ¦RESTART REQUESTED                          ¦te
---------------------------------  ¦                                           ¦--
1    Since 'Silent spring'.        ¦You can do :                               ¦72
2    The mutagenicity of pest¦                                                 ¦71
3    Pesticide microbiology :¦     New (start a new search)                    ¦78
4    Ecological effects of pe¦     View your list of chosen books              ¦77
5    Environmental dynamics o¦     Edit or repeat your last search             ¦75
     ** END OF LIST **         ¦        (your selections will be lost)          ¦
                               ¦                                                ¦
                               ¦Please choose one VIEW                          ¦
                               ¦                                                ¦
-------------------------------¦or type Quit if you have finished              ¦--
                               ¦     or Back to go back where you were before  ¦
Type its number to see if a    ¦                                                ¦
                               ¦                                                ¦
Type Back to return to the b¦                                                   ¦
                               ¦                                                ¦
                               ¦                                                ¦
                               -------------------------------------------------

or type Restart/new search/quit RESTART
================================================================================
```

Fig 3.15 Reviewing the list of chosen books

```
================================================================================
LIST OF BOOKS YOU HAVE CHOSEN                           Books 1 to 9 of 13
(Original search: 'insecticides and the environment')


--------------------------------------------------------------------------------
No. Title                            Author          Classmark  Date
--------------------------------------------------------------------------------
1** Biochemical insect control : its impact.. QURAISHI M S    668.651    1977
2** Organochlorine insecticides : persisten.. (MORIARTY F)    632.95042  1975
3** Environmental pollution by pesticides.    EDWARDS C A     632.95042  1973
4** Pesticides in aquatic environments.       (KHAN M A Q)    632.95042  1977
5** Environmental toxicology of pesticides.   (MATSUMURA F)   632.95042  1972
6** Since 'Silent spring'.                     GRAHAM F        632.95042  1972
7** Persistent pesticides in the environment. EDWARDS C A     632.95042  1975
8** A critical review of the techniques for.. BUSVINE J R     668.651    1971
9** Ecological effects of pesticides.         (PERRING F H)   632.95042  1977
--------------------------------------------------------------------------------

Type its number to see further details of one book

Type Back to go back to where you were before viewing the 'chosen' list

or type Down (next), Restart/new search/quit
================================================================================
```

Fig 3.16 Input screen for second or subsequent search

```
===============================================================================
                              SUBJECT SEARCH                        ** OKAPI



      -------------------------------------------------------------------
      ¦ Previous search(es)                              Books found ¦
      -------------------------------------------------------------------
      ¦ 1  'safety in the workplace'  . . . . . . . . . . . . . . .    31 ¦
      ¦ 2  'factories act'  . . . . . . . . . . . . . . . . . . .     7 ¦
      ¦ 3  'workshop practice'  . . . . . . . . . . . . . . . . .   137 ¦
      -------------------------------------------------------------------


Type a word or a phrase which describes the books you want :


      -------------------------------------------------------------------
      ¦ █                                                                 ¦
      -------------------------------------------------------------------




===============================================================================
```