a guide to

PERSONAL INDEXES

a guide to PERSONAL INDEXES

using edge-notched and peek-a-boo cards

by

A C FOSKETT FLA

SENIOR LECTURER, COLLEGE OF LIBRARIANSHIP WALES

CLIVE BINGLEY LONDON

FIRST PUBLISHED 1967 BY CLIVE BINGLEY LTD 16 PEMBRIDGE ROAD LONDON W11 · SET IN 11 ON 13 POINT BASKERVILLE AND PRINTED IN GREAT BRITAIN BY THE CENTRAL PRESS (ABERDEEN) LTD ·
85157 001 1

CONTENTS

ILLUSTRATIONS

INTRODUCTION

IT IS NOW COMMON practice for research workers to keep a note of periodical articles, pamphlets and reports which they have found interesting, so that they can refer back if necessary. Once such a file begins to grow it becomes tedious to look through every reference each time a search is made, and some form of organisation has to be introduced. A conventional card file requires a card for each method of approach: one or more for the author(s) and one for each subject. There is also the problem of choosing a suitable system of subject headings or classification to arrange the subject cards. To avoid these difficulties many people turn to more recent forms of index, for example edge-notched cards or ' peek-a-boo ' systems, only to find that these methods present their own problems. This book is an attempt to show in simple terms how to maintain a personal index on either edge-notched cards or peek-a-boo. The intention is to give practical advice without going deeply into theory, though for those who wish to pursue this side some further readings are given. The theory of subject approach can be highly sophisticated, but for the personal index, or even for the file in a small information service, the size of the file is not great enough to present many acute problems.

Edge-notched and peek-a-boo cards are available from a number of suppliers, in different sizes and printings. The principles of use remain the same, however, and the methods outlined in this book may be applied to any of these products. The major point of difference lies in the size of the cards, and care must be taken to choose a large enough card to begin with, while at the same time avoiding the storage problems arising with an unnecessarily large card.

Finally it should be realised that personal indexes, however useful, can only cover a small proportion of the literature. They will reveal articles which the compiler has himself already seen, but for more thorough searching it is essential to turn to published abstracts and indexes which have a far wider coverage.

A C FOSKETT

PART ONE: EDGE-NOTCHED CARDS

EDGE-NOTCHED (or edge-punched) cards consist of thin cards with holes punched round the edges. They may vary in size from small, *eg* 4 in by $2\frac{1}{2}$ in, to relatively large, say 12 in by 9 in, and may have holes along all four edges or perhaps only one. The spacing of the holes may vary, but there will usually be four or five holes to the inch. A popular size of card is 6 in by 4 in with 81 holes, 75 of them numbered; another size often found is 8 in by 5 in, with 134 holes all told. It is best to avoid using holes too near the corners, so these two cards have respectively 75 and 128 useful punching positions. It is possible to have cards specially printed to suit one's own requirements, but this is rather expensive, and most users will find that the basic card supplied by most manufacturers, with holes numbered in one sequence, is quite satisfactory. It is also possible to buy cards completely unprinted and duplicate them oneself, but this can be rather tedious and give unsatisfactory results in unskilled hands. A typical card is shown in figure 1.

In use, the cards are notched at the required positions using a special punch, so that those holes are now open to the edge of the card. To sort out the cards which have been notched at a particular hole, a needle is passed through the pack at that hole and the pack is shaken; cards which have been notched will drop out, leaving the rest still on the needle. The cards do not have to be kept in any order and after use cards are simply replaced with the rest of the pack.

The equipment required is very simple, consisting only of a punch and a needle to manipulate the cards and some sort of container to store them. This may be a cardboard box for

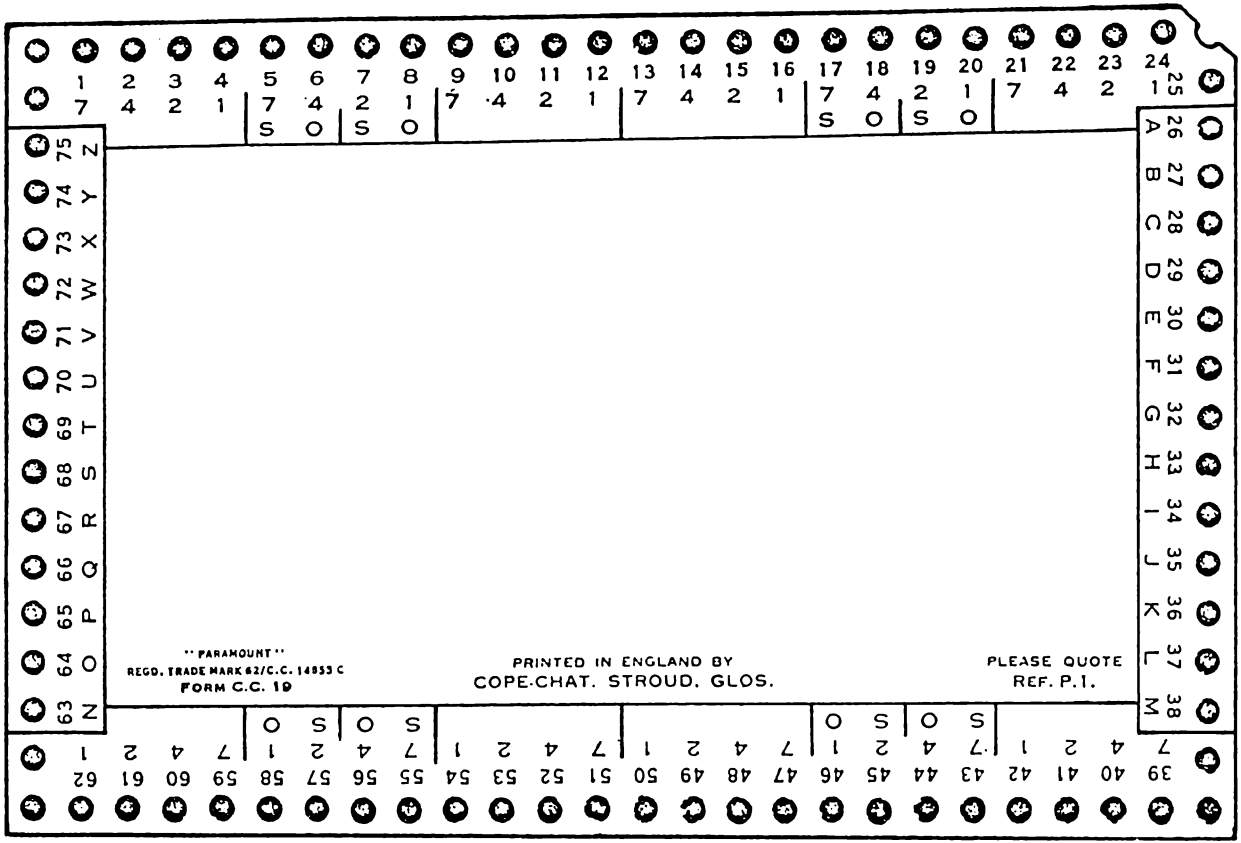

FIGURE 1: A typical edge-notched card (illustration by courtesy of Copeland-Chatterson Co)

a limited number of cards; at the other extreme is a lockable steel cabinet, which may be necessary for a large collection of cards relating to material which is to be kept private.

The information which it is desired to record is entered on the body of the card. There is plenty of room for author and title of an article, the bibliographical reference, and an abstract if required. The back of the card can be used as well as the front, but it is more convenient to keep to one side of the card if possible. This information has now to be translated into notches round the card if it is to be retrieved; this process is known as *coding*.

CODING

If each hole represents one specific item, then we can code as many items as there are holes: say 75. To retrieve an item we should only need to search that one hole. This is known as direct coding, and has the advantages of speed and precision; we can go straight to the item we want. However, 75 is not a very large number if we are thinking in terms of authors or subjects, so for many purposes it is necessary to resort to 'indirect' coding. Here, the holes are allocated in groups of, for example, four, five or six, each group being known as a *field*. In each field there are a number of different methods of coding, each having its own advantages and disadvantages. All of them require the items to be coded to be expressed in terms of numbers, so authors' names and subject terms must be converted in some way. (It is possible to code letters, but very wasteful of space.)

The simplest method of indirect coding is the 7-4-2-1 system. It requires a four-hole field for each digit, the four holes being labelled 7, 4, 2 and 1; to code from 0 to 9 we punch as follows:

0 — no punching	5 — 4 and 1
1 — 1	6 — 4 and 2
2 — 2	7 — 7
3 — 2 and 1	8 — 7 and 1
4 — 4	9 — 7 and 2

To code numbers up to 999 we should need three four-hole fields, one for units, one for tens and one for hundreds. If we arrange them so that the hundreds field is at the left, tens in the middle and units at the right, we can sort cards into numerical order very simply. First needle the 1 hole in the units field, then the 2, then the 4, and so on until we finally needle the 7 hole in the hundreds field; each time the cards that fall are placed at the rear of the pack, taking care that they do not get out of order. With the final sort it will be found that the cards are in order.

7-4-2-1 coding is the most economical for most purposes, but it has a disadvantage. If we have punched a card for, say, 853, it will drop if we search for 853, but it will also drop if we search for 852, 851, 843, 842, 841, 753, 742, 53, 41 etc. Conversely, if we search the pack for, say, 421, we shall drop 633, 521, 432 and so on if they are present. Simple needling for, say, 700 will give us not only 700 800 and 900 but in fact every number from 700 to 999! These unwanted possibilities are known as false drops. Often they will not present any particular problem, since it will be possible to sort finally by hand. If we want to find 700 we can needle the hundreds 7, then by inspection find the most frequently notched holes in the tens and units field and needle these, keeping this time the cards which do *not* fall, until we have reduced the number of cards to a batch which can be handsorted. However, the problem must be borne in mind.

We can solve it by using six-hole fields instead of four; the holes are now labelled 7, 4, 2, 1, S and 0, and to code from 0 to 9 we punch as follows:

0 — 0	5 — 4 and 1
1 — 1 and S	6 — 4 and 2
2 — 2 and S	7 — 7 and S
3 — 2 and 1	8 — 7 and 1
4 — 4 and S	9 — 7 and 2

We can still sort into numerical order by needling the 7-4-2-1 holes as before, but we can now also sort unambiguously for any number we want without the problem of false

drops. The cost of this positive searching is that we now use six holes instead of four in each field, so eighteen holes are needed to code from 0 to 999 instead of twelve, and we have to needle twice for every number instead of sometimes only needing to sort once.

Another commonly used method of coding is the pyramid, shown in figure 2. Each number from 0 to 9 is coded by punching two holes; provided that the numbers are entered in the pyramid as shown, cards can be sorted into numerical order by needling across from right to left, placing cards which fall at the back of the pack as usual. We have to sort on two holes each time we want a particular number, but there is no possibility of false drops. The pyramid is actually a special case of two-hole coding, described later.

Binary coding may be used. In this method, holes are labelled 1, 2, 4, 8, 16, 32, 64, 128, 256, etc, to whatever limit is required. Thus to code numbers up to 1023 would require ten holes, as shown in figure 2e. To code, say, 853, find the largest number which will go into it: in this case 512. Punch this hole and subtract the number from 853, leaving 341. Take the next number which will go into this, here 256, and repeat the operation, leaving 85. 64 is the next highest number to go into the remainder, so this is punched; remainder now 21, giving 16 as the next hole to be punched. 5 is left, requiring 4 and 1 to be punched out. So 853 is coded by punching the holes for 512, 256, 64, 16, 4 and 1. Binary coding becomes very economical for higher numbers; for example, where 7-4-2-1 coding would require 16 holes (four four-hole fields) to code from 0 to 9999, binary coding will code from 0 to 65,535 in the same space ($\sum_{n=15}^{n=0} 2^n$). However, there will be a large number of false drops, and though it is possible to sort into numerical order in the usual way, binary coding is rarely used. It is mentioned because it is the method of coding which is used in computers and is therefore becoming familiar to many people.

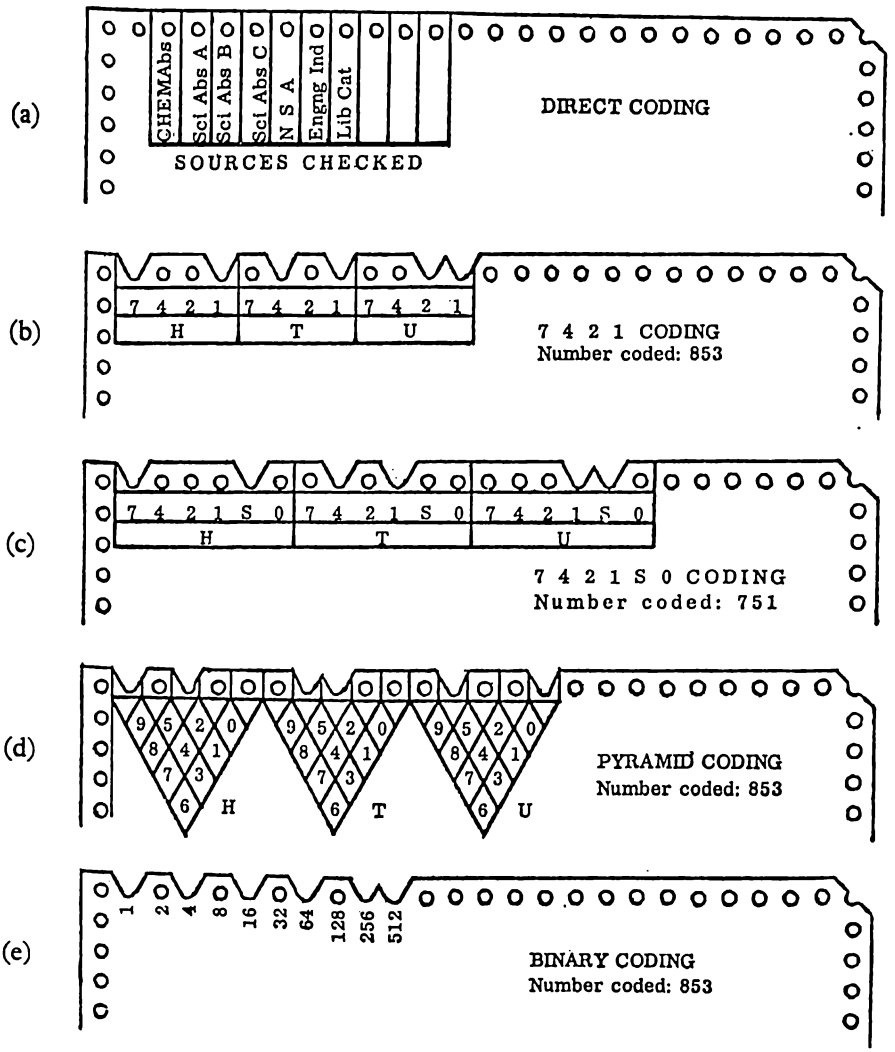

FIGURE 2: Various methods of coding

The methods of coding discussed so far have been suitable for certain kinds of information; for example, authors' names, year of publication and so on. It is not possible to punch more than one number into any particular field, because of the very rapid increase in false drops that occurs; for example, using three fields of 7-4-2-1 coding, it is possible to code numbers up to 999, but if one were to punch out, say, 742, 417, 221 and 174, the card would become virtually useless. If we are to be able to code subjects, we may need to punch several codes for each document, so we must have a method of coding which permits us to do this. We might use a different group of fields for each subject term; in this way we could have a set of up to a thousand terms, requiring three 7-4-2-1 fields for each, and code up to six descriptors on each card, 72 holes in all. This would leave no room on a 75-hole card for anything else, and might well be inadequate, since we may often need to code more than six descriptors.

A more satisfactory method of working is known as 'superimposed random coding'. In this method, we take a large field, say 40 holes. By direct coding we could cater for forty descriptors and we could punch out any or all of them on the same field at the same time. Forty is unlikely to be enough, but if we use two holes for each we can now code 780 terms in the 40-hole field. This figure is obtained quite simply: the first hole can be selected in forty different ways, and the second can now be selected in thirty-nine different ways; however, each pair of numbers will duplicate another pair (24 and 37 is identical with 37 and 24) so we have to divide by two. We can punch out more than one pair in the field at the same time, but at the risk of introducing false drops; for example, if we punch out 24 and 37 for one code, and 7 and 18 for another, we will drop that card when we are searching for a term coded 18 and 37, or 7 and 24. In order to keep the number of false drops to a reasonable level the number of codes punched into the field has to be limited; for a forty-hole field, using two-hole codes, eighteen holes (*ie* nine codes) is the useful maximum.

We can punch more than two holes for each code. This will increase the potential number of codes, but will also involve needling more than twice for each code. For example, if we use three-hole punching, our forty-hole field will give 9,880 possible codes ($40 \times 39 \times 38 \div 6$); however, we shall have to sort on three holes for each code, and it will be unwise to punch more than six codes at a time.

A PRACTICAL FILE

Having looked at general principles of coding we can turn to the more particular circumstances of a personal file. To maintain a useful file we should be able to approach the articles indexed through the author, date, source and subject, and we must therefore code all of these items on our card. The methods proposed here should be found to be widely applicable, but there may be special circumstances which call for a different approach, for example in dealing with large numbers of chemical compounds. The important thing is to avoid needless elaboration; it is very easy to try to make an index perform tasks which are more efficiently done by consulting general or special reference books.

Author: As has already been pointed out, it is usual to convert names into numbers for coding purposes. It would be possible to code the first letter of the author's name directly, using twenty-six holes; or to use six holes with a coding similar to the 7-4-2-1. It is more effective however to use a conversion table in which one can look up the name and find a three-digit equivalent. Such a table is given in appendix one (page 61), and the name to be coded is found in this and the desired code punched into the card. The first twelve holes of the card should be reserved for this purpose to give the three four-hole fields required. The coding should be 7-4-2-1; this will mean that the file can be sorted into alphabetical order of author if need be, and though there will be the risk of false drops when searching for a particular author, this will not be a very serious problem in practice. Some examples of coding are given in the appendix.

Because the risk of false drops increases unacceptably if we code more than one name in this way, we can only find an article through one author. This will not normally be a problem, as it is still common to find single authorship, but in some fields of science and technology *multiple authorship* is becoming quite usual. In such cases it will be necessary to find the most significant author of the group, often the first. An alternative is to code the name of the parent body, and this solution is recommended for technical report literature, where the firm or government department carrying out the project is usually more significant than the individuals concerned.

If it is essential to be able to find an article through more than one author, it is possible to make out additional cards. All that is necessary is to put the name of the secondary author followed by the word *see* and the name of the principal author. The name of the secondary author is coded but nothing else. When the file is searched, these extra cards will fall if the secondary author's name is sought, but not at any other time. It will not normally be necessary to make out these additional cards, but the possibility exists. To try to provide space on the card for more than one author to be coded would be very wasteful of coding space, since most documents are still the product of a single author.

Source: The next four holes can be used to indicate the kind of document involved: for example, book 13 and 14; periodical 13 and 15; pamphlet 13 and 16; patent 14 and 15; technical report 14 and 16; others 15 and 16. In some subject fields it would be desirable to modify this outline; for example in the social sciences it would be pointless to have a coding for ' patents ', so this combination (14 and 15) could be used for some other kind of source, *eg* statistics, film strips.

When the source is a periodical it is often useful to be able to trace it through the title, since one can often remember where one saw an article without being able to recall the author's name. There are two different approaches to the coding of periodical titles which might usefully be

adopted for a personal file; the choice will depend to a certain extent on the existence of a convenient list, for example a list of periodicals available within one's own organisation. Where such a list exists it is simple to go through it numbering the titles; to code a title we simply punch the number corresponding to that in the list. If the list contains less than 100 titles we need two fields; three fields if it is below 1,000 but above 100. It will be possible to use 7-4-2-1 coding, and we may find that we do not need to use all four for the hundreds field; if our list contains 500 titles we need only have 4-2-1 in the hundreds field, as we shall never need to use the 7. We shall only be noting a selection of articles, and experience shows that this will largely come from a relatively small number of titles. Since these will be scattered in the list, almost certainly, the problem of false drops will not be acute.

The alternative is to compile our own list as we make a note of articles that we wish to record. For example, if the article we note first comes from *American documentation,* this will be coded 1. The second title, *eg Journal of documentation,* will be 2; the third, *eg Journal of the Association for Computing Machinery,* will be 3; and so on. Since the list of titles will be growing all the time, it will be necessary to keep a record of both the titles and their codings and the number of the past coding allocated, and this record will have to be on cards so that it can be kept in alphabetical order; otherwise it will be difficult to find the coding for a given title. It could consist of ordinary catalogue cards (12·5 cm by 7·5 cm), each bearing the title of a periodical and its number; at the front of the file is another card on which is noted each number as it is allocated. We are unlikely to need more than two fields to code titles if we use this method, as most of our references will probably be to half a dozen journals. To set against this, we may find the problem of false drops more acute, as every code from 1 to the highest will be used; it will probably be worth using two pyramids rather than two 7-4-2-1 fields, as this eliminates the risk of false

drops at the cost of two extra holes. Here again we have to remember that two fields, enabling us to code up to 99, will almost certainly be enough if we list the titles ourselves, but three fields may be needed if we use an already existing list, containing several hundred titles of which only a selection are of interest to us.

A further consideration could arise if many of our references are to report literature. If we allocate three fields of 7-4-2-1 coding to the title when the reference is a periodical, it is possible to use these same three fields to code the name of the issuing body when the reference is to a report. Provided we sort our cards first by kind of source material before searching for a particular periodical title or report issuing body, there will be no confusion. Even this preliminary sort may not be essential if we are prepared to handsort in order to remove unwanted cards, as there will probably be very few of these. Figure 3 shows how the two purposes may be combined.

It will not be necessary to code such information as report or patent numbers. These are primarily filing devices rather than sources of information, and in the context of a personal file it will be very rare indeed that one will know a report number yet know nothing else about it. In a small information service it will be necessary to keep a separate file of cards in report (or patent) number order showing number of copies held, and any restrictions on availability imposed by security considerations, but this is essentially a library record and not important for a personal index.

The coding suggested so far, for author, source and title, has taken up to seven four- or five-hole fields. We also need to code the date of publication, as this is a useful factor to be able to search by. For most purposes it will not be necessary to show the decade; the year alone will do, since we can quickly handsort to remove unwanted cards. If it is found that a high proportion of references fall outside the current decade, it will prove useful to give a more detailed coding; this could be one extra hole coded 'non-current decade' or

it could be a whole field so that the decade can be coded exactly. It is important not to allocate unnecessary space to this, as every hole used for coding dates cannot be used for

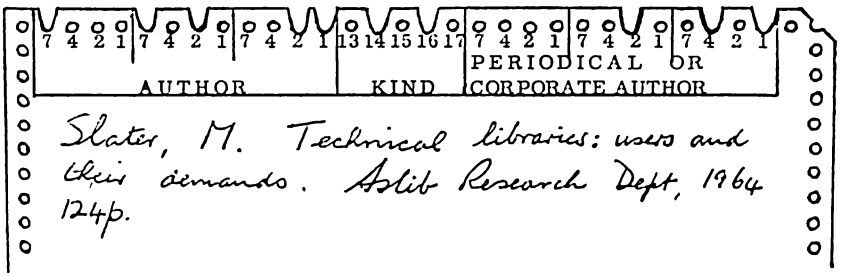

FIGURE 3(a)

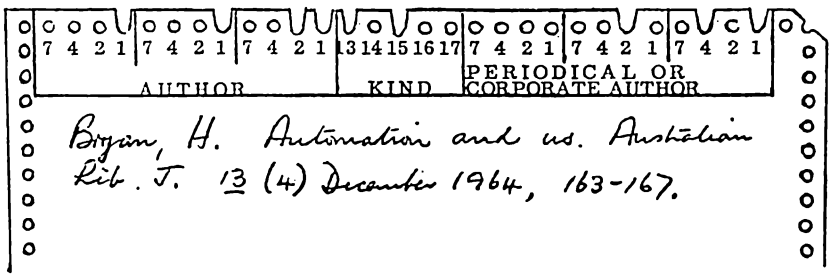

FIGURE 3(b)

In figure 3(a) holes 14 and 16 indicate 'report', and the coding in the next section 025 indicates a corporate author beginning ASL (in this case Aslib). In figure 3(b) 13 and 15 indicate 'periodical' and the coding in the next section therefore indicates periodical number 25 in the list being used: in this case, *Australian library journal*. Provided that the field for 'kind' is needled first, there will be no confusion between the two uses of the same fields for coding two different items

coding subjects, potentially a far more significant search factor.

Having used 35 holes we must now consider how large a card we need to be able to code for subject. With a 75-hole card we have a maximum of 40 holes left; as already shown, this is enough to code up to 780 subject terms using a two-hole coding, and allows us to use nine or ten subject terms

per document without getting too many false drops. This could well be adequate for a small collection (say a few hundred at the most) in a fairly limited field, but might prove too cramped for a larger collection covering wider subject areas and thus requiring more terms. It is possible to keep the number of subject terms low by using the principle of 'minimum vocabulary' but this requires a great deal of effort to be devoted to its compilation if it is to be of value, and it is not usually possible to spend much time on a personal index, quite apart from the special skill required. It is simpler to use a larger card.

A card with 128 holes will give plenty of room for the items specified so far, while leaving 100 holes for subject coding. In this field we can code up to twenty or more subjects per document without raising the number of false drops to an unacceptable figure, and there will be room for some 5,000 different two-hole codes: $(93 \times 92 \div 2)$. This is likely to be enough for quite a large collection of documents, even if they cover a number of subject areas. Once again a choice has to be made which depends on the precise situation to be handled, and some guidance is given in figure 4, which shows how cards of the two sizes mentioned might be laid out. A method of obtaining the requisite random codings is given in appendix two.

Subjects: It would be possible to devote a large book to the problems of subject approach; indeed, a great many authors have already done so, with varying success. Once again, in this book no attempt will be made to go deeply into theory; instead, practical suggestions will be given. However, it may be useful to introduce some theoretical ideas to justify the suggestions made, in particular the ideas of recall and relevance.

Let us assume the existence of a set of documents to which we pose a question. We could discover all the information in the collection relevant to our question by reading every document in it, but if we did so we should at the same time have been obliged to look at a lot of items that were not relevant.

Instead of looking through the whole collection each time we wish to make a search, we describe the documents in

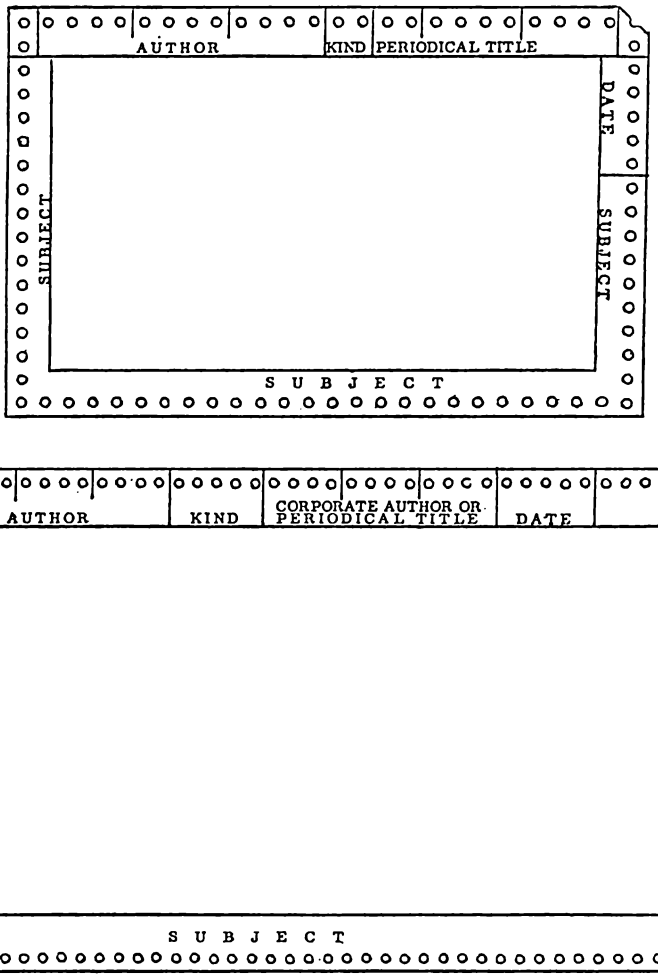

FIGURE 4: A comparison of possible lay-outs for 75- and 128-hole cards

terms of subjects; now we only need to look at those documents of which the description corresponds to a similar description of the subject we are interested in. However, the transfer of information is not a precise process; authors write

from one point of view, readers study their works from a different point of view. So our subject descriptions of documents for our information file will never correspond exactly with the descriptions that we might wish to use when searching. In consequence, when we put a question to our file, we shall find a proportion of the relevant documents, but not all; and we shall at the same time turn up a number of documents which for one reason or another are not what we want. If the total number of relevant documents in the file is N, the number of documents we find R and the number of those we find that are actually relevant is P, then we have two ratios:

$$\text{recall ratio} = \frac{P}{N} \times 100\ (\%) \qquad \begin{matrix}\text{relevance ratio} \\ \text{(or precision ratio)}\end{matrix} = \frac{P}{R} \times 100\ (\%)$$

It has been shown that if we try to improve the recall ratio we do so at the cost of a decrease in the relevance ratio; and if we try to improve the relevance ratio we shall do so at the cost of decreased recall.

A high proportion of the devices that are associated with the subject approach, for example classification schemes, thesauri and so on, are intended to improve recall; the rest, for example roles and interfixes, are aimed at higher relevance. However, in the context of a relatively small personal index or one used in a small information service, it appears likely that the improvements to be gained from such devices are not as great as is often claimed, and that it is possible to manage without them if one is willing to accept that results may be less than perfect. It must always be remembered that we are unlikely to search such a file unless we are fairly sure that we have put something in to start with.

What effect does this theoretical discussion have on the practical approach? The main point is that it shows that we can rely for our subject terms quite largely on the documents themselves, with a minimum of control. It is best to use only one form of the same word, or rather in the case of edge-notched cards to use the same coding for all forms: thus

anneal, annealed and *annealing* might all be given the same coding.

It may well be helpful to consult a classification scheme or thesaurus when trying to think of subject terms, since they will show related terms which ought to be considered. However, most of these schemes are rather expensive if one is thinking of a personal file, and it may not always be convenient to consult a library copy. If no scheme is available, the document should be studied carefully and any significant words underlined or noted on a sheet of paper. Any other terms that come to mind as likely to search terms should be noted at the same time. The useful limit on the number of terms imposed by the amount of space available on the card for coding must be borne in mind. When the terms to be used have been decided, the document is entered onto a punched card. The same order of the details should be followed each time: for example, author, title, bibliographical citation, then list of subject terms, and finally an abstract if necessary.

If a scheme is available, for example the UDC (Universal Decimal Classification) or EJC Thesaurus, after reading through the document look up the likely terms in the scheme. In UDC consult the index, which should lead to the relevant part of the schedules (list of terms in systematic order); then consider whether the terms chosen are the most useful or whether it looks as though the wrong place has been chosen. The scheme will also show related terms, either by simple juxtaposition or by the use of, *eg*, bold type face for broader headings. The schedules and index should help to clarify any uncrystallised ideas and suggest perhaps better or additional terms. With a thesaurus the same procedure is followed, except that the terms are arranged in alphabetical order instead of systematic, so related terms have to be listed under each term instead of being shown by the schedules. There is thus no need to refer from an index to the schedules, but on the other hand there is less systematic linking of related subjects.

To establish the coding for a given term the first time it is used, look up a pair of random numbers (see appendix two). Make out a card bearing this term and the pair of numbers, and another card bearing the numbers, always citing the lower first. The first card will be filed alphabetically so that the same coding will be used again the next time that term is wanted, and so that it will be possible to search the file for that term. The second card should be filed numerically so that it will show that that particular coding has been used. Both these files can be on a sheet of paper rather than on cards if the file is a small one, but cards do enable them to be kept completely up-to-date and in order. If a classification scheme or thesaurus is being used as an aid, it is possible to enter the coding in the scheme, but it will still be necessary to keep some note of the codes that have been used.

Within the limits of the coding, the subject terms should be chosen to give as precise a description of the subject content of the document as possible. The more material on a subject put in a system, the more exact the indexing must be if we are to be able to find items when we want to. It is often suggested that 'fringe' subjects—those of only limited interest to us—can be indexed quite broadly, if we are unfamiliar with the terminology, for example. However, precise indexing cannot make searching more difficult and may make it much easier, provided that one follows a sensible search procedure.

The advice to be precise means that if a document deals with, say, *annealing of tinplate,* we punch out the appropriate codes for *annealing* and *tinplate.* Another document dealing with *heat treatment of tinplate* would be coded for *heat treatment* and *tinplate.* If we later have occasion to search for all the documents we have on various forms of heat treatment, we shall have to needle not only the code for *heat treatment* but also that for *annealing.* The use of a thesaurus or classification scheme will help considerably here, and if we do not use one it will be necessary to link related subjects in our own list. This can be done quite simply by

anneal, annealed and *annealing* might all be given the same coding.

It may well be helpful to consult a classification scheme or thesaurus when trying to think of subject terms, since they will show related terms which ought to be considered. However, most of these schemes are rather expensive if one is thinking of a personal file, and it may not always be convenient to consult a library copy. If no scheme is available, the document should be studied carefully and any significant words underlined or noted on a sheet of paper. Any other terms that come to mind as likely to search terms should be noted at the same time. The useful limit on the number of terms imposed by the amount of space available on the card for coding must be borne in mind. When the terms to be used have been decided, the document is entered onto a punched card. The same order of the details should be followed each time: for example, author, title, bibliographical citation, then list of subject terms, and finally an abstract if necessary.

If a scheme is available, for example the UDC (Universal Decimal Classification) or EJC Thesaurus, after reading through the document look up the likely terms in the scheme. In UDC consult the index, which should lead to the relevant part of the schedules (list of terms in systematic order); then consider whether the terms chosen are the most useful or whether it looks as though the wrong place has been chosen. The scheme will also show related terms, either by simple juxtaposition or by the use of, *eg,* bold type face for broader headings. The schedules and index should help to clarify any uncrystallised ideas and suggest perhaps better or additional terms. With a thesaurus the same procedure is followed, except that the terms are arranged in alphabetical order instead of systematic, so related terms have to be listed under each term instead of being shown by the schedules. There is thus no need to refer from an index to the schedules, but on the other hand there is less systematic linking of related subjects.

To establish the coding for a given term the first time it is used, look up a pair of random numbers (see appendix two). Make out a card bearing this term and the pair of numbers, and another card bearing the numbers, always citing the lower first. The first card will be filed alphabetically so that the same coding will be used again the next time that term is wanted, and so that it will be possible to search the file for that term. The second card should be filed numerically so that it will show that that particular coding has been used. Both these files can be on a sheet of paper rather than on cards if the file is a small one, but cards do enable them to be kept completely up-to-date and in order. If a classification scheme or thesaurus is being used as an aid, it is possible to enter the coding in the scheme, but it will still be necessary to keep some note of the codes that have been used.

Within the limits of the coding, the subject terms should be chosen to give as precise a description of the subject content of the document as possible. The more material on a subject put in a system, the more exact the indexing must be if we are to be able to find items when we want to. It is often suggested that 'fringe' subjects—those of only limited interest to us—can be indexed quite broadly, if we are unfamiliar with the terminology, for example. However, precise indexing cannot make searching more difficult and may make it much easier, provided that one follows a sensible search procedure.

The advice to be precise means that if a document deals with, say, *annealing of tinplate,* we punch out the appropriate codes for *annealing* and *tinplate.* Another document dealing with *heat treatment of tinplate* would be coded for *heat treatment* and *tinplate.* If we later have occasion to search for all the documents we have on various forms of heat treatment, we shall have to needle not only the code for *heat treatment* but also that for *annealing.* The use of a thesaurus or classification scheme will help considerably here, and if we do not use one it will be necessary to link related subjects in our own list. This can be done quite simply by

noting them on the card showing the coding. Such a card might read:

Heat treatment 58, 97
related subjects:
annealing
quenching

Similar cards would be in existence for *annealing* and *quenching*. The EJC Thesaurus shows clearly how such links may be made for broader terms, narrower terms and related terms.

It is best to code single terms or concepts rather than composite terms. Thus in the previous paragraph it was suggested that *annealing of tinplate* should be coded for the two separate terms *annealing* and *tinplate* rather than the composite idea as it stands. It is possible to use such 'pre-coordinated' composite subjects, but this can lead to difficulties and it requires rather elaborate indexing if some of the terms are not to get 'lost'. The use of single terms is preferable for the personal index. Occasionally it may be necessary to use more than one word: for example, it would be pointless to split *electrical engineering* and *Great Britain* into their components.

Synonyms may present a problem. It is often found that there is a difference between the terminology used in the United States and that used in Great Britain, even in science and technology. Similar problems can arise from the use of trade names, *eg* polythene instead of polyethylene; or 'Mylar' (US) and 'Melinex' (GB) instead of polyethylene terephthalate. In the social sciences the problem is much worse, because many important concepts, such as 'democracy', are not clearly defined and their meaning may depend on the author's point of view. It is usual to avoid the use of synonyms in library catalogues by choosing one term and using only that; synonymous terms have a reference to the chosen term. For example, we may decide to use the term *mumps* rather than *parotitis,* or vice versa. However, it should be evident that we are in fact using synonyms all the

time, since we are using codes to represent terms, *ie* to act as synonyms. There is therefore no reason for avoiding the use of synonyms in our alphabetical guide to codes used, provided that we give them the same coding, for example:

mumps 43, 76
parotitis 43, 76

If we use more than one word for the same idea, for example if we depend too heavily on the words used in the document we are indexing, all that will happen is that we shall decrease recall when we search the file later on. Using the same example, if *mumps* is coded 43, 76 and *parotitis* 54, 62, then if we search for *mumps* we shall not drop the cards on which we have coded *parotitis,* and of course the converse is true. This could be a disadvantage, particularly if we want to find everything. However, it might be advantageous; if the articles using the word *mumps* and thus coded for that term are 'popular' in their approach, while those using *parotitis* are more technical, then we might want to find one set of documents at one time and not the other. So not merging synonyms will decrease recall, but may improve relevance. This is generally a desirable state of affairs; as has already been pointed out, in a personal index or a one-man information service, memory can play a useful part in reminding us that we put something into the system and therefore ought to be able to get it out.

Homographs may be a problem, though in a personal index we may not find this serious. Such words as 'pitch' are ambiguous unless given a context, but the fact that the index is being compiled for a particular purpose will often provide such a context. Again, failure to distinguish homographs will not prove disastrous, but will merely lead to a decrease in relevance—if we are interested in the slope of an aircraft's wings we will probably not wish to read about a black tarry substance—but this will not be a major difficulty, as we shall be able to handsort the unwanted cards quickly.

It may be useful to repeat here the advice to enter subject terms by which a document is coded on the punched card

just after the bibliographical citation. If this is not done, it becomes much more difficult to remove cards which, for one reason or another, have fallen during a search but are not wanted. If the subjects are entered on the card in a standard position every time, it becomes much simpler to look through a batch of cards to find the required ones.

MANIPULATING THE FILE

As we start building up our file of references so we shall start wanting to make use of it. An edge-notched card file is very simple to use, but there are various points which are worth noting.

The top right hand corner of each card is removed, showing which way up the cards should go. If the pack of cards gets mixed up, for example if they are dropped, gather the pack together and put them straight, then needle through the top right-hand corner. Cards which are the right way up will fall; any which are upside down or back to front will remain on the needle; remove them and turn them through 180 degrees so that what was the bottom left-hand corner is now at the top right. Needle again through the top right-hand corner; again cards which are now correctly positioned will drop, leaving on the needle only those which are back to front. They should be removed from the needle and turned round, and once again needled through the top right-hand corner, cards which fall being added to the previous two sets; those remaining should be turned through 180 degrees and will now be correctly positioned.

When searching the pack, do not try to needle more than about an inch and a half of cards at the time, otherwise it is highly probable that not all the cards that ought to fall will do so. After passing the needle through the appropriate hole, fan the cards out so that they hang loosely on the needle, then lift. When the required cards have dropped, close up the remainder of the pack and inspect it to see that all the cards punched for that hole have dropped; any that have not fallen will show up clearly as a gap above the needle. To

help in getting the pack straight for each needling, it is possible to make a simple board with a stop on it at the right-hand edge; the cards are butted against this stop to get them into position.

To find what documents we have indexed by a particular author, we look up the name in the coding table. As an example, let us take *J Russell,* coded 753. First needle the 7 hole in the hundreds field; this will drop all cards coded with numbers from 700 to 999. Next needle the tens field, first on 4 and then on 1. In each case cards which do not drop are put back in the container, so we now have all cards coded 750-759, 850-859 and 950-959. Needle the units field, first the 2, then the 1; now we have all cards coded 753, 853 and 953. The false drops may now be removed by inspection, or we can needle the 2 and 1 in the hundreds field, this time *retaining* the cards which remain on the needle.

Suppose the name wanted had been *Hewitt,* coded 400. By needling the 4 hole in the hundreds field we shall obtain any cards coded 400, but also any cards coded 401 to 699. Take this pack of cards and by inspecting the top it will be possible to see which of the other holes is most frequently punched. Needle this hole, retaining the cards which do not fall; repeat this until the set of cards remaining on the needle is small enough to be conveniently handsorted. Practice will soon indicate short cuts for eliminating false drops in similar circumstances.

The procedure for sorting the whole pack into alphabetical order has already been described in the section on coding. Some modifications may be necessary if the pack is larger than can be handled at one time. Needle the 7 hole in the hundreds field through the whole pack, putting all the cards that fall into one pile. Then needle the remaining cards on the 4 hole in this field, putting all these cards into a second pile. There will now be three packs: the cards still remaining on the needle, including 001 to 399; 400 to 699; and 700 to 999. If these three sets are manageable, they can be sorted in the usual way, beginning at the right-hand end (1 in the

units field) and working across to the left. There will of course be no need to needle beyond the 2 hole in the hundreds field, as all the higher cards have been removed already, or else already sorted. If the three packs are still too large it will be necessary to sort them further into individual hundreds. If they are still too large it would probably be useful to start reducing the size of the pack, for example by weeding out some of the older references; but this stage will only be reached after some considerable time in most personal indexes.

To find an article when we only know that it was published in a particular periodical, we must first sort out the cards for periodical references, by needling 13 and 15 in the example given earlier. The code for that periodical is now found and the pack sorted for this, using the kind of technique described for searching by author. If we are numbering periodical titles as they come, rather than working from an existing list, we may well have used pyramid coding for this, in which case there will be no problem with false drops. Whichever method of coding we have used we can sort into numerical order; this will not be helpful if we are compiling our own list of periodicals, but if we are simply numbering an existing list, then numerical order will correspond to alphabetical order of title and this could be useful.

If we are searching for reports issued by a particular organisation, and have used the method suggested earlier of coding these names in the space used for periodical titles, we must again first sort out the cards relating to the right kind of material, in this case by needling 14 and 16. We can then look up the coding for the name and search for this in the usual way. Note that if we do not make the preliminary sort by kind of material, we shall still get the cards relevant to our search, but we run a considerable risk of getting a lot of unwanted cards which refer to periodicals with the same coding. We can of course sort the other way round; first by the code for the name wanted, then by kind of material to eliminate the unwanted periodical references; here again

the best way to carry out the overall operation will depend on factors such as the relative numbers of references to reports and to periodicals, and practice will soon show which is the most satisfactory way in any given situation.

To search by date is a straightforward operation and requires no comment, except to point out that we can arrange our file in chronological order if we wish by sorting across the pyramid in the way described. Such a sort may well be useful from time to time for weeding purposes; it helps us to remove older items from the current file. These older references may be kept in a separate file which we can consult if need be, while their removal from the current file will make this much more manageable.

It is searching by subject which is likely to cause most problems, for the reasons already outlined. We must first translate our enquiry into the terms we have used in putting material into the file, and here our list of subject terms on cards will prove invaluable. Having decided on the terms we are going to search for, we must find the appropriate codes. Let us assume that we are searching for *annealing of aluminium,* being fairly sure that there is something in the file. We find that the codes are *annealing* 42, 98; *aluminium* 65, 79. To find the composite subject we want, we must sort out the cards punched for holes 42, 65, 79 and 98. A quick inspection will show which of these holes has fewest punches in the pack of cards; this is the hole to sort first. The order does not matter as far as the end result is concerned, but the fewer cards we have to handle for any sort, the quicker it will be, so it is useful to reduce the number by the largest possible amount at each sort, *ie* to needle the least-punched hole.

If our search is successful, we shall probably not want to go further, but if we find nothing at the first attempt we have to broaden our search strategy. We can do this by searching the pack for *annealing* on its own, or *aluminium* on its own; either of these may well give us a set of cards that is not too large to look through. If this is unsuccessful,

we may be able to try *heat treatment* and *non-ferrous metals;* our list of terms used will show if we have employed these related terms. Again, if the two terms together reveal nothing, it is possible to search for them singly.

In a fairly small file it is possible to go through every card if the position becomes desperate, but this is not usually necessary if the original subject coding was adequate. However, subject specification is not exact in the way that an author's name can be specified exactly, and some failures must be expected. In a small index of the kind described, the likelihood of failure is much smaller than in a large library catalogue.

SYSTEMS FOR THE SMALL INFORMATION SERVICE

Edge-notched cards are like any other kind of mechanisation in that it is important to exploit the potential to the full by making one record serve several purposes. The following account shows how a card may be designed to serve as a book order record, temporary catalogue card and finally an entry in a ' current articles and books ' file.

The book order system was designed for a situation in which several people are entitled to suggest books for purchase, the actual ordering being coordinated by one man. When a member of staff wants a book ordered he fills in the details on one of the edge-notched cards (figure 5): author; title; publisher; date of publication; price; edition; requested by (their own initials); number of copies required; source of details (so that it is possible for the coordinator to check back for confirmation if need be); and, possibly, kind of material (research report, periodical) and language. Cards are handed to the coordinator, who decides which bookseller to order from (for special materials this might be, *eg* The Clearing House for Federal Scientific and Technical Information) and notches the card:

author, holes 1-12: three four-hole fields using the name-to-number coding list;

kind of material, holes 13-14: book, neither; research report, 13; periodical, 14; other, 13 and 14;

bookseller, holes 15-19: a five-hole pyramid giving codings for nine named booksellers and a tenth ' other ' code;

month ordered, holes 20-25: January and July 20, February and August 21, . . . June and December 25;

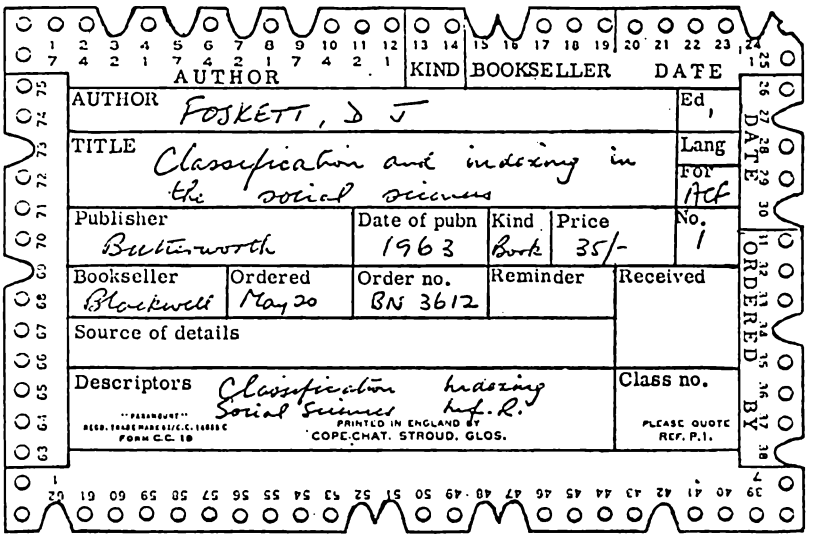

FIGURE 5: Book order system card

year of publication 26-30; five-hole pyramid showing year only;

requested by, holes 31-38: this gives 28 2-hole combinations. (If more people than this are entitled to order it will be necessary to use more holes; 10 holes give 45 codes, for example.)

The cards for books to be ordered are gathered together, say once or twice a week as a routine, depending on the numbers concerned and the urgency of the requests. They are then sorted by bookseller by needling from right to left across holes 19 to 15. The typist can now make out the orders from the details on the cards, filling in the order number at the

same time. The cards are then placed in the 'on order' file, which does not need to be kept in any particular sequence.

If the book is supplied promptly, the card is found in the 'on order' file by needling for author or bookseller; if a batch is received from one bookseller it is obvious that the quickest way to find the cards is to sort out all those for that particular bookseller rather than by sorting for each individual author. The cards are then placed in the corresponding books which are passed to the librarian for processing. When a call number has been allocated to a book, it is entered on the card, which is then given back to the coordinator for entry in the 'current interest' file.

If the book is not supplied promptly, this will be revealed by the 'overdue' procedure. At the beginning of each month the 'on order' file is needled for the month before last; for example, at the beginning of May the pack is needled on hole 22 for March. Cards which fall are for books which have not been supplied; by sorting these into order of bookseller it is simple to present the typist with all the information needed to send a reminder notice, including of course the order number and date ordered. When a reminder has been sent, hole 38A (the unnumbered hole after 38) is punched out; each month this hole is also needled to sort out cards for books which are further overdue and for which at least one reminder has been sent. It is rare that a book will be as much as six months before it is supplied or some other action taken, so it is possible to use the set of six holes twice—a useful economy, as the number of holes is limited.

As already mentioned, it is not necessary to keep the 'on order' file in any particular sequence. It is however simple to find out what has happened to any book requested, by needling for it under author or requestor. It is also possible to keep a check on expenditure on behalf of any individuals or department, by sorting out all the cards for books ordered by them. The system gives access to the 'on order' file through a number of approaches. It is typical of a mechanised

system that in some ways it requires more input (in the punching of holes) than a conventional system, but in return gives a much greater output.

Current interest file: It was suggested that cards for books received should have the call number added and then be placed in a 'current interest' file, similar to the kind of personal index already described. The object of this file is to

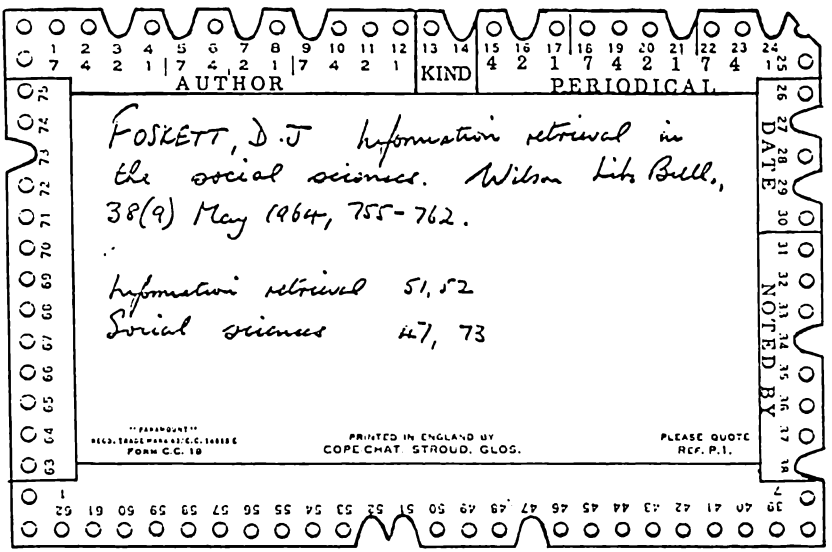

FIGURE 6: Current articles file card (illustrations by courtesy of Copeland-Chatterson Co)

supplement published abstracts and indexes and to provide a guide to the holdings of the library. Most of the entries in this file will be for periodical articles, but cards from the book order file form a useful addition to the content.

For periodical articles, unprinted cards are used, and the details of any article of interest are entered in standard form, as recommended. Holes 15-25 are used to code the periodical title; these holes are used in the book order card for 'bookseller' and 'date ordered', so they will not be used for this purpose for periodical articles. The two kinds of card are compatible; to sort one kind from the other all that is

needed is to needle holes 13 and 14, retaining cards which do not *fall*—these are cards for books. It will not be difficult to go through this file removing the few cards which are for books which have not passed through the book order system.

The subject approach is catered for in holes 39 to 75, using the superimposed random coding method described earlier. All kinds of material can be indexed in this way, and for the small information service this may replace the conventional library catalogue completely.

The exact mode of use of such a system will of course depend on local circumstances. This description has been included to show how a complete system was designed for a particular organisation, where it is possible to restrict the number of holes used for 'kind of material' to two (13 and 14) rather than the four (13-16) suggested earlier, and where the number of periodical titles to be coded was less than 300. The principles outlined in earlier chapters, together with the practical outline given here, should be sufficient to indicate the lines along which the design of a card for any similar set of circumstances should proceed.

SOME SUPPLIERS OF EDGE-NOTCHED CARDS

UK

Copeland-Chatterson Co Ltd,
Stroud,
Gloucestershire

Manifoldia Ltd,
Manisort Division,
Bromford Lane,
West Bromwich,
Staffs

Royal McBee (UK) Ltd,
Lilton House,
Goswell Road,
London E C 1

USA

Acme Visible Records Inc,
Crozet, Virginia

E-Z Sort Systems, Ltd,
45 Second Street,
San Francisco 5, California

Royal McBee Corp,
Port Chester, New York

PART TWO: OPTICAL COINCIDENCE CARDS

OPTICAL COINCIDENCE CARDS are commonly known by the name 'peek-a-boo'. They differ in many ways from edge-notched cards, the most significant being that they are not 'item' cards but 'term' cards: each card relates to a descriptor of some sort which may apply to a number of items, rather than giving all the information about one particular item, as does an edge-notched card. This fundamental difference implies a quite different approach to the various problems.

A typical peek-a-boo card is shown in figure 7. The body of the card consists of a grid of numbered positions; each square contains 100 positions, and there may be from ten to a hundred squares, giving a maximum of 1,000 to 10,000 positions. It is possible to obtain special cards giving more positions than this, but they are much more expensive, as is the ancillary equipment needed, and in any case are of interest only in very large collections of documents. Each position corresponds to the number of a document; when the term on a card is relevant to a particular document, the position for the number of that document is punched out, using a tool which again may vary from a relatively cheap hand drill to a much more expensive and accurate punch which can go through several cards at once.

The term for a given card is written at the top, usually in the right-hand corner along the cut-off edge. Along the bottom of the card are a series of notches which fit on to rods in a filing box; by means of these notches, each card is set off one notch from its neighbours so that the filing term can be seen easily. There is also on each card a row of dots or the

word OUT printed in such a position that it is possible to see at once when a card has been removed; the object of this is to make refiling easier after the index cards have been used. The method is shown in figure 8. However, the special filing cabinets are relatively expensive, and while their use in an information service may justify the cost, the individual may well decide to forgo the convenience in order to keep down

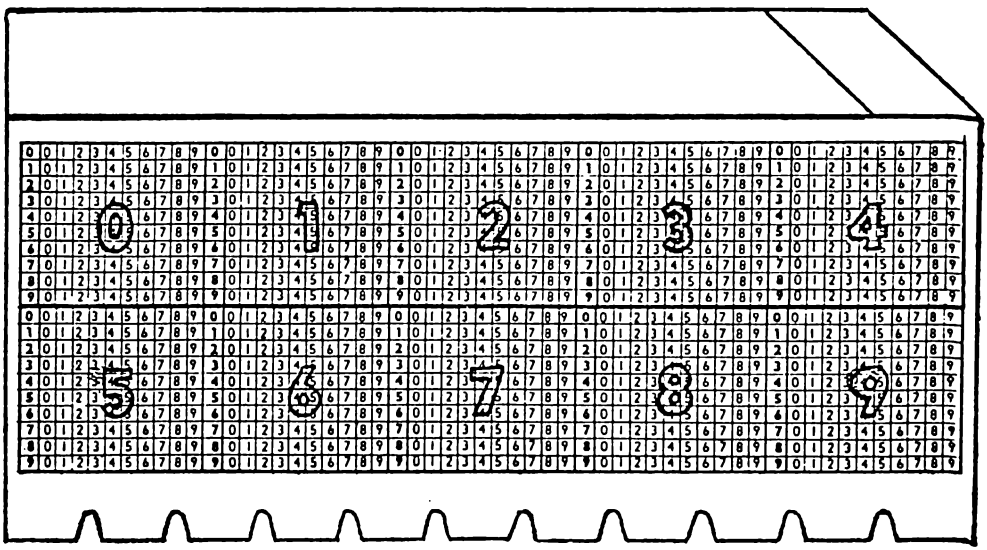

FIGURE 7: Diagram of a 1,000 position peek-a-boo card

expenditure. It is quite possible to keep the cards together in a shallow box or tray; the filing terms will not be immediately visible but this is only a minor problem.

Most cards also have room for notes, though in general there will not be much need for this; a possible use is suggested later. Some cards also have one edge ready for edge-notched punching, but this seems to be an unnecessary elaboration for most purposes.

It is essential to choose a card that will be large enough for the collection of documents envisaged. Since the number of positions on the card is limited, once the number of documents in the collection exceeds this limit, a completely new

set of cards must be started. While 1,000 positions may be enough for the individual, it will not be sufficient for even a small information centre, and a larger card should be used, for example a 5,000-position. One point may be borne in mind here: if the information is scientific and technical, or commercial, it could be useful to start a new index every year, since only recent information will be of value. In this case, a smaller card may be chosen than would otherwise be

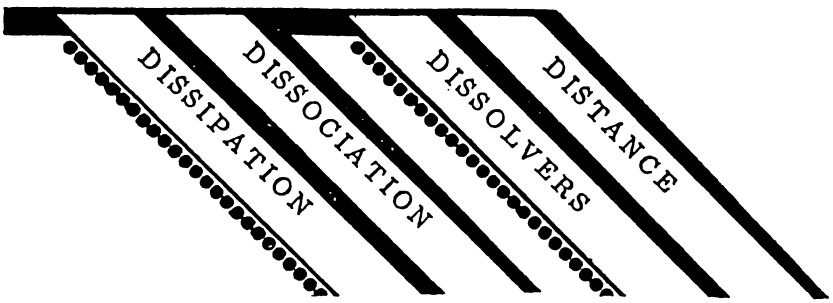

FIGURE 8: Peek-a-boo file with a card removed

necessary, since it has only to provide room for a year's accessions.

Practical aspects of the use of peek-a-boo cards are quite straightforward. The descriptors for a document are chosen and the cards corresponding to those descriptors removed from the file; the position for the number of that document is then punched out on each of the cards, which are then refiled. If a hand-punch is used, care must be taken to ensure that the holes are accurately placed. If a larger punch is used, the cards can all be punched at once, so that the holes are sure to coincide. The greater the number of punching positions on any card, the smaller each individual position will be, and the more accurately the punch must be positioned; however, even with 10,000-position cards a reasonably

steady hand is all that is needed to obtain the necessary accuracy.

To search the index the process is reversed. Once again the desired descriptors are established and the cards for these removed from the file. They are then held up to the light; any positions where all the cards are punched will immediately show up as points of light. (If the original punching was badly done, the operation will be rather more difficult, but things will have to be very bad before it becomes impossible.) It is important to align all the cards exactly, of course. If the index is to be used a lot, it is worth buying or making a ' light box '. This consists basically of a sheet of translucent material —plastic or ground glass—with a light source behind it, the whole being assembled into a convenient box. More sophisticated versions may have a control for the brightness of the light, which can make searching easier, by making it possible to broaden the search strategy.

Suppose that we frame our question in such a way that five descriptors are involved: for example, we may wish to find out whether we have any information on *foundations* for the *power supplies* for the *electromagnet* of a *proton synchrotron,* and we therefore take out the cards for the five terms underlined. If we had not previously had any information on this combination of descriptors, no holes will show up and the search will have failed. However, if we had already had a document on ' *foundations* for the *power supplies* for the *electromagnet* of an *electron synchrotron* ', four of the descriptors will coincide, the only card in our search not being punched for that document being the card for *proton.* By increasing the brightness we can make the light shine through the single thickness of card sufficiently brightly for us to see it; we now know that the document with that number is very nearly what we are looking for, though not precisely. By increasing the brightness in controlled steps we can find the positions where all, all but one, all but two, etc of the cards are punched. Some manufacturers are now making

their cards on special translucent card to facilitate this search-broadening process.

It will be obvious that the information revealed by the index is a document number, corresponding to a position on the cards; it is therefore essential to have a subsidiary file of ordinary catalogue cards in document number order, so that we can look up the numbers revealed by the index and find further information on author, title and source. It will not usually be possible to keep the documents themselves in number order, but there will be certain instances where this is feasible, for example a collection of pamphlets or research reports. In these cases, the subsidiary file will not be necessary, though it may be a convenience. Normally, the collection of references in the index will be to all sorts of documents —periodical articles, books, conference proceedings and so on —and the subsidiary file in document number order cannot be dispensed with.

Equipment required for a peek-a-boo index will thus be as follows: peek-a-boo cards, of adequate size (the number of cards required is discussed later); a container to hold them—either a ready-made holder supplied by the manufacturer or a shallow drawer or tray, depending on the amount of money available and the use to be made of the index; ordinary catalogue cards (12·5 cm by 7·5 cm) or similar cards, and a container for these; a punch for drilling out the holes, either a simple handpunch or a more elaborate machine, again depending on the two factors already mentioned. Optional for the personal index, but worthwhile for the information service, is the light-box, available in varying degrees of sophistication (and price).

A PRACTICAL FILE

It was suggested in part one that for an adequate index we may need to be able to find documents through their authors, date, source and subject, and that all these things must therefore be catered for by the system. In addition we have seen that for a peek-a-boo file we need a subsidiary file in docu-

ment number order. As with edge-notched cards, there are various ways in which these approaches can be catered for.

Author: One way of entering a document under author is simply to compile an author index on catalogue cards. Each card will bear the name of an author and the numbers of any documents with which he is associated, either as primary or as a secondary author. This file can be kept in the same container as the accessions file in document number order. If the subject field is one in which the author approach is important, it will be better to compile an author *catalogue* rather than an index. Each document will have a card made out for each author, as before, but instead of merely giving the document number, these cards will give details of title, date and source, just as do entries in a conventional library catalogue. Such a catalogue will take longer to compile, but when a search is made to see what documents in the collection are by a particular author, the catalogue will give the answer directly, whereas with the index it will be necessary to make a note of the numbers and then look up the details in the accessions file. A decision can usually be made at the outset as to whether the additional effort involved in compiling a catalogue will be justified, but even if the wrong decision is made, it is possible to change it at a later stage without too much confusion. If a catalogue is begun, but proves to be little-used in practice, then it is very simple to revert to simple index entries, ignoring any catalogue details already in the file. If a start is made with an index, which later proves to be a nuisance, the change to catalogue entries should be made and a decision taken as to whether to go to the trouble of converting the index entries already in existence into catalogue entries. In general, it is best to make catalogue entries for an information service, since the kind of enquiry that may have to be answered will be more varied than those likely to occur in a personal index.

To compile an author index or catalogue along the lines indicated, involves setting up an extra file of cards, and this may not appeal for the personal index, since it means addi-

tional complexity as well as extra effort. It is possible to use peek-a-boo cards to make an author index, and this has the advantage that the cards for the author index can be punched at the same time as those for the other descriptors. Four sets of twenty-six cards are needed, each set running from A to Z. To index a work by a particular author, take the first four letters of his name; then from the first set of twenty six take the card corresponding to the first letter of the name, from the second set the one corresponding to the second letter, and similarly for the third and fourth sets. The document number is then punched into these four cards. It is possible to use the first three letters of the name and the first initial, instead of four letters from the surname, but when searching one does not always know the initials of the author one is looking for, so the first method is to be preferred. When searching, the appropriate cards are taken from the four sets and compared; holes through all four show the numbers of documents written by the wanted author. There is a possibility of false drops: WHITe and WHITwick, for example. It is also not desirable to index more than one author for each document, otherwise the potential number of false drops increases very rapidly; for example, if one indexes both BELLamy and SANDers for the same document, that number will appear as a false drop if one searches for BALL, BALDock, BANDler, BENDix, BELDing, BENLey, SALLis, SELLars, SENLac, SENDall, SELDon, or any other similar names. This of course will be in addition to the inevitable false drops for names such as BELL, BELLairs, BELLaston, SANDS, SANDerson, and so on, which would be coded in the same way in any case.

An alternative method would be to use the coding table for names drawn up for edge-notched cards, given in appendix one. Here, instead of 104 cards, one would need thirty-three sets of ten, one for units, one for tens and one for hundreds. To code *Bellamy* one would look up the table and find that the coding was 61; one would therefore select the 0 card from the hundreds, the 6 from the tens and the 1 from the units set. Apart from the fact that only thirty cards are

used instead of over a hundred, there seems to be little advantage in this method, since it involves a further operation to code the name or to search for it. The possibilities of false drops are also present, and it is not possible to index more than one author per document.

These two methods involve a definite set of cards, thirty or 104 depending on which is chosen. It is, however, possible to treat authors' names like any other descriptor and make out a card for each individual, and this is probably the easiest solution. It does involve an indefinite number of cards, since nearly every document indexed will mean adding to the cards for authors' names. As many authors as required can be indexed for any given document, for with this form of direct coding there is no possibility of false drops. For the personal index, where the number of documents is likely to be fairly small in any case, this is certainly the simplest method and should not prove intolerably expensive. The small information service is probably best served either by a conventional kind of author catalogue, or by the four-letter coding system.

Date: Date is very simply dealt with by having a card for each year of interest. For older works it will usually be adequate to have a card for each decade; in some subject fields where older works are important, it may be necessary to have a card for centuries other than the present. No difficulties are likely to be encountered here.

Source: Periodical titles may be treated in the same way as authors; either each title can be given an individual card, which is probably the best solution in a personal index, or four-letter coding can be used. It will be necessary to use another set of 104 cards in the latter case, but there is a set of four-letter codes for scientific and technical periodicals known as CODEN which may be used. It is not feasible to use the straightforward method of taking the first four letters of the title, since this would lead to merging of all the periodicals beginning JOURNAL of . . .

Bodies issuing reports (corporate authors) may be coded in the same way as authors: direct coding, one card for each corporate author, or four-letter coding. In the latter case, the set of 104 cards used for periodical titles may be used, since a report will not be coded for periodical title, nor will a periodical article be coded for corporate author.

It may be thought unnecessary to code either the title of the periodical or the report-issuing body, but merely to indicate the kind of source material. In this case, the words *report, periodical, pamphlet,* etc may be used as descriptors and cards made out for them on which the appropriate positions are punched out. The larger the collection of documents, the less likely it is that it will be feasible to index such features as periodical title. If a search under *Journal of the American Chemical Society* reveals 500 document numbers it will not have been particularly useful as a retrieval factor. If experience shows that this particular section of the index is not serving any useful purpose there is no point in continuing it; it is better to cut one's losses than to put more effort into maintaining a tool of doubtful value.

Subject: Many of the problems of subject approach are common to all retrieval systems, and were indicated in the discussion in part one. However, peek-a-boo cards are arranged alphabetically and so no problems of coding for subjects arise; we are concerned only with the choice of terms and with links between related or synonymous terms. Also, since we are using the terms themselves, it is not essential to compile a separate authority file, though if the file is likely to grow to any great size, such a list can be of value.

To index by subject, read the document carefully, selecting terms which are likely to be valid as searching factors at a later date. It is important not to overlook significant ideas. but equally important not to index insignificant pieces of information which one will not wish to retrieve later on. Reference to a classification scheme or thesaurus may be helpful in choosing terms to index by. Once the desired terms have been selected, look in the index to see if they have been

used already; if so, withdraw the card from the file, if not, make out a new card. When all the required cards have been assembled, punch out the position corresponding to the document number. With a hand punch it may be necessary to deal with each card separately, but the heavier punches will go through a number of thicknesses of card at one time. The cards can now be refiled in the alphabetical sequence.

As suggested in part one, it is useful to merge different forms of the same word: *weld, welding, welded* should all be treated as one. The space on the peek-a-boo card for notes may be used to enter information of this kind, as well as relationships of various kinds. To draw attention to related terms, one may use the abbreviations found in the EJC Thesaurus: BT broader term *ie* a wider subject; NT narrower term *ie* a more specific subject; RT related term *ie* a subject of equal importance probably forming part of the same wider subject; UF used for *ie* the term chosen is the preferred form of synonym and is used for the others. There will be references from unused forms for synonyms in the form, for example, CELLS (biology) *use* CYTOLOGY. So one may find cards in the index as follows:

CELLS (biology) *use* CYTOLOGY

CYTOLOGY
UF CELLS (biology)
BT BIOLOGY
NT MEIOSIS
RT HISTOLOGY

BIOLOGY
UF LIFE
NT CYTOLOGY
MICROBIOLOGY
etc.

If the index is likely to be large, it may be desirable to keep a separate authority file, or to use a thesaurus such as the EJC, because a separate file is easier to consult when documents are being indexed. However, for indexes of small to medium size a separate file is not necessary, its place being

taken by the main file. Additional cards can be inserted for synonyms, and can be left unpunched as suggested above. An alternative, which gives more work but which may make searching quicker, is simply to punch document numbers into all the cards for synonyms: in the example above, use both CYTOLOGY and CELLS (biology) and simply punch both together. This alternative is feasible in subject fields where terminology is well established and synonyms are the exception; in areas where terminology is ill-defined it is best to select a preferred term and refer from the others in the usual way.

In a large file it may happen that some terms begin to occur too frequently; if an index term leads to more than about fifty references, it may help if we can refine the idea to cut down the number at that particular index term by splitting it. One way of doing this is to use roles. For example, if we have a metal such as *copper* or *aluminium,* the reference may lead us to an article on that metal as a raw material, *eg manufacture of copper pipes,* or it may lead to an article on the metal as an end product, *eg electrolytic extraction of aluminium.* We can refine our indexing by having two cards instead of one for each material: one as raw material, the other as end product. On each card we only enter the appropriate document numbers, thus reducing the number of references punched in each card. The EJC Thesaurus has a set of roles, but it may well be possible to devise a set for one's own particular index.

In general, it is best to use single uncoordinated terms, nouns in the plural form. There are times when it is useful to precoordinate terms, if they regularly occur together. It would probably be better to use *particle accelerators* as a descriptor (either directly or in the inverted form *accelerators, particles*) rather than have to pick out two cards from the index every time we want this particular subject. As with edge-notched cards, and indeed any system for subject analysis and coding, it is usually best to keep the vocabulary as simple as possible, but experience will show when it is wise to depart from this rule.

Other terms that may be encountered in articles about indexing systems include links and interfixes. It is wisest to ignore these devices altogether; their use is in many cases incorrect, and unless they are fully understood they are unlikely to be of any help in compiling an index.

Accessions file: This is an essential part of a peek-a-boo index. Each item indexed must have a card made out for it

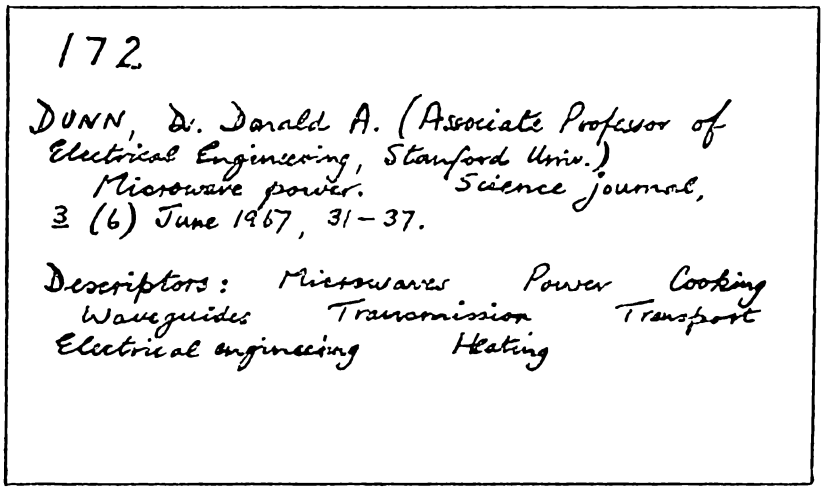

FIGURE 9: A typical peek-a-boo accessions card

in this file showing its running number, author, title, bibliographical reference, and the descriptors by which it has been indexed. The peek-a-boo index itself will give only a number, and full details can only be found through the accessions file, kept in document number order. The accessions card will normally be the first to be made out, and then serves as a basis for the subject and author indexing. It is also valuable to have the subject descriptors listed on the cards when searching; if one reference has been found, for example through the author's name, it is very useful to be able to see at once a suitable search strategy to find further items (figure 9).

Although cards have been suggested as the best medium for the accessions file, it is quite possible to use a book in-

stead, since no question of interfiling arises. A small notebook could be used, with a page for each entry; this would be better than a larger size, with more than one entry per page. Notebooks have of course a limited number of pages, so as the file grows more and more books might have to be brought into use. For a file of a few hundred items or more, cards are cheaper and more convenient.

SEARCHING THE FILE

All searching in peek-a-boo indexes is basically a matter of comparing cards to see which positions let light through. This will give the numbers of the appropriate documents, which should be noted on a sheet of paper. By turning up these numbers in the accessions file we can find the details of the original documents and decide whether they are relevant or not. Having established which are relevant to the enquiry in hand, we can consult the originals.

The exact search procedure will, of course, depend on how the various index factors have been dealt with in indexing. To find a document by a particular author, either 1) look in the author index and note the document numbers, or the author catalogue, which will give full details, depending on which has been made; or 2) from the peek-a-boo file take the relevant card, if direct entry has been used, or cards if a coding method has been employed, and from these find the relevant document numbers. Having the document numbers, details of the documents can be found from the accessions file.

To find a document in a particular periodical, or a report issued by a particular body, a similar procedure is followed. One would not normally want to find a document solely by date, but this can be done. Normally one would use the date card to narrow down a search by other factors: for example, to find a paper by a given author published in 1965, or to find articles on a given subject published in 1967. The only point worth noting here is that to find all the documents on a given subject published since, say, 1964, we have to search

the subject cards with the 1965 card, noting all the relevant numbers, then with the 1966 card, then with the 1967 card. There is no quick way of merging 1965, 1966 and 1967 into a single search.

Searching by subject is also straightforward. We have to formulate the search in the correct terms and then take the appropriate cards from the peek-a-boo index and compare them in the usual way. If our first search reveals nothing, we can either increase the brightness of the light box and thus reveal positions where all but one card are punched, or we can broaden the search by removing one or more of the cards. For example, the search for *foundations* for the *power supplies* for the *electromagnet* of a *proton synchrotron* was quoted earlier to show how increasing the light would show up a position punched for all the terms except *proton,* but exactly the same result could have been obtained by removing the card for *proton* altogether. The difference lies in the fact that increasing the light is the equivalent of removing each of the cards one at a time, but does it in one operation.

We can also broaden the subject searching by choosing different terms. If, again quoting an earlier example, we do not find what we want by searching the card for *cytology,* we can search instead the card for the broader term *biology*. The notes on related terms on the cards will help us here, or the thesaurus if we are using one.

It must be stressed again that any search of the peek-a-boo index will only yield a set of document numbers, which must then be followed up in the accessions file. This means that we shall not know whether we have found any relevant documents until we look through this file. If our search of the peek-a-boo index reveals any great quantity of document numbers, it may be worth trying to narrow down the search to some extent before checking the accessions file, rather than looking up all the numbers revealed by the first search. Experience in using the file, and also personal preference, will usually indicate the choice of strategy to be followed in such cases.

If a search proves fruitless, but we nevertheless feel sure that there is something in the index, it is possible in a personal index to look through the whole of the accessions file. This begins to get impractical when there are more than a few hundred cards in the file, and in any case it is a counsel of desperation.

SOME SUPPLIERS OF PEEK-A-BOO CARDS

UK

George Anson & Co Ltd,
Solway House,
Southwark Street,
London S E 1

Carter-Parratt (VISIRECORD) Ltd,
Orchard Road,
Sutton,
Surrey

Manifoldia Ltd,
Manisort Division,
Bromford Lane,
West Bromwich,
Staffs

Royal McBee (UK) Ltd,
Lilton House,
Goswell Road,
London E C 1

USA

Visirecord Inc,
375 Park Avenue,
New York 22, New York

Royal McBee Corp,
850 Third Avenue,
New York 22, New York

PART THREE: CHOICE OF SYSTEM

THE ADVANTAGES AND disadvantages of the two systems have now been outlined and it should be possible to make a choice of the kind of system one would prefer in a given situation. In the long run, personal preference may well be the deciding factor.

EDGE-NOTCHED CARDS may be summarised as follows:

Advantages

No limit to number of documents a given size of card will take.

One file gives all the information required in searching.

Each card carries full details of a document.

Very little equipment is needed.

Disadvantages

Limit to number of descriptors that can be coded.

Any system of coding other than direct leads to false drops (but the importance of this should not be overestimated).

Separate record needed to show codings for names and other descriptors.

Searching a large file can be very tedious.

PEEK-A-BOO CARDS may be summarised:

Advantages

No limit to number of descriptors that may be used.

Searching the file is simple no matter how large it is.

Disadvantages

Limit to number of documents that can be indexed by a given size of card.

Separate accessions file must be maintained, and searches lead to this, not direct to details of the documents.

Equipment needed can be quite expensive.

Assuming that sensible decisions are taken about the size of card in each case, the most important factors are probably 1) the need for fairly elaborate coding systems for edge-notched cards, and the difficulty of searching a large file; 2) the use of natural language and ease of searching peek-a-boo. These two factors have led to a tendency to adopt peek-a-boo rather than edge-notched cards in libraries and information services; but for personal indexes where the numbers of documents concerned are not great, edge-notched cards can prove very satisfactory and take up less space, as well as being less expensive.

It is very difficult to give an exact idea as to how many cards will be needed in either case. It is rather expensive per card to buy edge-notched cards in lots of less than 1,000, but there is not a great deal of difference in cost between different sizes of card, so it is worth buying a size that is sure to be large enough. 1,000 cards will index up to 1,000 documents, ample for most personal files; the information service will probably need more than this, but can buy in lots of 1,000 at a time. A large file, which it would be time-wasting to search *in toto* frequently, may be pre-sorted into groups which are of more manageable size.

The number of peek-a-boo cards that will be needed depends on the number of descriptors, and here again it is almost impossible to give any valid estimate. However, it has been shown that detailed indexing of a collection of between one and two thousand research reports gives rise to some six to eight hundred subject terms. Once the thesaurus reaches this size there is little further increase, even if the collection of documents grows, provided that there is no change of subject field. For a personal index two hundred cards would almost certainly be enough for some time. If names are coded as suggested using four sets of cards, another 104 cards must be added for this purpose, and a similar number to code sources if desired.

SOME USEFUL REFERENCE BOOKS

THE MAJOR WORK in this field is undoubtedly CASEY, R S and others: *Punched cards: their application to science and industry*. Reinhold, second edition 1958. This covers every aspect, including the mathematics of coding, and gives several case histories. It is however a little out of date and does not cover recent developments in the theory of subject approach. *American documentation* is a useful source of ideas and case histories.

Probably the best example of a large-scale thesaurus is ENGINEERS JOINT COUNCIL: *Thesaurus of engineering terms*. EJC, 1964. This not only contains the thesaurus itself, but also has several introductory pages explaining how it was compiled and how it should be used, including the EJC system of roles.

The classification scheme most widely used in special libraries is: BRITISH STANDARDS INSTITUTION: *Universal decimal classification: abridged English edition*. British Standards Institution, third edition revised 1961. (BS 1000A: 1961; FID no 289). Until the full English edition appears this is the most convenient form in which to use the UDC, but expansions in some subject fields are available in English, and some others in German or French. The four-letter codes for periodical titles mentioned in the text are to be found in: AMERICAN SOCIETY FOR TESTING AND MATERIALS: CODEN *for periodical titles: an aid to the storage and retrieval of information and to communication involving journal references; prepared and maintained for Committee E-13 by the Wyandotte-ASTM Punched Card Project; L E Kuentzel, editor*. ASTM special technical publication no DS23, two volumes 1966.

DEFINITIONS

THE FOLLOWING IS a list of definitions that may prove helpful. No attempt has been made to provide an exhaustive vocabulary, but terms used in the text likely to be unfamiliar to the general reader are included. There is unfortunately no convenient and satisfactory list of terms used in information retrieval.

Bibliographical references (citations): It is important to have some system of uniquely specifying books, periodical articles, pamphlets and other printed documents. Such specifications are known as references or citations, and must be stated according to a set of rules, so that adequate information is always given in a conventional form. An article in a periodical may be given a reference in the form:

WATERMAN, Alan T: 'The National Science Foundation'
Impact of science on society 9 (4) 1961, 199-221

or alternatively:

'The National Science Foundation' by Alan T Waterman. *Impact of science on society,* 1961, vol. 9, no 4, pp. 199-221.

There may be many other variants. Whichever style is adopted should be adhered to strictly, and it is best to use one of the recognised forms. There is a British Standard: BRITISH STANDARDS INSTITUTION: *Bibliographical references.* BSI, 1951. (BS 1629 : 1950, incorporating amendments issued May, 1951). For those who require fuller guidance the Library of Congress has published a useful manual: US LIBRARY OF CONGRESS (Reference department, general reference and bibliography division): *Bibliographical procedures and style: a manual for bibliographers in the Library of Congress.* Library of Congress, 1954 (reprinted 1966 with list of abbreviations).

Call number: A number given to a document in a library so that it may be easily located. The call number often consists of a combination of class number and author number in the case of books. For reports or patents the number of the document itself is normally used. A call number should identify a particular book, whereas a class number only identifies a group of books on the same subject.

Catalogue card: The standard size of library catalogue card is 12·5 cm × 7·5 cm, but is usually known as 'five by three' (inches). Cards may be obtained plain or ruled, or with margin rules.

Classification scheme: A systematic tabulation of subjects. An enumerative scheme tries to list all subjects, simple or composite (*qv*); an analytico-synthetic scheme lists single concepts and provides methods for linking these to specify composite topics. The Universal Decimal Classification lies somewhere midway between the two extremes. There are now quite a number of special schemes in existence covering particular subjects. Many of these are listed in BARDEN, B R and DENISON, B: *Guide to the* SLA *loan collection of classification schemes and subject heading lists.* New York, Special Libraries Association, fifth edition 1961. The collection has now been transferred to a new centre at Western Reserve University, and a new catalogue is to be published.

Clearing-House for Federal Scientific and Technical Information (CFSTI): The major source of unclassified (*ie* non-secret) reports in the USA. It publishes a list, *US Government research and development reports* (USGRDR) and also a government-wide *Index to reports.* Reports may usually be purchased from CFSTI at Springfield, Virginia.

Composite subject: A subject consisting of more than one element, *eg extrusion of copper; influence of PAS on the resistance to streptomycin of tubercle bacilli.*

Descriptor: A term or group of terms used to describe a document. Usually refers to subjects, but may include authors, titles, etc if taken in the broad sense. Descriptors are normally single concepts, not composite.

Extra cards see *False drops.*

False drops: Unwanted cards which fall during a sort owing to some form of ambiguity in the coding used. Known also as *extra cards,* a name in which the implication of 'something wrong' is less explicit.

Field: A group of holes in an edge-notched card devoted to a particular purpose.

Homograph: A word spelt in the same way as another but having a different meaning, *eg* pitch (aeronautics); pitch (music); pitch (football); pitch (bitumen); sole (fish); sole (shoe). Sometimes called *homonyms,* though these are strictly words which are spelt differently but sound the same, *eg* sole, soul; sew, sow, so.

Interfixes: If a document deals with two separate composite subjects, each complete in itself, it is possible to 'tie' the subjects together by means of interfixes. Suppose document number 791 deals with *heat treatment of aluminium and of magnesium alloys;* this is in effect two subjects: *heat treatment of aluminium* and *heat treatment of magnesium alloys.* A simple indexing of this will mean that if a search is made for documents on magnesium-aluminium alloys, this one will be retrieved, though it is not relevant. If we make out two entries, number one 791 and the second 792, we can index the two subjects separately; such a device is known as interfixing.

Links: A similar kind of device to interfixes, but this time linking elements *within* composite subjects. If a document deals with *plastic tops for metal containers,* then by a simple indexing process it will be entered under each of the four terms; thus it will be retrieved if we search for *metal tops for plastic containers,* though it is not relevant. If *plastic* and *tops,* and *metal* and *containers,* could be linked, this kind of false drop would be eliminated. However, M Taube showed in a useful article in *American documentation* 12 (2) 98-100 April 1961, that this kind of linking is logically indefensible, and that it is better to use roles (*qv*).

National Lending Library for Science and Technology (NLL): The major source of unclassified reports in the UK. It publishes a list, *British research and development reports,* and also serves as a source for US reports. Documents may be borrowed, or photocopies may be purchased, through a voucher system. The NLL is, however, mainly of importance for its collection of scientific and technical periodical literature, of which it claims to have the best collection of current material in Western Europe. Materials may be obtained on loan from NLL, Boston Spa, Yorkshire, direct to approved borrowers or through an approved library.

Research reports: In many fields of science and technology the research report rather than the periodical article is now the main form of publication. Reports are generally poorly covered by abstracting and indexing services, and the personal index is often a useful way of keeping track of those of interest. Reports may be obtained by the National Lending Library or Clearing-House for Federal Scientific and Technical Information in many cases.

Roles: A material may be treated in more than one way in the literature; for example, one might have an article on *copper pipes* in which the copper is used for making pipes; and another on *the extraction of copper from industrial scrap metal* in which the copper is the end product. By simple indexing, using only one card for *copper* in a peek-a-boo index, all these senses will be merged. By making out a card for each sense, and showing the context by means of roles, we can avoid this, while still having the power to find everything on copper if we wish. The EJC Thesaurus contains a useful system of roles.

Schedules: Lists of terms. In this context, normally used for the systematic lists found in classification schemes.

Thesaurus: An alphabetical tabulation of subjects, normally restricted to single concepts, and showing relationships with other subjects, whether broader, narrower or coordinate.

APPENDIX ONE: NAME-NUMBER CODINGS

THE FOLLOWING LIST provides a simple method of converting names to numbers for coding into three fields; normally 7-4-2-1 coding in each field will be satisfactory. Such a list is relatively simple but very tedious to compile. This one is derived from the (British) Library Association Yearbook, which gives a list of members. There are some 12,000 names listed, and not more than 999 codes available, so on average each code will represent twelve names, though it has to be remembered that in the case of common names the twelve may be differentiated only by initials or even full forenames. Another way of working the ratio of names to codes out is to count the number of pages—about 340 in this case; this gives about three codes per page. Any similar list may be used: a telephone directory, or a large list of authors such as Whitaker's *British books in print . . .* (formerly *Reference catalogue of current literature . . .*), or the American equivalent, R R Bowker's *Books in print . . .*, or the catalogue of a large library. In each case the number of names per code has to be worked out, and allowance has to be made for names which do not appear but conceivably might.

Some examples of names and the appropriate codings from this list may be helpful:

ANLIKER	17	ERICSSON	269
BENCKHUYSEN	61	FOREMAN	295
BROWN, P W	109	GATES	316
CICALA	152	KOVALSKI	501
CURRAN	199	LIPPERT	528
DEJKA	223	METCALF	594
DONALDSON	235	NOBLE	637

KOZIK	618	TAIT	839
PRASAD	699	TOUPIN	868
RANGER	716	VOREL	886
SATTLER	761	YOUNG, J	969
SKYGGE	792	ZHUKOV	971

These names are taken at random from WADEX, the index to *Applied mechanics reviews* 16 1963, and illustrate how the list works.

It is of course possible to obtain any desired degree of refinement if one is drawing up a list. Two fields only may be used, giving 99 potential codes, or four giving 9,999, if this is thought necessary. An alternative method which does not involve a list of this kind is to code each name in sequence, keeping an index of the code used. However, this is not so satisfactory as using a list, for it is no longer possible to sort into alphabetical order, since this no longer corresponds to the numerical order. Also, there will only be room for a limited number of names in any given number of fields, and though 999 may be enough for a personal index it is unlikely to be adequate for anything larger.

A - Abn	1	Anu - Ard	18
Abo - Adam	2	Are - Armss	19
Adama - Adams	3	Armst	20
Adamsa - Af	4	Armsu - Arn	21
Ag - Aitc	5	Aro - Asg	22
Aitd - Ak	6	Ash - Asht	23
Al - Ald	7	Ashu - Ask	24
Ale - Allarc	8	Asl - Atkinr	25
Allard - Allem	9	Atkins - Atkz	26
Allen, A - L	10	Atl - Aur	27
Allen, M - Z	11	Aus - Auz	28
Alleo - Allz	12	Av - Az	29
Am - Anderson, D	13		
Anderson, E - Z	14	Ba - Baf	30
Andersp - Andrewr	15	Bag - Baild	31
Andrews	16	Baile	32
Andrewt - Ant	17	Bailf - Bakeq	33

APPENDIX TWO : RANDOM NUMBERS

THE EASIEST WAY to establish the random codings required for superimposition is to use a table of random numbers. A convenient set is to be found in KENDALL, M G and SMITH, B Babington: *Tables of random sampling numbers.* Cambridge University Press, 1939 (reprinted since then). This contains pairs of two-figure numbers, so that if a 100-hole field is available, these pairs may be used as they stand. If a more limited number of holes is available, then any pairs containing unusable numbers must be discarded.

Another source which is widely available is FISHER, *Sir* R A and YATES, F: *Statistical tables for biological, agricultural and medical research.* Oliver & Boyd, sixth revised edition 1963 pp 134-139. This contains a list of 7,500 two-figure numbers in random order. Again, one would either take pairs as they come or discard unusable numbers. Both these sources give advice on how to use the tables, though for information on how the tables are derived and tested it is necessary to go back to the first edition of Fisher and Yates. Kendall and Smith is more helpful in this respect.

If neither of these or any similar sources is available, it is possible to use other tables of numbers. Fisher and Yates used a table of 20-figure logarithms (taking the last figures, not the first); it would be possible to use a table of cubes such as that found in *Handbook of chemistry and physics;* editor-in-chief C D Hodgman, Chemical Rubber Publishing Co, annual. Again, one would select figures other than the first in any given number.

If no tables of this kind are available at all, less satisfactory methods must perforce be adopted. It is possible to number

a set of cards with the numbers available, shuffle them well and then take two. After these have been noted, they are returned to the pack and the operation is repeated. For this to work really well, shuffling must be complete each time, so

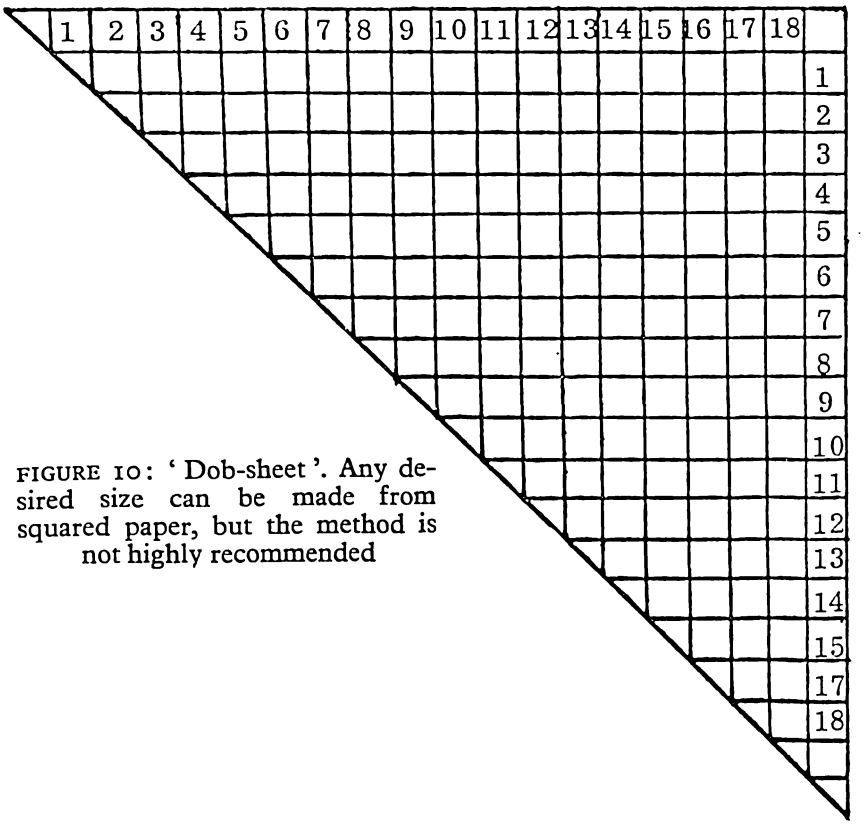

FIGURE 10: 'Dob-sheet'. Any desired size can be made from squared paper, but the method is not highly recommended

in practice it is less than perfect. Another home-made method is to compile a sort of 'dob-sheet' in the form of a triangle (figure 10). Each box represents a pair of numbers, and selection is made by the time-honoured method of sticking a pin in. A friend with a computer may also be able to help as it is relatively easy to use this to generate and print out a set of random numbers.

Whichever method is adopted, it is important to observe the precautions mentioned earlier (page 27). A file must be

kept of terms used showing the coding used for each; this file will be in alphabetical order. A second file must also be kept showing which codings have been used; each card should have the pair of numbers (lower first) and the terms for which that coding has been used, and this file will be kept in numerical order. By these means duplication of codings can be avoided.

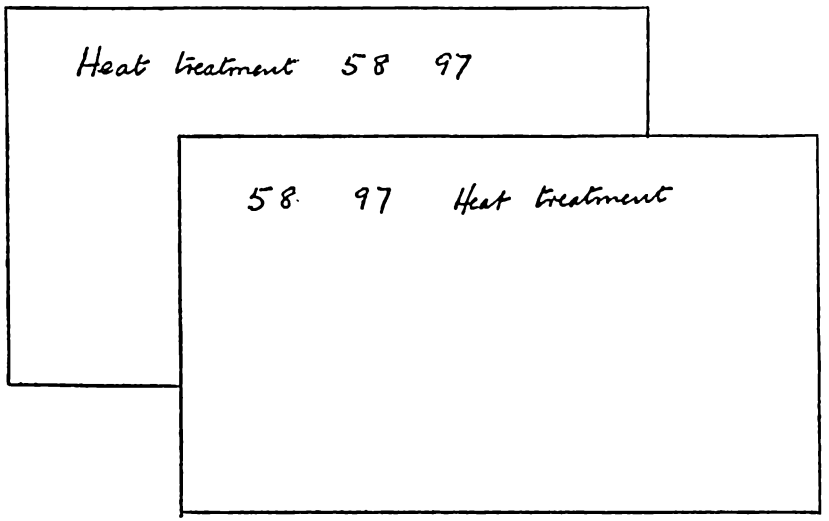

FIGURE 11: A typical pair of record cards for random number coding

INDEX

(*e*) *edge-notched* (*p*) *peek-a-boo*