A program suite for the production of articulated subject indexes

M. F. Lynch and J. H. Petrie

Postgraduate School of Librarianship and Information Science, University of Sheffield, Western Bank, Sheffield S1O 2TN

A suite of computer programs has been devised to generate articulated subject indexes with machine assistance. The suite has been written for the ICL 1900 series of machines in PLAN assembly language. The application of the index generation algorithm necessitates a double sorting operation. The first of these is straightforward. The strategy of the second is determined by the nature of the algorithm for rearranging index phrases, which is context-sensitive, and therefore requires special procedures for organising the data in core storage. The programs incorporate both tree and letter-table systems for dictionary look-up, and use simple list processing techniques.

During the last decade the production of subject indexes by machine has become a widely accepted technique. Some computer manufacturers now offer software packages for this purpose in their ranges of software. The most popular and by far the best known form of machine produced index is the KWIC or KeyWord In Context index. These indexes can be simply and economically produced from document titles and the display is usually limited to the standard line-printer character set. The display is sufficient for small indexes but for retrospective searching, when large numbers of documents are involved, the method has a number of deficiencies. The index relies on the information content of the titles of documents, the structure of the display is designed to save space rather than be helpful to the searcher and the use of an all upper case line printer output limits its acceptability to users.

Traditionally produced indexes are, in general, of a higher quality than the KWIC system for they often incorporate controlled vocabularies, displays with helpful structures and high graphic quality in printing. One form of index that is widely used by abstracting services for their retrospective indexes is the Articulated Subject Index (ASI). An example of such an index, that to Chemical Abstracts, is shown in **Fig. 1**. It is made up of entries which comprise a subject heading, a modification and a reference. The subject headings are the alphabetised entry points to the index. The modifications are phrases which augment the descriptive subject heading and they are alphabetised within each subject heading. The reference links the searcher to the bibliographic details of the original article.

Production of high quality subject indexes by manual techniques is a time consuming and hence a costly operation. The time delay between the publication of a document and its appearance in an index is an important factor in an assessment of the effectiveness of the index. The use of computer techniques can substantially reduce this time lag. Several systems are now in operation and the area has recently been reviewed (Lynch, 1969).

This paper describes a computer-based system to produce ASI's which takes over many of the former manual tasks. Instead of creating the index entries for each document manually, the indexer produces one or more title-like phrases and the index entries are generated from these. Alphabetic sorting and display production operations are also undertaken by the machine.

The index entry generation algorithm

A study of the index entries in Chemical Abstracts Subject Index (Armitage and Lynch, 1968) revealed that certain index entries could be rearranged in a simple way to form title-like phrases. It seemed that the indexers intuitively did the opposite,

i.e. started with English language phrases and generated the index entries from them. It was found that these intuitive processes could be formalised and thus it became possible to devise an algorithm to generate a number of index entries from one title-like phrase. In the phrases shown in **Fig. 2** the indexer uses the '<' and '>' characters to indicate which words in the phrase are to become subject headings. The phrase is analysed by comparing each word with a list of function words OF, BY and ON are such words. When a subject heading has been removed from a phrase the remainder can be articulated i.e. broken and rearranged to form a modification. A phrase can be broken only at the function words. The phrases are processed and the index entries shown in **Fig. 3** are produced Each input phrase produces a number of index entries equal to the number of assigned subject headings and can be regarded in effect, as a number of phrases each containing a single subject heading. Six rules (detailed in Appendix 1) make up the phrase rearrangement algorithm and are applied in turn to each of the single subject heading phrases.

Those phrases which satisfy one of the first four conditions have an invariant modification order. For instance, if a subject

---

Aroclor 5460
adhesives from vinyl butyral polymers containing Castorwax and, for paper terephthalic acid polyethylene ester laminates, P 14027v supporting components, for machining of thin walled articles, P 113297g

Aromaticity
R 30774j, R 30775k
of alkoxyboroxins, 127059t
of annulene polyoxides, 3280p
of benzene, Faraday effect in relation to, 131863t
chem. definition of, quantum chem. approach to, 103921s
of cycloheptadecapentaenetriyne derivatives, 66572f
of cyclopentadiene, i.r. spectrum in relation to, 116341c
of dihydrodibenzodiazocine, 12696f
hetero-, of diazepines, 78994h
of isopropylmethylphenylphosphole, 110645h
of 5-membered heterocycles, 89603f
of metal complexes, magnetooptical rotation in relation to, 131869z
of metal pentanedione complexes, 8768p
of nonbenzenoid conjugated cyclic hydrocarbons, R 31481y
of phenylene carbonate, 94900c
quasi-, of cobalt complexes, 89437c

Aromatization
of androst-4-en-3-one derivatives, 12952m
of C_6-isomers, mechanism of catalytic, 78210z
of cyclohexane-heptane, 81060f
of ethanothebainone methine derivatives, mechanism of, 3607a

Fig. 1. Example from 'Subject Index to Chemical Abstracts, Vol. 72'

```
COAL> FOR OPEN <GRATES># WHEELER R.V. 43 ##
SHOCK>TEMPERATURES IN <COAL MEASURES> OF <GREAT BRITAIN># GRAHAM J.I. 1 ##
DUST> CONTROL IN COAL MINES IN <GREAT BRITAIN> # MAGUIRE B.A. 3 ##
INFLUENCE OF RANK OF <COAL> ON RESPIRABLE<DUST> IN COAL MINES IN <GREAT BRITA
IN># RAMSAY H.T. 4 ##
MANNER AND FREQUENCY OF <IGNITION> OF <FIREDAHP> IN MINES IN <GREAT MINES IN <GREAT BRITA
IN># RAMSAY H.T. 2 ##
COAL-DUST EXPLOSIONS> IN MINES IN <GREAT BRITAIN># SHRE 18 ##
RECENT EXPERIMENTS IN <COAL-DUST EXPLOSIONS> IN <GREAT BRITAIN># TIDESWELL F.V
. 16 ##
COMPOSITION AND PATHOGENICITY OF AIRBORNE <DUST> IN MINING <EXPLOSIVE>. IN <GREAT
BRITAIN>#BERGHAM I. 1 ##
<DEFLAGRATIONS AND TRANSMISSION OF <DETONATION> IN MINING <EXPLOSIVE>. IN <GREAT
BRITAIN>#PLANT J. 1 ##
<FIRST AID> ORGANISATION AT COLLIERIES IN <GREAT BRITAIN># CRONIN A.J. ##
HAZARDS FROM MINE <EXPLOSIONS> IN<GREAT BRITAIN># TIDESWELL F.V. 24 ##
RESEARCHES ON <EXPLOSIONS> IN MINES IN <BELGIUM>,<FRANCE>,<U.S.A>, B<GREAT BRIT
AIN>;#COWARD H.F.##
RECENT RESEARCH INTO <SILICOSIS> IN <GREAT BRITAIN>#BERGHAN I. 2 ##
FAILURES IN <WIRE ROPES> IN COLLIERY PRACTICE IN <GREAT BRITAIN>#MCCLELLAND A.E.
'USE OF-<WIRE ROPES> IN MINES AND QUARRIES IN <GREAT BRITAIN>#GREENOUGH G.K.  10
<GROUND>#SEE ALSO FALLS OF GROUND#
<EXPLOSIVES> AND USE IN BREAKING <GROUND>#RITSON J.A.S. 3 ##
<GROUND> FAILURE AROUND <EXCAVATIONS>#DINSDALE J.R. 1 ##
MOVEMENTS OF <GROUND> IN ADVANCE OF <LONGWALL WORKING>#MUDSPETH H.M. 3 ##
MOVEMENTS OF <GROUND> IN MINES#MCTRUSTY J.H. 1 ##
```

Fig. 2. Input phrases to articulated subject index programs

heading is the first part of the phrase, the rest of the phrase becomes the modification with the original order retained. The citation order of the modifications of phrases which do not obey the first four conditions is influenced by the content of other phrases with the same subject heading. These are said to have a variant structure. Referring to Fig. 3 and the subject heading GREAT BRITAIN, the first index entry is derived from the invariant phrase:

ROCK TEMPERATURES IN COAL MEASURES OF
⟨GREAT BRITAIN⟩

The subject heading is preceded by the function word 'of' (see Rule 4, Appendix 1). The next three entries are all derived from variant phrases and these are indicated with an * in Fig. 3. A frequency list is used to decide the order of the modifications derived from variant phrases. The contenders for the first components of modifications from these phrases, together with the chosen first components of modifications from invariant phrases make up this list. Only phrases which are to appear under the same subject heading are to influence the decision process. The list generated for the subject heading GREAT BRITAIN would include

AIRBORNE DUST	1
COAL	1
COAL MEASURES	1
COAL MINES	3
COLLIERY PRACTICE	1
DUST CONTROL	1
INFLUENCE	1
RANK	1
RESPIRABLE DUST	1

the figures indicate the frequencies of occurrence.

'COAL MINES' is chosen to be the first element of the second, third and fourth modifications of GREAT BRITAIN (see Fig. 3). In each variant phrase the chosen component is that which occurs most frequently. This choice gives the index display a structure which aids the searcher. The display is, consequently, context-sensitive, because the word order of index entries can

---

```
GRATES   OPEN --, COAL FOR, WHEELER R.V. 43

GREAT BRITAIN
       COAL MEASURES OF, ROCK TEMPERATURES IN, GRAHAM J.I. 1
       COAL MINES IN, DUST CONTROL IN, MAGUIRE B.A. 3
*      COAL MINES IN, INFLUENCE OF RANK OF COAL ON RESPIRABLE DUST IN, CARVER J. 2
                       MANNER AND FREQUENCY OF IGNITION OF FIREDAMP IN, RAMSAY H.T. 4
       COAL-DUST EXPLOSIONS IN MINES IN, SHRE 18
                       RECENT EXPERIMENTS IN, TIDESWELL F.V. 16
       COMPOSITION AND PATHOGENICITY OF AIRBORNE DUST IN, BERGMAN I. 1
       DEFLAGRATION AND TRANSMISSION OF DETONATION IN MINING EXPLOSIVES IN, PLANT J. 1
       FIRST AID ORGANISATION AT COLLIERIES IN, CRONIN A.J.
       HAZARDS FROM MINE EXPLOSIONS IN, TIDESWELL F.V. 24
       RESEARCHES ON EXPLOSIONS IN, COWARD H.F.
       RECENT RESEARCH INTO SILICOSIS IN, BERGHAN I. 2
       USE OF WIRE ROPES IN COLLIERY PRACTICE IN, FAILURES IN, MCCLELLAND A.E. 5
                       MINES AND QUARRIES IN, USE OF, GREENOUGH G.K. 10

GROUND   SEE ALSO FALLS OF GROUND
       BREAKING --, EXPLOSIVES AND USE IN, RITSON J.A.S. 3
       FAILURE AROUND EXCAVATIONS, DINSDALE J.R. 1
       MOVEMENTS OF, IN ADVANCE OF LONGWALL WORKING, MUDSPETH H.M. 3
                       IN MINES, MCTRUSTY J.H. 1
```

Fig. 3. Computer generated articulated subject index

---

change, as may happen, for instance, if indexes are merged into cumulated editions.

Design of the subject index production system

One of the advantages of the KWIC index technique is the comparative simplicity of the algorithm; the design of a typical program can be broken down into the stages shown in Fig. 4. The simple algorithm involves an input stage, the creation of an inverted file, an alphabetic sort and finally a printing stage. The creation of the ASI display, on the other hand, involves more complex processing; for example, two sorting operations are needed. A flowchart of the process is shown in Fig. 5 and a detailed flowchart appears in Appendix 2. Initially a file of records, containing subject headings and their associated phrases, is created and these records are sorted into alphabetic order of subject headings. Only then can modifications from variant phrases be generated because of the context-dependency described above. The phrase rearrangement algorithm is then applied and modifications are created. These are sorted into alphabetic order within subject headings and the print stage completes the process.

The title-like phrases are initially input and subjected to format checks. An inverted file is then created and a number of records are generated from each phrase, each containing a subject heading, a subject heading sort key and the phrase with the

1. Input document titles.
2. Create index entries comprising a keyword and context.
3. Sort index entries by keyword.
4. Print index.

Fig. 4. Keyword in context index production

1. Input stage.
2. Format checks.
3. Create index entries comprising a subject heading, a subject heading sort key and the phrase.
4. Sort index entries into subject heading order.
5. Analyse each phrase to produce a list of elements using the function word list (of, by etc.) until the subject heading changes.
6. Rearrange the list of elements into the order required for printing, if a phrase is invariant. Generate a modification sort key. Add the first element of the modification to the frequency list.
7. Add the possible first elements of the modification to the frequency list, if a phrase is variant.
8. When an index entry with a new subject heading is encountered rearrange the variant phrase element lists and create a modification sort key for each.
9. Alphabetise the modification sort keys.
10. Print the subject heading followed by the modifications introducing indentation when necessary.
11. Go to rearrange the index entries with the next subject heading.

Fig. 5. Stages in articulated subject index production

subject heading removed. Thus n records are generated from a phrase containing n subject headings. Sort key generation is a simple process in which the subject heading is placed in a fixed length, zero filled area. No special account is taken of non-alphanumeric characters and it is recognised that in certain applications, where, for example, chemical nomenclature is used, a more sophisticated sort key generation routine will be needed. The organisation of the modification display cannot be undertaken until all phrases with the same subject heading are together on the file. The phrase rearrangement algorithm can then be applied to each phrase in turn in order to create the required modifications.

The application of this algorithm presents some interesting implementation problems. As mentioned above, each index phrase is analysed by comparing those words that are not part of subject headings against a list of function words. The look-up table which contains these words has a tree structure. The number of words in the table is small, typically 20 to 30, and hence the relatively high storage demands of the tree are acceptable.

A string of fixed length elements is generated during the phrase analysis, and each element contains length, type and position data of each word or word group within the phrase. There are six element types and these are:

H—subject heading.
P—articulating point
M—the rest of a word group when a subject heading is removed
N—a word group bounded on one or both sides by articulating points
Q—the word OF
R—the reference—the code leading to the bibliographic details

The following input phrase:

⟨DUST⟩ CONTROL IN COAL MINES IN
⟨GREAT BRITAIN⟩, MAGUIRE B.A.(3)

after analysis and removal of the heading GREAT BRITAIN appears as follows:

DUST CONTROL IN COAL MINES IN, MAGUIRE
B.A.(3)

A string of six elements is generated to describe it:

N13 A1 : P3 A2 : N11 A3 : P3 A4 : H2 A5 : R16 A6

N13 A1 describes DUST CONTROL which is a type N element with a length of 13 characters and a character address A1 within the phrase. The fixed length elements can be rearranged into the order required by the algorithm (see Appendix 1) and the modification can then be constructed when required.

The new order of the elements becomes:

N11 A3 : P2 A4 : H2 A5 : N13 A1 : P2 A2 : R16 A6

and when printed the entry becomes:

GREAT BRITAIN
COAL MINES IN, DUST CONTROL IN, MAGUIRE
B.A. (3)

Thus, N11 A3 is 'COAL MINES'; P2 A4 is 'IN'; H2 A5 is ' ,';
N13 A1 is 'DUST CONTROL'; P2 A2 is 'IN';
R16 A6 is ', MAGUIRE B.A. (3)'

A terminal space is removed by reducing the respective element count by one. The comma and space preceding the reference, in this case an author, are added to the reference and the address is modified.

The processing of variant phrases adds to the complexity of the system because the elements of the analysis cannot be rearranged immediately. Instead, the contenders for first position in the modification must be entered into the frequency list described above, and both the phrase and its analysis must be retained. The frequency list has a letter-table structure giving a maximum of 64 forward-threaded lists, or one for each

character. When the order of a particular modification has been decided, a sort key is generated which is used to alphabetise the modifications within each subject heading. Function words are subordinated in the sort so that modifications beginning with, for example, COMPUTERS OF and OF COMPUTERS sort together, although the distinction is retained in printing. The final stage of the process is the display organisation which is at present via line-printer with line-width and page-size options.

Use of storage

The manipulation of variant phrases involves the use of dictionary look-up techniques and necessitates a double sorting operation. A number of program strategies are thus possible. An important consideration is that the programs can be implemented on simple machine configurations, e.g. without direct access facilities. The initial sort of records into subject heading order involves the use of magnetic tape files. However, two approaches can be made to the problem of sorting the modifications. The file is already partitioned by subject heading and providing the size of the resulting sections is not too large, the secondary sort can be performed in core. Otherwise a second file is created which is sorted in a separate operation.

The former strategy, i.e. a core sorting technique, can best be described by using a set of four phrases P_1-P_4 with a common subject heading as an example. Assume that P_1 and P_4 are invariant in form and that P_2 and P_3 are variant. The main storage areas and their contents are summarised in **Fig. 6**. Initially, phrase P_1 is analysed and the resulting analysis A_1 is put into area CLASS. The analysed phrase AP_1 appears in the area FINAL: the storage areas are summarised in the following diagrams. An arrow in a diagram indicates a stored address and its position in core.

| CLASS | A_1 |

| FINAL | AP_1 |
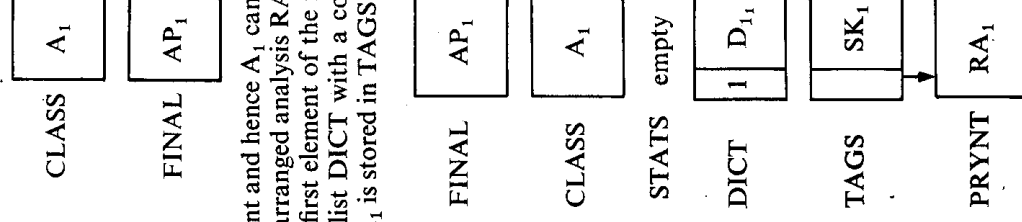
Phrase P_1 is invariant and hence A_1 can be rearranged immediately to form the rearranged analysis RA_1, which is stored in the area PRYNT. The first element of the modification D_{1_1} is put into the frequency list DICT with a count of one. The modification sort key SK_1 is stored in TAGS. The storage areas then become:

| FINAL | AP_1 |

| CLASS | A_1 |

| STATS | empty |

| DICT | 1 | D_{1_1} |

| TAGS | SK_1 |

| PRYNT | RA_1 |

CLASS	Elements of the phrase analyses.	Each phrase produces a variable length list of elements.
FINAL	The phrases after analysis.	Phrases have variable length.
DICT	Possible first elements of modifications of invariant phrases and chosen first elements of modifications of invariant phrases.	Element has a variable length.
TAGS	Modification sort keys.	Each sort key has a fixed length.
STATS	Addresses of possible first elements of modifications, both in DICT and in CLASS.	Number of possible first elements is variable.
PRYNT	Phrase analysis elements after rearrangement into order required for printing.	See CLASS.

Fig. 6. Main storage areas

The magnetic tape record containing the phrase P_2 is input and the phrase is analysed. However, it has a variant form and hence cannot be rearranged immediately. Assume there are three possible first components (D_{21}, D_{22}, D_{23}) of the modification. These are entered in the frequency list and their positions in this list and the positions of the elements of the analysis are stored in STATS. The storage areas then become:
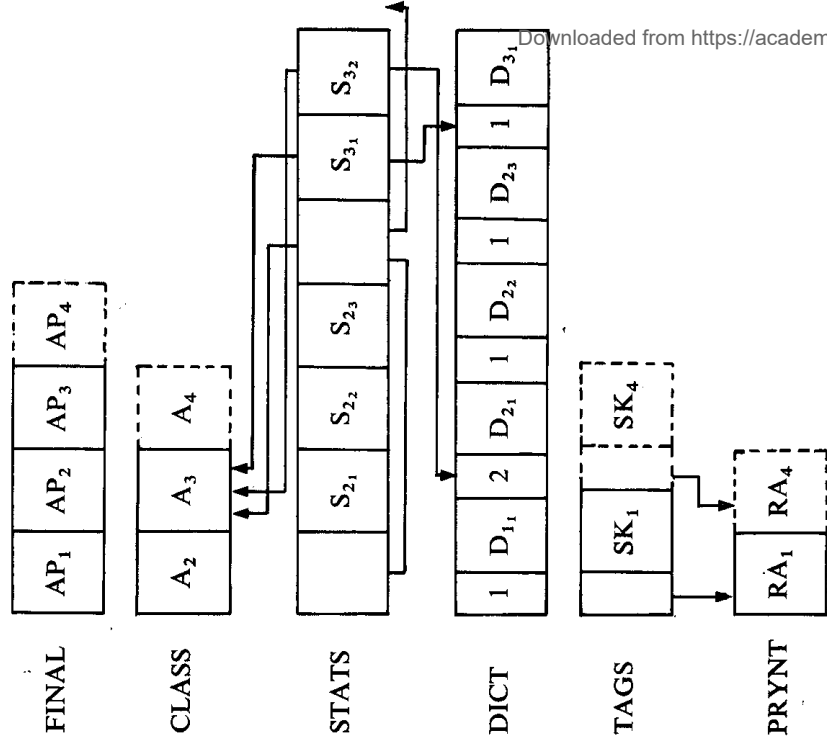
FINAL	AP_1	AP_2		
CLASS	A_2			
STATS	S_{21}	S_{22}	S_{23}	
DICT	1 D_{11}	1 D_{21}	1 D_{22}	1 D_{23}
TAGS	SK_1			
PRYNT	RA_1			
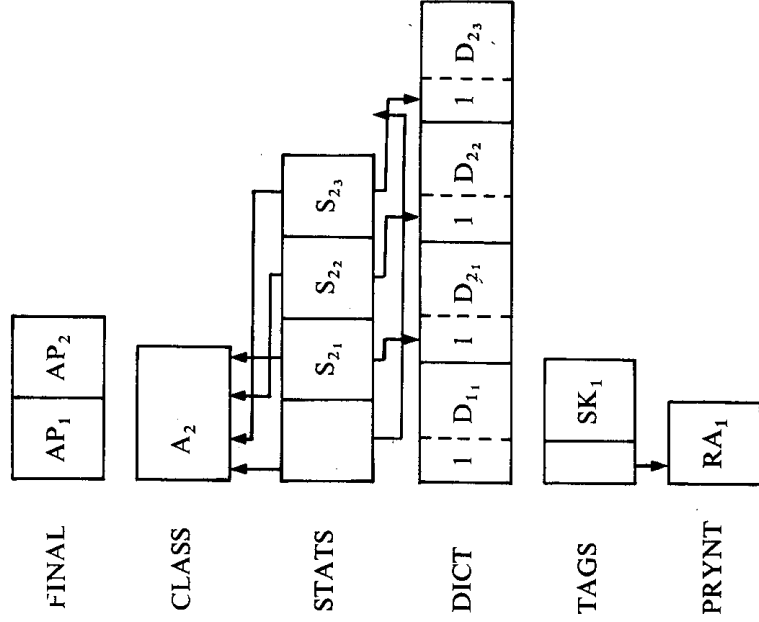
Phrase P_3 is similarly processed. It is variant; assume that there are two possible first components of the modifications, D_{31} and D_{32} but D_{33} is identical with D_{21}.

Ignoring the areas enclosed with dotted lines, the areas become as in **Table 1.**

P_4 is finally input and processed. This phrase is invariant and information relating to it is added to the core storage as indicated by the dotted lines in the previous diagram.

There are no further phrases to be processed for the current subject heading and hence the variant phrase analyses of P_2 and P_3 must now be rearranged. The address list S_{21}, S_{22}, and S_{23} enables the occurrence counts of D_{21}, D_{22} and D_{23} to be compared. D_{21} occurs more frequently than D_{22} or D_{23} and hence is chosen to be the first element of the modification for phrase P_2. Similarly D_{32} (D_{21}) occurs more frequently than D_{31}, and hence the same element is chosen in each case. Sometimes

Table 1

FINAL	AP_1	AP_2	AP_3	AP_4	
CLASS	A_2	A_3	A_4		
STATS	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}
DICT	1 D_{11}	2 D_{21}	1 D_{22}	1 D_{23}	1 D_{31}
TAGS	SK_1	SK_4			
PRYNT	RA_1	RA_4			

the elements occur with equal frequencies and in such cases the order of the modification is the same as the original order of the phrase.

The elements of the analysis A_2 and A_3 can now be rearranged, and the sort keys SK_2 and SK_3 formed thus:

TAGS	SK_1	SK_4	SK_2	SK_3
PRYNT	RA_1	RA_4	RA_2	RA_3
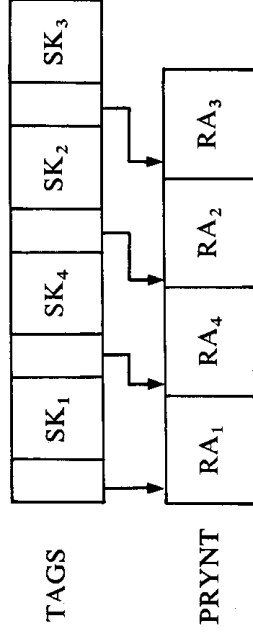
The sort keys in area TAGS are then sorted into alphabetic order and finally the subject heading, with its related modifications in alphabetic order, is printed.

The strategy just described creates the index entries entirely within core and may, on occasion, lead to the restrictions described above. The storage of the data can make heavy demands on core storage and hence there is a limit to the numbers of phrases with a common subject heading, that can be processed.

The alternative approach using the secondary magnetic tape file (tape 3, Appendix 2) mentioned above involves the removal of invariant phrase information from core as soon as it is processed. However, the frequency list and the information relating to variant phrases are retained in core. Once variant phrases have been rearranged, these records are also written to tape three. The procedure is repeated for all the subject headings and phrases; hence a partially ordered file, consisting of subject heading records followed by their associated modification records, is generated. A running number is used as a primary sort key to retain the predefined order of the file when alphabetising the modifications. Standard software which takes account of partial ordering of file is used for sorting.

In practice the use of the secondary magnetic tape file in the

ASI system reduces the core size of the program or, alternatively, increases the number of phrases that can be processed. The limiting factor in this case is the number of variant phrases that can be stored.

Summary

The production of high quality articulated subject indexes is no longer the arduous task that it was formerly, for, as is shown in this system, the machine can take over many of the functions that were formerly performed manually. Certain operations that were hitherto regarded either as intuitive or intellectual have been formalised and hence mechanised.

The ASI program suite has been used to produce indexes in several production environments with output via line printer. A valuable elaboration would be to interface them with a computer typesetting system. A larger character set would bring out the full potential of the structure of the display and hence further aid the searcher in using the index.

The ASI production system uses a simple language analysis in its procedures. The function that it performs are limited and are less ambitious than certain sophisticated language-processing systems, but it has already lead to a useful working system.

Note

Information on the availability and commercial exploitation of the programs may be obtained from the National Research Development Corporation.

4. Is the subject heading preceded by an 'of'?

The phrase is rearranged to form a modification in which the noun phrase preceding the 'of', together with the 'of', are placed at the beginning of the modification. These are followed by that part of the phrase which follows the subject heading and finally the rest of the phrase from the beginning up to the noun phrase which is the first part of the modification.

5. Does one of the noun phrases in the derived phrase occur more frequently than any other in the derived phrases with the same subject heading?
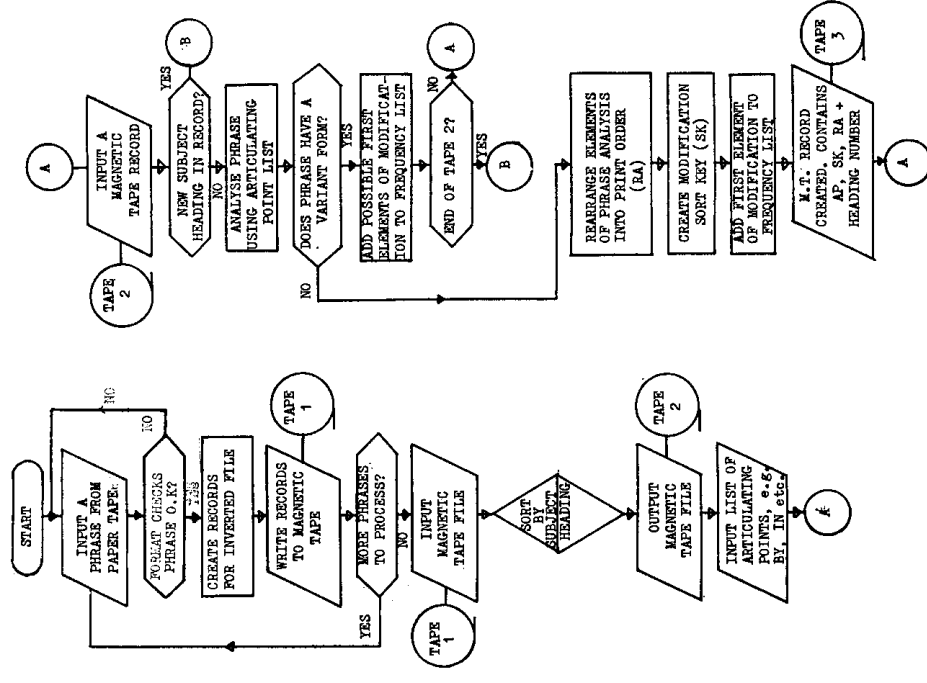
Any of the elements (i.e. type N or M) preceding, and the one immediately following the subject heading can be placed at the beginning of the modification. These elements are placed in a frequency list. All variant phrases with the same subject heading are treated similarly. The first elements of modifications generated from invariant phrases (i.e. invariant phrases with the same subject heading) are also entered into the frequency list. When all the phrases, both variant and invariant, with the same subject heading, have been processed, the order of the modifications to be derived from variant phrases can be decided. If any contender for first place in the modification occurs more often than any other contender in the frequency list, that contender is chosen.
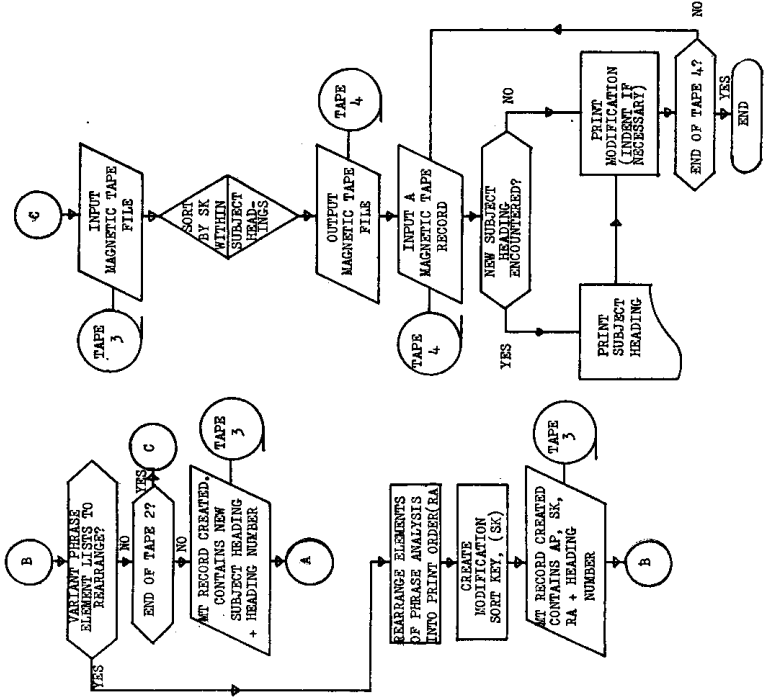
6. Do rules 1 to 5 fail?

If rule 5 fails, because possible candidates for first place in the modification occur equally frequently, the modification assumes the same order as the original phrase.

Appendix 1

Index entry generation algorithm

The index entry generation algorithm comprises a series of rules which are applied to each derived phrase. These rules are applied in a systematic way until one rule applies; they are formulated in terms of questions. A phrase satisfies a particular rule if the answer to the question is affirmative. Phrases which can be rearranged using any of rules 1 to 4 are called 'invariant' because they change the order of a modification, formed by the application of one of the above rules, is fixed. Other phrases are called 'variant'.

1. Is the subject heading the first component of the phrase?

The phrase is rearranged to form a modification in which the modification follows the subject heading in its original order.

2. Is the subject heading only part of a noun phrase?

The phrase is rearranged to form a modification in which the rest of the noun phrase containing the subject heading is placed at the beginning of the modification, then that part of the phrase which follows the subject heading and finally that part of the phrase preceding the subject heading.

3. Is the subject heading followed by an 'of'?

The phrase is rearranged to form a modification in which that part of the phrase following the subject heading is placed at the beginning of the modification. The part of the phrase preceding the subject heading then follows.

Appendix 2

Flowchart for articulated subject index programs (with magnetic tape secondary sort)

START → INPUT A PHRASE FROM PAPER TAPE → FORMAT CHECKS PHRASE O.K? → NO (loop back); CREATE RECORDS FOR INVERTED FILE → WRITE RECORDS TO MAGNETIC TAPE (TAPE 1) → MORE PHRASES TO PROCESS? → YES (loop back), NO → INPUT MAGNETIC TAPE FILE (TAPE 1) → SORT BY SUBJECT HEADING → OUTPUT MAGNETIC TAPE FILE (TAPE 2) → INPUT LIST OF ARTICULATING POINTS, e.g. BY, IN etc. → A

A → INPUT A MAGNETIC TAPE RECORD (TAPE 2) → NEW SUBJECT HEADING IN RECORD? → YES → B; NO → ANALYSE PHRASE USING ARTICULATING POINT LIST → DOES PHRASE HAVE A VARIANT FORM? → YES → ADD POSSIBLE FIRST ELEMENTS OF MODIFICATION TO FREQUENCY LIST; NO → END OF TAPE 2? → NO → A; YES → B

B → REARRANGE ELEMENTS OF PHRASE ANALYSIS INTO PRINT ORDER (RA) → CREATE MODIFICATION SORT KEY (SK) → ADD FIRST ELEMENT OF MODIFICATION TO FREQUENCY LIST → M.T. RECORD CREATED, CONTAINS AP, SK, RA + HEADING NUMBER (TAPE 3) → A

References

ARMITAGE, J. E., and LYNCH, M. F. (1968). Some structural characteristics of articulated subject indexes, *Information Storage and Retrieval*, Vol. 4, pp. 101-111.

LYNCH, M. F. (1969). Computer-aided production of printed alphabetical subject indexes; A review. *Journal of Documentation*, Vol. 25, pp. 244-252.

Errata

There is an error in the paper 'A rapidly convergent iterative method for the solution of the generalised nonlinear least squares problem' by D. R. Powell and J. R. Macdonald (this *Journal*, Vol 15, No. 2, p. 148). The formula on the second line of Table 8 (p. 154) should read

$$y = \alpha_1(1 + \alpha_2^{-1}\alpha_3)^{-1/\alpha_3} .$$

Also, there are a few misprints in the paper involving 'a's' where 'a's' should appear. Since 'a' is used for parameter estimates and 'α' for parameter designation, it is hoped these misprints will be obvious from the text and will cause no misinterpretations.

There is an error in the paper 'LR algorithm with Laguerre shift for symmetric tridiagonal matrices' by J. Grad and H. Zakrajšek (this *Journal*, Vol. 15, No. 3, p. 268). On page 262, second column, seven lines from the foot of the page, the line should read:

shift:= **if** v **greater** 0.0 **then** $(n - 1)/(sw + sw * \text{sqrt}(v))$