

## 索引の自動化システム MISA

### MISA: An Automatic Indexing System

安 西 郁 夫

*Ikuo Anzai*

#### *Résumé*

The computer system planned and developed by the author for indexing journal articles is called MISA (**M**echanized **I**ndexing **S**ystem by **A**nzai). The system is composed of the following six segments written in KOS (Keio Operating System) Assembler:

1. INIMAS  
reads and processes initial data, produces Initial Master File, and prints out Sequential List.
2. ALTMAS  
alters the contents of Initial Master File and prints out Altered Sequential List.
3. KWICED  
reads and processes Master File, and produces unsorted KWIC File.
4. KWOCED  
reads and processes Master File, and produces unsorted KWOC-type File.
5. SORPRI  
sorts KWIC or KWOC-type File and prints out KWIC Index, KWOC Index, Author Index and/or WADEX.
6. WORDFR  
counts the frequency of word occurrence and prints out a frequency table.

A usual automatic indexing system prepares a separate bibliography in addition to a KWIC or KWOC index. And for detailed bibliographic information, a user must look up the bibliography. In MISA each index itself gives necessary data in a simplified form making it unnecessary to look up a separate bibliography.

The machine configuration required for the operation of MISA are as follows:

TOSBAC 3400	16K Words
Card Reader	1
Line Printer	1
Magnetic Tape Units	4

(Keio University Mita Library and Information Center)

序

- I. システムの構成と特色
- II. INIMAS の機能
- III. ALTMAS の機能
- IV. KWICED の機能
- V. KWOCED の機能
- VI. SORPRI の機能
- VII. WORDFR の機能
- VIII. 辞書のリスト処理

序

本稿は、筆者が慶応義塾大学大学院図書館・情報学専攻に提出した論文に基づき、筆者の設計した自動索引システム MISA (Mechanized Indexing System by Anzai) を概説したものである。

H.P. Luhn が KWIC を創始して以来、各種索引の自動作成が試みられ、多くのシステムが開発されてきた。その盛況は百花斉放にもたとえられよう。今日、一つの自動索引システムを設計することは、まさに山頂に 1 個の小石を積むに等しい。とはいえ、従来の自動索引システムは主として計算機専門家の手によって作られたものであり、図書館・情報学の立場から見た場合には、必ずしも満足すべきものではない。とすれば、IR サイドの者が経済的でコンパクトなシステムの設計を試みることに、それなりの意義が認められよう。

MISA の作成に当って使用した計算機は、慶応義塾大学工学部中央試験所計算センター（現情報科学研究所）の TOSBAC 3400 であり、使用したプログラミング言語は KOS (Keio Operating System) の統制下にあるマクロ・アセンブリ言語 KMAP である。

筆者が MISA を設計しえたのは、図書館・情報学科前主任橋本孝教授ならびに現主任澤本孝久教授の御配慮により、大学院において研究する機会を与えられたからであり、また、indexing について中村初雄教授から、計算機科学について浦昭二教授から御指導を得たからに他ならない。KMAP 使用上の技法については学友三輪行雄氏（三菱総合研究所）と小川昭易氏（情報科学研究所）から多くの御教示を受け、計算機の使用については大西邦夫氏（情報科学研究所）の御好意にあずかった。学友細野公男、渡部満彦、小沢弥の諸兄からは終始変ることなきサポートを受けた。また、本研究に対して慶応義塾

からは学事振興資金による研究助成金を与えられた。MISA の設計はこれらの方々によくを負っている。ここに記して深甚なる謝意を表したい。

I. システムの構成と特色

A. 構成

自動索引システム MISA は、1969 年 12 月 9 日現在の version で 3177 ステップからなるプログラムであるが、このプログラム全体を一度に主記憶装置に load すると作業用エリアが不足するので、6 個の loading segments に分割されている。それらセグメントの名称と機能は次のとおりである。

① INIMAS

文献データを入力して整形し、Initial Master File を作成するとともに、Sequential List を印刷する。

② ALTMAS

Master File の内容を変更（追加、削除、取替）するとともに、新しい Sequential List を印刷する。

③ KWICED

Master File を入力して整形し、未分類の KWIC File を作成する。

④ KWOCED

Master File を入力して整形し、未分類の KWOC File を作成する。

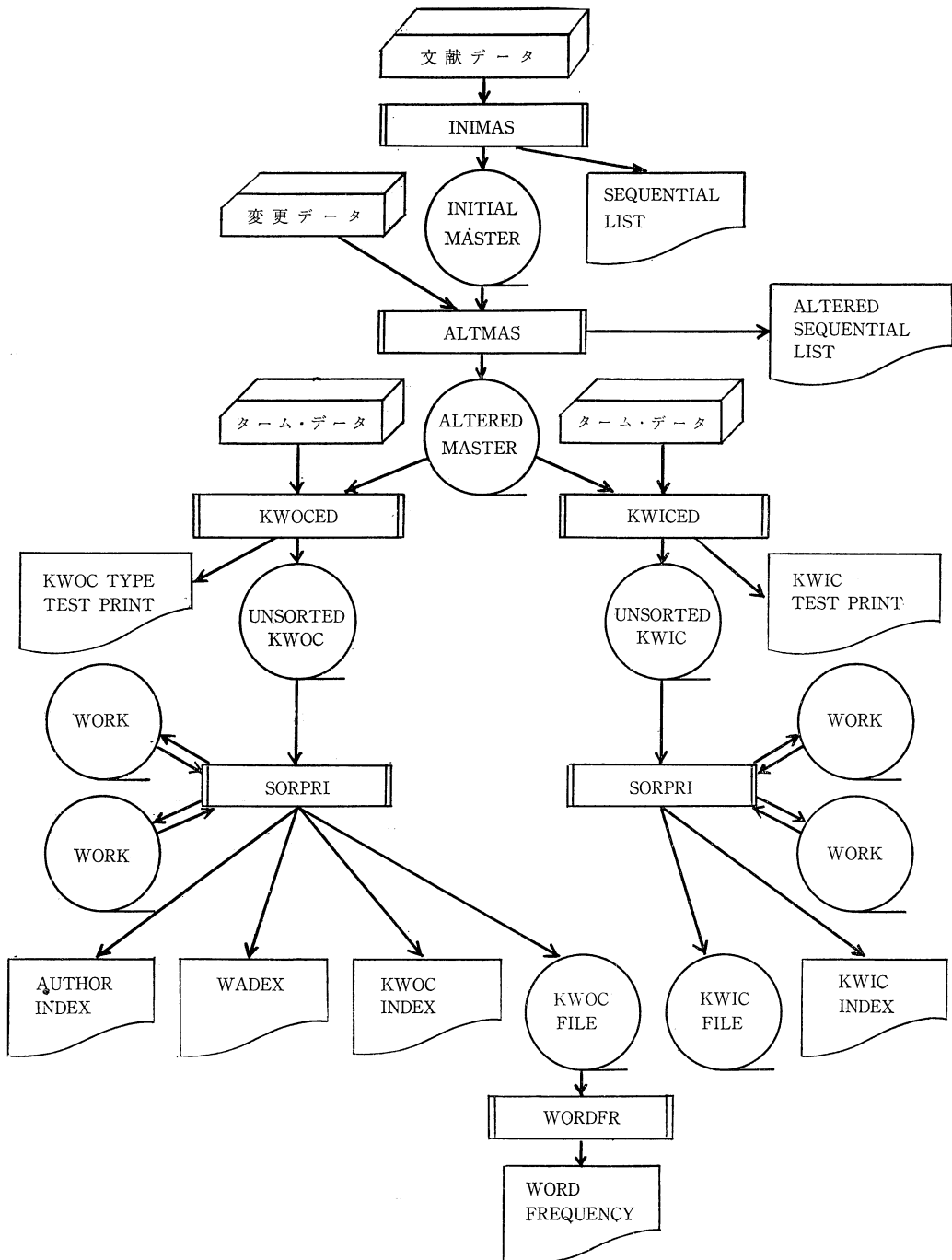
⑤ SORPRI

未分類の KWIC File または KWOC File を入力して分類し、KWIC Index, KWOC Index, Author Index または WADEX を印刷する。

⑥ WORDFR

分類ずみの KWOC File を入力して、keyword またはすべての語の生起頻度を計算し、ABC 順の

図 1. MISA のシステム・フロー



## 索引の自動化システムMISA

頻度表を印刷する。

システムのフローは図1の示すとおりである。通常自動索引システムでは、KWIC, KWOC等の索引とは別に bibliography または文献リストと呼ばれるものを作成し、索引には論文標題と参照コードのみを与え、論文掲載誌、巻号、ページ等の書誌的データは文献リストに収録し、参照コードによって見られるようにしている。MISAでは、経済的観点から索引と文献リストの2本建制を廃止し、個々の索引中に必要な書誌的データを簡略形で示すようにしている。KWIC Indexの場合でも著者名(第1著者のみ)を付随させている。この“1本建制”がMISAの大きな特色となっている。

### B. ローディング方式

MISAを構成する6個のセグメントは、オブジェクト・プログラムとして磁気テープに記憶されているが、このような構成を持つプログラムをロードして実行するには、monitor方式をとる場合と、overlay方式による場合とがある。

Monitor方式では、システムを統制するmonitorがまずロードされて、コントロール・カードを解読し、テープの中から指定されたセグメントを探し出してロードし、テープを捲き戻す。第2のジョブ指定があれば、またテープを廻して必要なセグメントを探し出してロードする。Monitorはその間主記憶装置内に常駐している。

Overlay方式には統制用monitorがなく、第1セグメントがまずロードされ、実行が終ると次のセグメントがロードされるという風に、セグメントは順次ロードされる。

システム・フローを見ても明らかなように、MISAによって索引を作成したり、頻度表を印刷したりする場合には、必ず一定の順序に従っており、あるセグメントを実行した後で、逆戻りしてそれ以前のセグメントを実行するということがない。逆戻りや、単一セグメントの選択実行が多い場合にはmonitor方式をとらねばならないが、MISAではその必要がないので、overlay方式を採用している。ただし、ジョブによっては不必要なセグメントがあるので、その場合は、コントロール・カードの指定によって、直ちに次のセグメントをロードできるようにしてあることはいうまでもない。

### C. 使用機器

MISAの実行に必要な機器は次のとおりである。

本体 TOSBAC 3400 16K 語  
カード・リーダー 1台

ライン・プリンター 1台  
タイプライター 1台  
磁気テープ装置 4台

## II. INIMASの機能

カードにパンチされた文献データを読み込み、そのsequenceを検査しながら作業エリアに転送し、必要な整形を施して論理レコードを形成し、ブロックして磁気テープ#1にInitial Master Fileを作成する。さらに、必要があれば、作成したFileを読んでSequential List(入力文献の一貫番号リスト)を印刷する。

### A. 入力データ

MISAの原始データはindexingの対象となる雑誌記事の書誌的事項であるが、INIMASはこの原始データを入力データとする。データはIBM 80欄カードにパンチされる。カード枚数は可変であり、1文献につき2枚ないし4枚が使われる。データのformatは図2のとおりである。

#### 1. レコード番号

図2. 文献データ・フォーマット

欄	1~6	7	8	9~28	29~48	49~68	69~80
項目	レコード番号	ブランク	カード番号	第1著者	第2著者	第3著者	ブランク

欄	1~6	7	8	9~22	23, 24	25~80
項目	レコード番号	ブランク	カード番号	ソース・コード	予備	標題(56字)

欄	1~6	7	8	9~80
項目	レコード番号	ブランク	カード番号	標題続き(72字)

欄	1~6	7	8	9~80
項目	レコード番号	ブランク	カード番号	標題続き(72字)

各文献には、入力順に 6 桁の一貫番号がレコード番号として付与される。仮に累積ファイルを作成する必要がある場合でも、100 万件以内であれば処理できることになる。

この番号は identification code の役割を果たすと同時に、データ・カードの配列検査のキーとなる。この種のコードに対しては、各カードの右端の数欄が割り当てられることが多いが、標題は通常次のカードにまでまたがるから、標題をパンチしながら、レコード番号用フィールドに侵入しないよう絶えず注意しなければならない。このようなパンチ作業上の不利を避けるため、本システムでは先頭のフィールドを配当してある。

## 2. カード番号

1 個のレコードを構成する各カードには順次 0 から 3 までのカード番号が付与される。この番号は、レコード番号と相まってデータ・カードの配列検査を可能にするばかりでなく、レコードの整形作業を容易にする機能を持っている。

## 3. 著者名

著者は 3 名まで記入することができる。各著者名の長さは 20 字以内である。記入形式が転置形であることはいうまでもない。

筆者が作成したデータでは、著者名を構成する要素が 2 個 (first name と family name) の場合にはフルに記入し、middle name 等を含むためその要素が 3 個以上ある場合には、family name 以外は頭文字のみで略記することにしている。

〔例〕

James Smith → Smith, James

James L. Smith → Smith, J. L.

## 4. ソース・コード

通常の自動索引では、標題の後に参照コードが与えられているのみで、必要な書誌的事項を知るためには、参照コードを手がかりとして、別に作られた bibliography を見なければならないが、MISA では、参照コードの代りに 14 桁の source code を付与し、そのコード自体で書誌的事項を知ることができるようにしてある。このコードは、かつて筆者<sup>1)</sup>が論じた簡略化形式に基礎を置くもので、Bishop<sup>2)</sup>の Coden の形式に近い。

ソース・コードは次の 4 要素によって構成される。

- ① 略誌名……4 字。その原則は Coden 方式による。ただし誌名のキー・ワードが 3 語の場合、Coden 方式では、初語と次語の初字および第 3 語の初めの 2

字をとるが、筆者が作成したデータでは、初語の初めの 2 字および次語と第 3 語の各初字をとっている。

② 巻次……2 桁

③ 刊年……西暦年の末尾 2 桁をカッコでくくる。

④ 初頁……論文の最初のページ番号を 3 桁で表示する。

〔例〕

*American documentation*, vol. 20, 1969, p. 143-

AMDO-20 (69) 143

仮に巻次が 3 桁であったり、*Physical review* の如く巻次以外に series があり、例えば ser. 2, vol. 24 を 2/24 と表記することにすれば、4 桁を必要とする。その場合には、予備欄が 2 桁とってあるので、16 桁まで拡張してもよい。

## 5. 標 題

標題の長さのマキシマムは 200 字である。標題が 200 字を越えることはないといっても過言ではない。仮にそのようなことが例外的にあるとしても、恐らくは冗長なサブタイトルを伴っているであろうから、適当にカットすればよい。

標題の末尾には 1 個のブランクに続けて \* (アスタリスク) をパンチしなければならない。また、この記号は、標題の末尾以外で使用してはならない。原標題中にアスタリスクが使用されている場合には (脚注参照符合として使われることがある)、取り除かねばならない。

1 文献当りのカード枚数が可変であるため、レコードの区切りを表示するものとして、このアスタリスクは処理上必要なものであるが、アウトプットにもそのまま印刷されて、標題の末尾を明示する。これは循環形の KWIC 索引において特に必要な措置である。

NEAC の IR システム<sup>3)</sup>では、終了マークとして '=: ' を KWIC で使用している。ALISS<sup>4)</sup>では、'\*' を用いているが、印刷した時に目立ちすぎる嫌いがないでもない。

## B. コントロール・カードとオプション

コントロール・カードの format は表 1 のとおりである。

表 1. INIMAS コントロール・カード

Col. 1 ~	13 ~	25 ~
INIMAS	SKIP NO PRINT [PRINT]	mm/dd/yy

第1フィールドには必ず INIMAS というセグメント名をパンチしなければならない。

このセグメントを使用しない場合には、第2フィールドに SKIP とパンチすればよい。

このセグメントのアウトプットに Sequential List があるが、これを印刷しない場合には、第2フィールドに NO PRINT と指定する。印刷する場合には、同じフィールドに PRINT とパンチしてもよいし、ブランクのままでもよい。

印刷の場合は、さらに第3フィールドに8桁以内で日付をパンチする。筆者は 12/09/69 の如く、月/日/年の形式を用いているが、別の形式であっても8桁以内であればよい。この日付は Sequential List の各ページの上部に印刷される。もしも第3フィールドが空白のままであれば、日付は勿論印刷されない。

### C. セグメントの作業

#### 1. コントロール・カードの解説

作業の第1はコントロール・カードの解説である。

SKIP 指定があれば、直ちに operating system の loader routine を実行して、次のセグメントである ALTMAS をロードする。

NO PRINT の指定があれば印刷用スイッチを0に、ない場合は1にセットし、第3フィールドの日付を heading 用エリアに転送し、'SET MTI' というメッセージをタイプライターで打ち出し、pause 状態に入る。コントロール・カードに誤りがあれば、'CONTROL CARD ERROR. RESET CARDS.' というメッセージをタイプアウトして pause 状態に入る。正しいコントロール・カードを置き直して、タイプライターを carriage return すると、セグメントの始点に戻って実行を再開する。ただし、この場合カードを1枚カラ読みさせている。Operating system がカードを1枚先読みしているので、その分を吐き出させるためである。

#### 2. 配列検査

コントロール・カードの処理が終ると、outfile (出力テープ #1) を open し、データ・カードの読み込みを開始し、レコード・シーケンスとカード・シーケンスが正しいかどうかをチェックする。前者に誤りがあれば、'REC SEQ ERROR. RESET CARDS.' というメッセージが、後者に誤りがあれば、'CARD SEQ ERROR. RESET CARDS.' というメッセージをタイプアウトする。この場合にも、配列を正してデッキをカード・リーダーにセットし直し、タイプライターの C/R (carriage

return) キーを押せばよい。

#### 3. レコードの整形とテープへの出力

シーケンスが正しければ、カード・イメージを順次作業エリアに転送し、不要の部分を切り捨てて繋ぎ合わせる。その一方、標題の末尾記号を探し、発見すれば標題の字数を計算して、論理レコード中の標題部の直前に記録する。最後のカードに末尾記号がない場合には、'RECORD END MARK NOT FOUND. EXECUTION SUSPENDED.' というメッセージを打ち出し、実行を打ち切る。

カード枚数が2または3の場合には、作業エリアに残っている直前のレコードを消去するために、ブランクを埋め込んだうえで、テープ用バッファに整形済みの1論理レコードを転送する。論理レコードのフォーマットは図3のとおりである。

図 3. 論理レコードの構成

語数	1	5	5	5	4	1	50
項目	レコード番号	第1著者名	第2著者名	第3著者名	ソース・コード	標題字数	標題

1 論理レコードの長さは 71 語であり、10 論理レコードをブロックして 1 物理レコードを形成し、磁気テープに書き込む。バッファは 2 つ用意する。

データ・デッキの末尾には 1~4 欄に '++++' とパンチしたカードを置く。このカードを読み込むと、データ終了符号として用いる @ を作業エリアに埋め込み、必要な回数だけ MT に put する。さもなければ、ファイルを close する命令を出すと、その時点でバッファに残留しているレコードがすべて書き込まれるため、同じレコードが重複してファイルに加えられるからである。

記号@の埋め込み作業の結果、ファイルの末尾には@で埋められたレコードが重複して書き込まれるが、@は 76 (8 進) という高い値の内部コードが与えられているため、ソーティングの段階で常に最後尾に廻されるので、正常なレコードに紛れて印刷されることはない。

#### 4. 文献リストの印刷

Initial Master File の作成が終ると、印刷用スイッチによって印刷の必要の有無を確かめる。

各ページの頭部には heading を 1 行印刷する。Head-

## 出力例 1. Sequential List

MECHANIZED INDEXING SYSTEM BY AN7AI	SEQUENTIAL LIST	10/18/69	PAGE 1
1 SAUNDERS, W.L. THE 'SHEFFIELD REPORT' *	ASPR-21(69)009		
2 RICHNELL, D.T. DEVELOPMENT IN EDUCATION LEADING TO FORMAL QUALIFICATION *	ASPR-21(69)013		
3 WILSON, T.D. THE HSC DEGREE IN INFORMATION SCIENCE AT NEWCASTLE UPON TYNE *	ASPR-21(69)018		
4 SCHUR, HERBERT UNIVERSITY OF SHEFFIELD MSC COURSE IN INFORMATION STUDIES *	ASPR-21(69)024		
5 FARRADANE, J. STANDARDS IN EDUCATION IN INFORMATION SCIENCE *	ASPR-21(69)032		
6 HILL, D.W. INFORMATION SERVICES AND MARKET RESEARCH *	ASPR-21(69)059		
7 LEES, R. SOURCES OF INFORMATION ON MARKET RESEARCH USED BY INDUSTRIAL LIAISON CENTERS TO ASSIST SMALL FIRMS *	ASPR-21(69)064		
8 HANDOVSKY, F.L.A. COLLECTION AND UTILIZATION OF ECONOMIC INFORMATION: SOME OF ITS PROBLEMS *	ASPR-21(69)071		
9 HUTCHESON, B.N.P. INFORMATION IN MARKET RESEARCH *	ASPR-21(69)075		
10 ADLER, L.S. CURRENT TRENDS IN GOVERNMENT STATISTICS FOR THE CHEMICAL INDUSTRY *	ASPR-21(69)079		
11 SAWAMOTO T A DETAILED PLAN FOR THE PROPOSED TECHNICAL PROCESSING CENTER OF KEIO UNIVERSITY LIBRARY SYSTEM *	LIIS-06(68)223		
12 CLARKE N F CATALOGING, CLASSIFICATION AND STORAGE OF GOVERNMENT PUBLICATIONS WHEN INCORPORATED INTO THE GENERAL LIBRARY COLLECTION *	LITR-15(66)058		
13 VICKERY, B.C. RESEARCH AT ASLIB *	ASPR-21(69)156		
14 HARTMAN, R.R.K. LINGUISTICS AND TRANSLATION *	ASPR-21(69)190		
15 WRIGHT, J.E. STEPS IN THE DEVELOPMENT OF HS 1000 *	ASPR-21(69)195		
16 LLOYD, G.A. THE IDIC IN ITS INTERNATIONAL ASPECTS *	ASPR-21(69)204		
17 WOOTTON, BARBARA EXPERT AND LAYMAN IN CONTEMPORARY SOCIETY *	ASPR-21(69)225		
18 EARLE, PENELOPE VICKERY, BRIAN SUBJECT RELATIONS IN SCIENCE/TECHNOLOGY LITERATURE *	ASPR-21(69)237		
19 WOOD, D.N.	ASPR-21(69)262		

ing は左端からシステム名、アウトプット名、日付、ページ番号の順になっている。

Heading の次には 2 行分のスペースを置いて detail を印刷する。Detail には 56 行を使用する。

Master File から読み込まれた各論理レコードは、印刷エリアに転送して標題の長さによって 2 行または 3 行に分割して印刷する。印刷される項目は、レコード番号、著者名、ソース・コードならびに標題である。

記号 @ からなるレコードを読み込むと印刷を打ち切り、リストの末尾にレコード数の total を印刷し、ファイルを close するとともに、タイプライターに 'OUTPUT IN MTI' というメッセージを出す。リストを印刷しない場合にもこのメッセージはタイプアウトする。

以上でこのセグメントの作業を終了し、loader によって次のセグメント ALTMS をロードする。(出力例 1 参照)

### III. ALTMS の機能

ALTMS は INIMAS によって作成された Initial Master File の内容をオプション・カードの指定によって変更する機能を持っている。

変更には次の 3 種類がある。

- ① ADD.....File の中間にレコードを挿入したり、末尾にレコードを追加する。
- ② DELETE.....File 中のレコードを削除する。
- ③ REPLACE.....File 中のレコードを新しいレコード

ドと取り換える。

変更されたファイルは磁気テープ #2 に作られる。その内容は、コントロール・カードで指定することにより、Sequential List (Altered) として印刷することができる。

このセグメントは、Sequential List によって発見した Initial Master File 中の誤りを訂正するのが本来の機能ではあるが、追加の機能を持っているので、累積ファイルの作成にも利用することができる。

#### A. コントロール・カードとオプション

コントロール・カードは狭義のコントロール・カードとサブコントロール・カード（以後オプション・カードと呼ぶ）によって構成されている。それらのフォーマットは図4のとおりである。

図 4. ALTMAS コントロール・カード

CONTROL CARD	Col. 1~	13~	25~
	ALTMAS	SKIP NO PRINT [PRINT]	mm/dd/yy
OPTION CARD ①	Col. 1~	13~	25~
	ADD	××××××× (追加の直前の レコード 番号)	×××× (追加するレ コード数)
OPTION CARD ②	Col. 1~	13~	25~
	DELETE	××××××× (削除する先 頭のレコー ドの番号)	×××× (削除するレ コード数)
OPTION CARD ③	Col. 1~	13~	25~
	REPLACE	××××××× (取換えられ る先頭のレ コードの番 号)	×× (取り換える レコード数)
END CARD	Col. 1~		
	+++++		

コントロール・カードの第1フィールドは、必ず ALTMAS とパンチされていなければならない。

このセグメントの実行を必要としない場合には、第2

フィールドに SKIP と指定する。

Sequential List (Altered) を印刷しない場合には、第2フィールドに NO PRINT と指定しなければならない。印刷する場合には、ブランクのままでもよい、PRINT と指定してもよい。印刷する場合には第3フィールドに日付をパンチする。この形式は INIMAS の場合と同じである。

オプション・カード①の第1フィールドは ADD であり、追加（挿入）するレコードの個数を第3フィールドに4桁の数字で指定しなければならない。挿入の場合には、例えばレコード番号 000375 の次に挿入したい場合には、000375 を第2フィールドで指定し、追加の場合には、ファイルの末尾のレコード番号を指定すればよい。このカードの後には、第3フィールドで指定した数字に等しいレコード数のデータ・カードが配列されなければならない。

オプション・カード②の第1フィールドは DELETE であり、第2フィールドには削除する先頭のレコード番号を指定し、第3フィールドには削除するレコードの個数を4桁の数字で指定する。このカードの後にはデータ・カードは続かない。

オプション・カード③の第1フィールドは REPLACE であり、第2フィールドには取り換えられる先頭のレコード番号を指定し、第3フィールドには取り換えられるレコードの個数を2桁の数字で指定する。このカードの後には、第3フィールドで指定した個数に等しいレコードのデータ・カードが続かねばならない。

オプション・カードは、その種別を表わす①②③の順序に配列するのではなく、第2フィールドで指定するレコード番号の正順（小→大）に配列しなければならない。同種のオプションは1回だけしか指定できないというような制限はまったくない。

最後のオプション・カード（データ・カードがある場合は、最後のデータ・カード）の次には‘+++++’をパンチした end card を置かなければならない。

データ・カードのフォーマットその他はすべて INIMAS の場合と同一である。

#### B. セグメントの作業

##### 1. コントロール・カードの解読

セグメント名が正しいかどうかを最初にチェックする。誤っていれば、‘CONTROL CARD ERROR. RESET CARDS.’ というメッセージをタイプアウトする。

SKIP 指定があれば直ちに次のセグメントである KW-



ICED をロードする。その指定がなければ 'ALTMAS LOADED' とタイプライターに出力される。

NO PRINT の指定があれば印刷用スイッチ 0 をにセットする。指定がなければ 1 にセットし、日付を heading 用エリアに転送し、'SET MT 2 (INP), 1' という指示をタイプアウトする。ここでオペレータは Initial Master File を MT #2 にセットし、出力用テープを #1 にセットする。タイプライターの C/R キーを押すと、次の作業に移る。

## 2. オプション・カードの解説

入力・出力両ファイルを開いてから、オプション・カードを読み、要求される作業を判読して、必要な処理ルーチンを実行し、終われば次のオプション・カードを読む。End card を読むと、入力ファイルの残余を読み込んで、そのまま出力ファイルに書き込み、しる後に両ファイルを close する。

## 3. 追加・挿入作業

入力ファイルを読んで、入力レコード番号と、オプション・カードで指定したレコード番号とを比較し、出力ファイルに入力レコードを書き込む。入力番号と指定番号がマッチするまでこの操作は繰り返される。両者がマッチする以前に、指定番号よりも大きな番号が現われた場合には、'RECORD NO ERROR. RESET CARDS.' というエラー・メッセージがタイプされる。

マッチした場合には、データ・カードを読んでレコードを整形するサブルーチンを実行して、新しいレコードをファイルに出力する。出力のたびにカウンターから 1 を引き、カウンターが 0 になるまでこの作業を繰り返す。

## 4. 削除作業

入力ファイルを読んで、入力レコード番号と指定レコード番号とを比較し、マッチしなければ入力レコードを出力する。マッチすれば、カウンターから 1 を引き、

## 出力例 2. Sequential List (Altered)

MECHANIZED INDEXING SYSTEM BY ANZAI		SEQUENTIAL LIST (ALTERED)	11/01/69	PAGE 1
	1	SAUNDERS, W.C. THE 'SHEFFIELD REPORT' *	ASPR-21(69)009	
	2	RICHNELL, D.T. DEVELOPMENT IN EDUCATION LEADING TO FORMAL QUALIFICATION *	ASPR-21(69)013	
	3	WILSON, T.D. THE HSC DEGREE IN INFORMATION SCIENCE AT NEWCASTLE UPON TYNE *	ASPR-21(69)018	
	4	SCHUR, HERBERT UNIVERSITY OF SHEFFIELD MSC COURSE IN INFORMATION STUDIES *	ASPR-21(69)024	
	5	FARRADANE, J. STANDARDS IN EDUCATION IN INFORMATION SCIENCE *	ASPR-21(69)032	
	6	HILL, D.W. INFORMATION SERVICES AND MARKET RESEARCH *	ASPR-21(69)059	
	7	LEES, R. SOURCES OF INFORMATION ON MARKET RESEARCH USED BY INDUSTRIAL LIAISON CENTERS TO ASSIST SMALL FIRMS *	ASPR-21(69)064	
	8	HANDOVSKY, F.L.A. COLLECTION AND UTILIZATION OF ECONOMIC INFORMATION: SOME OF ITS PROBLEMS *	ASPR-21(69)071	
	9	HUTCHESON, B.N.P. INFORMATION IN MARKET RESEARCH *	ASPR-21(69)075	
	10	ADLER, L.S. CURRENT TRENDS IN GOVERNMENT STATISTICS FOR THE CHEMICAL INDUSTRY *	ASPR-21(69)079	
REPLACE の 134	11	HUMPHRY, H.G. INTERNATIONAL CHEMICAL STATISTICS: HISTORICAL BACKGROUND AND CURRENT PROBLEMS *	ASPR-21(69)098	
	12	HATTY, C.D. PROGRAMMED INSTRUCTION IN CLASSIFICATION AND INDEXING *	ASPR-21(69)127	
	13	VICKERY, B.C. RESEARCH AT ASLIB *	ASPR-21(69)156	
	14	HARTMAN, R.R.K. LINGUISTICS AND TRANSLATION *	ASPR-21(69)190	
	15	WRIGHT, J.E. STEPS IN THE DEVELOPMENT OF BS 1000 *	ASPR-21(69)195	
	16	LLOYD, G.A. THE UDC IN ITS INTERNATIONAL ASPECTS *	ASPR-21(69)204	
	17	WOODTON, BARBARA EXPERT AND LAYMAN IN CONTEMPORARY SOCIETY *	ASPR-21(69)225	
	18	EARLE, PENELOPE VICKERY, BRIAN SUBJECT RELATIONS IN SCIENCE/TECHNOLOGY LITERATURE *	ASPR-21(69)237	
	19	WOOD, D.N. DISCOVERING THE USER AND HIS INFORMATION NEEDS *	ASPR-21(69)262	

## 索引の自動化システムMISA

カウンターが0になるまで入力ファイルの読み込みのみを続ける。

マッチする以前に、指定番号より大きな入力レコード番号が現われると、追加の際と同じエラー・メッセージがタイプされる。

### 5. 取 換 作 業

入力ファイルを読んで、入力レコード番号と指定レコード番号とを比較し、マッチしなければ入力レコードを出力する。マッチすれば、サブルーチンによってデータを読み、レコードを整形して出力ファイルに書き込む。カウンターから1を引き、カウンターが0になるまで入力ファイルを get しては、データを読んで put する操作を繰り返す。

### 6. 文献リストの印刷

End mark を読みこむとファイルを close し、印刷用スイッチを調べ、その内容が1であれば、印刷ルーチンを実行し、Sequential List (Altered) を印刷する。印刷作業やフォーマットは INIMAS の場合とまったく同じである。

印刷が終了すると、‘OUTPUT IN MT 1’ というメッセージがタイプアウトされ、次のセグメントがロードされる。

印刷をしない場合にも前記のメッセージは出される。(出力例2参照)

## IV. KWICED の機能

ターム・データをカード・リーダーで読みとり、リスト処理によって木構造の辞書をコア・メモリー内に作成し、Master File (Initial または Altered) を MT # 1 から入力して keywords を選別し、レコードを整形して KWIC File を MT # 2 に出力すると同時に、指定によりラインプリンターにも出力させる機能を持っている。

### A. コントロール・カードとオプション

コントロール・カードのフォーマットは表2のとおりである。

コントロール・カードの第1フィールドは必ず KWI-

表 2. KWICED コントロール・カード

Col. 1~	13~	25~	37~
KWICED	SKIP STOP WORD KEYWORD ALL WORDS	PRINT [NO PRINT]	mm/dd/yy

CED とパンチされていなければならない。

このセグメントの実行を必要としない場合は、第2フィールドに SKIP と指定すればよい。

辞書を作成する場合に、ストップ・ワードを登録する方式とキーワードを登録する方式とがあるが、前者を選ぶ場合は STOP WORD、後者を選ぶ場合は KEYWORD と、第2フィールドで指定する。すべての語を keyword として処理する場合には、当然辞書は作成されないの、ターム・データは存在しない。

第3フィールドで PRINT と指定すれば、ソートされていない KWIC 索引が印刷される。印刷の場合は、さらに第4フィールドに日付をパンチしなければならない。第3フィールドがブランクであれば、印刷は行なわない。この場合 NO PRINT と指定してもよい。

### B. セグメントの作業

#### 1. コントロール・カードの解読

コントロール・カードを読みとり、第1フィールドに SKIP 指定があれば、直ちに次のセグメントである KWOCED をロードする。SKIP 指定がなければ ‘KWICED LOADED’ のメッセージをタイプライターに出力する。

ワード処理方式の指定を判読し、STOP WORD 指定であれば、ワード方式指定用スイッチを1にセットし、KEYWORD 指定であれば、スイッチを0に、ALL WORDS 方式であればそれを -1 にセットする。

次に印刷指定の有無をチェックし、指定があれば印刷用スイッチを1に、なければ0にセットする。

タイプライターに ‘SET MT 1 (INP), 2’ という指示を出力して pause 状態に入る。

コントロール・カードの指定に誤りがあれば、‘CONTROL CARD ERROR. RESET CARDS.’ というメッセージをタイプアウトする。正しいコントロール・カードをセットし直して、C/R キーを押せば、解読ルーチンが再び繰り返される。

#### 2. 辞書の作成

ワード方式指定用スイッチをテストし、その内容が0か1であれば、ターム・データを読んで辞書を割り付ける。-1 であれば辞書を作成せず、次のステップに移る。

リスト処理による辞書の作成は、技術的に多くの問題を内包しているので、これについては「Ⅷ. 辞書のリスト処理」で論ずることにする。

#### 3. デリミターとキーワードの処理

入力ファイルである Master File (Initial または

Altered) ならびに出力ファイルを open する。(入力 MT # 1, 出力は MT # 2)

入力ファイルを get し、ソース・コードと第 1 著者名を出力エリアに転送する。

標題を展開して、まず語の抽出を行なう。語は、初語を除き、デリミターに囲まれているので、デリミターの次の文字を語頭とし、次に現われるデリミターの直前の文字を語尾とすることによって抽出する。ただし、デリミターは複数個連続することがあるから、デリミターを発見しても、直ちにその次の文字を語頭と断定してはならない。

デリミターには、ブランク、コンマ、ピリオド、コロ、セミコロン、左カッコ、右カッコ、クォーテーション・マーク等があるが、デリミター判定の最も単純な方法は、文字リテラルによる減算を繰り返し、減算の結果が 0 になるかどうかで判定する方式であるが、非能率的なので、64 ワードからなる判別用ベクトルを用意し、ダイレクトにアクセスして判定できるようにする方がよい。

計算機で使用する記号のうち、いずれをデリミターとし、いずれをノン・デリミターとして扱うかは、かなり難しい問題である。MISA においては、アルファベット以外のすべての文字をデリミターとして処理する。(ただし標題末尾記号の\*および、ファイル終了記号の@は別扱いである。) この場合問題となるのはアポストロフィとハイフンであろう。アポストロフィは通常計算機ではクォーテーション・マークを兼ねているので、アポストロフィをノン・デリミターとすれば、クォーテーション・マークまでもノン・デリミターとなって不都合を生ずる。しかしながら、一方クォーテーション・マークをデリミターとすれば、IT'S のような語が IT と S の 2 語に分割されてしまう。アポストロフィの後には通常 S や T や D なので、MISA では、これら単独の文字をストップ・ワード辞書に登録し、ストップ・ワードとして処理

している。'S や 'T は容易に多くの語に付着するので、ストップ・ワード辞書に一々それらの語形をも登録するのは煩わしく、辞書のサイズを不必要に大きくする。

ハイフンは合成語を形成する機能を持つ記号であるが、ハイフンによって連結される 2 要素のいずれからアクセスできるようにしておきたい場合がしばしばある。このような場合には、ハイフンがデリミターであることが望ましいが、一方において、ON-LINE のような語の場合には、ON がストップ・ワードとして登録されるのが普通なので、不都合な事態を生ずる。筆者が実際に入力したデータ中にも ON-LINE は幾度か出現しているので、やむをえず、ONLINE とパンチする特別の措置を講じなければならなかった。

さらに MISA では、数字をデリミターとして扱う。標題中にはしばしば数字が現われるが、これらの数字語をストップ・ワードとして処理することは難しい。数字の組み合わせは無限にあり、ストップ・ワード辞書に登録することは事実上不可能であるからである。そこで考案されたのが、0 から 9 までの数字をデリミターとして処理し、ストップ・ワードとして除去するのと同じ結果をもたらす方策である。

いずれにせよ、デリミターの判別によって語が抽出されると、辞書を引いて、その語が登録されているかどうかを調べなければならない。ALL WORDS の指定があればその必要はない。STOP WORD 方式の場合には、登録語とマッチすれば捨て、逆に KEYWORD 方式では、登録語とマッチしなければ捨てることになる。

#### 4. レコードの整形

KWIC 索引は特殊な形式を持つので、レコードの整形は最も煩雑な作業となる。図 5 は出力形式の骨組を示したものである。

1 行の印刷字数は 120 で、標題 (I) には 38 字、標題 (II) には 42 字が割り当てられている。キーワードの先頭は 41 字目に揃えられ、その直前には 2 字分の間隙が

図 5. KWIC の出力形式

1	38		41	82		85 98		101 120
標 題 (I)	デ リ ミ タ ー	ブ ラ ン ク	キ ー ワ ー ド の 初 字	標 題 (II)	ブ ラ ン ク	ソ ー ス ・ コ ー ド	ブ ラ ン ク	第 一 著 者 名

## 索引の自動化システムMISA

置かれる。そのうちの1字分(40字目)はblankであるが、他の1字(39字目)は、キーワードが標題の中でその直前に持っていたデリミターである。通常はblankであるが、キーワードがカッコやクォーテーション・マークで囲まれている場合は、左カッコやクォーテーション・マークがこの位置に現われる。

標題(Ⅰ)は右詰め、標題(Ⅱ)は左詰めの形をとる。(Ⅰ)に入るべき部分が38字以内であり、(Ⅱ)に入るべき部分が42字以内である場合には問題ない。しかしながら、超過した場合には種々面倒な問題が生ずる。その場合、このセグメントは次のような処理を行なう。

- ① 両者ともに超過する場合は、溢れた部分は切り捨てる。
- ② (Ⅰ)が超過して(Ⅱ)に余白がある場合は、(Ⅰ)の先頭につながる文字を、余白部分に右詰めで入れる。それでも入りきれない場合には、標題末尾記号(\*)の後にblank1字とプラス記号を置いてカットする。
- ③ (Ⅱ)が超過して(Ⅰ)に余白がある場合には、(Ⅱ)の後尾につながる文字を余白部分に左詰めで入れる。ただし、それでも収容しきれない場合には、(Ⅰ)の先頭語との間にスラッシュ記号と1個のblankを置いてカットする。

以上の整形作業の主体は、超過・余白の字数計算と比較および転送であるが、転送も通常の転送では間に合わず、逆転送のサブ・ルーチンを必要とする。

整形されたレコードはテープ・ファイルに出力される。

### 5. 中間印刷

印刷指定がある場合には、未分類KWIC索引をラインプリンターに出力するが、出力テープ・ファイルの作成が終ってから、それを読みとって印刷するのではなく、テープに書き込みながら同時に印刷する方式をとっている。

この印刷はあくまでもテスト的なものであるが、ストップ・ワードやキーワードの選定の適否をチェックするには、ソートずみの索引よりも反って便利である。

## V. KWOCED の機能

KWOCEDは、ターム・データをカード・リーダーから読とり、リスト処理によって木構造の辞書をコア・メモリー内に作成し、Master File (InitialまたはAltered)をMT #1から入力してkeywordsを選別し、レコードを整形して未分類のKWOC型ファイルをMT #2に出力すると同時に、指定により、ラインプリンターにも出力させる機能を持っている。

KWOCには広狭の二義がある。KWOCはKey Word Out of Contextの略称であるから、subject termのみならず、著者名等の固有名詞もkeywordと称するとすれば、主題索引であれ、著者索引であれ、keywordを分離した形式で標目とする索引はすべてKWOC索引と呼んでも差し支えないようにも思われる。これら広狭二義のKWOCを区別するため、本稿では狭義のものをKWOC索引、広義のものをKWOC型索引と呼ぶことにする。

従って、KWOCEDは、KWOC型索引を作成する機能を持つ、と言い改めた方が正確である。

### A. コントロール・カードとオプション

コントロール・カードのフォーマットは表3のとおりである。

第1フィールドには必ずKWOCEDと指定しなければならない。このセグメントを実行しない場合には、第2フィールドでSKIPを指定する。

ストップ・ワード方式の場合には、第2フィールドでSTOP WORDと指定し、キーワード方式の場合は、KEYWORDと指定する。すべての語をキーワードとしてとる場合にはALL WORDSと指定し、すべての語をキーワードとして採らない(例えば、著者索引のみを作

表 3. KWOCED コントロール・カード

Col. 1～	13～	25～	37～	49～
KWOCED	SKIP STOP WORD KEYWORD ALL WORDS NO WORD	NO AUTHOR [AUTHOR]	PRINT [NO PRINT]	mm/dd/yy

成する) 場合には, NO WORD と指定する。

KWOC 索引のみを作成し, WADEX や著者索引を必要としない場合には, 第3フィールドに NO AUTHOR と指定すればよい。WADEX や著者索引を作成するために著者名を採る必要がある場合には, 第3フィールドはブランクのままでもよいし, AUTHOR としてもよい。

結果を印刷したいときは, 第4フィールドに PRINT と指定する。印刷しない場合は, このフィールドはブランクにしておくか, あるいは NO PRINT と指定する。印刷の場合には, さらに第5フィールドに日付をパンチする。

## B. セグメントの作業

### 1. コントロール・カードの解読

コントロール・カードを読み, 第1フィールドに SKIP 指定があれば, 直ちに次のセグメントである SORPRI をロードする。SKIP 指定がなければ, 'KWOCED LOADED' のメッセージをタイプライターに出力する。

ワード処理方式の指定を解読し, STOP WORD 指定であれば, ワード方式指定用スイッチを1にセットし, KEYWORD 指定であれば, スイッチを0に, ALL WORDS 方式であれば, スイッチを-1にセットする。NO WORD 指定であれば, 2にセットする。

さらに第4フィールドで印刷指定の有無をチェックし, 指定があれば印刷用スイッチを1に, なければ0にセットする。

タイプライターに 'SET MT 1 (INP), 2' という指示を出して pause 状態に入る。

コントロール・カードの指定に誤りがあれば, 'CONTROL CARD ERROR. RESET CARDS.' というメッセージをタイプアウトする。正しいコントロール・カードをセットし直して C/R キーを押すと, 解読ルーチンが再度実行される。

### 2. 辞書の作成

ワード方式判別用スイッチをテストし, その内容が0か1であれば, 辞書割付サブルーチンを実行する。-1か2であれば, 実行せず, 次のステップに移る。

### 3. Author Entry の作成

入力ファイルである Master File (Initial または Altered) ならびに出力ファイルを open する。(入力は MT #1, 出力は MT #2)。

入力ファイルを get し, 著者名, ソース・コードならびに標題を出力エリアに転送する。

著者記入用スイッチをチェックし, その内容が1であれば, 入力エリアの著者名を出力エリアの先頭 (heading

図 6. KWOC 出力レコード・フォーマット

0	4	5	6	10	11	12	16	17	18	22	23	24	27	28	
標 目 (5 語)		ブ ラ ン ク	第 1 著者名 (5 語)		ブ ラ ン ク	第 2 著者名 (5 語)		ブ ラ ン ク	第 3 著者名 (5 語)		ブ ラ ン ク	ソース・コード (4 語)		標 題 字 数 指 数	
29	33	58													
ブ ラ ン ク (5 語)		標 題 前 半 (100 字)													
59	63	88													89
ブ ラ ン ク (5 語)		標 題 後 半 (100 字)													標 目 種 別

(注) 上欄の数字は番地の修飾値

用エリア)に転送し、出力ファイルに put する。著者が 2 名または 3 名の場合にはこの作業が繰り返され、2 個または 3 個の author entry が作成される。Author entry は次のステップで作られる word entry と区別するため、出力レコードの末尾に 'A' マークを記入する。スイッチの内容が 0 であれば、この作業はスキップする。

#### 4. Word Entry の作成

Word の抽出に必要なデリミターとキーワードの処理については、KWICED とまったく同じ方法を用いているので、省略する。

ワード方式指定用スイッチの内容が -1 であれば、すべての word を出力エリアの先頭 (heading 用エリア) に転送するが、スイッチの内容が 0 であれば辞書の登録語とマッチした word のみを転送し、スイッチの内容が 1 であれば、辞書の登録語とマッチしない word のみを転送する。転送が終れば出力ファイルに put する。Author entry と区別するため、出力レコードの末尾に 'w' マークを記入する。

スイッチの内容が 2 であれば、この作業はスキップする。

KWOC 型索引は KWIC 索引に比較して、レコードの整形作業が簡単である。

図 6 は出力レコードのフォーマットを示したものである。

#### 5. 中間印刷

印刷指定がある場合には、未分類の KWOC 型索引をラインプリンターに出力するが、出力テープ・ファイルの作成が終了してから、ファイルを読んで印刷するのではなく、テープへの書き込みと平行して、印刷スイッチをチェックしながら印刷する方式を採用している。

KWICED の場合と同じく、この印刷はあくまでも中間のテスト的な性格を持つものである。

ファイルの作成が終ると、'OUTPUT IN MT 2' と

いうメッセージがタイプアウトされ、次のセグメントである SORPRI がロードされる。

### VI. SORPRI の機能

未分類の KWIC ファイルまたは KWOC ファイルを MT # 2 から入力し、MT # 1, MT # 4 をワーク・テープに使用しながら、ファイルの全レコードを標目をキーとして ABC 順にソートして、MT # 2 に KWIC File または KWOC File を出力すると同時に、KWIC 索引または KWOC 型索引をラインプリンターで印刷する。KWOC 型索引の場合には、KWOC Index, Author Index, WADEX のいずれか 1 つを印刷することも、2 種あるいは 3 種を連続して印刷することも可能である。

#### A. コントロール・カードとオプション

コントロール・カードの format は表 4 のとおりである。

コントロール・カードの第 1 フィールドには必ず SORPRI の指定がなければならない。

このセグメントを実行しない場合には、第 2 フィールドで SKIP を指定する。

ソートずみの索引を印刷する場合には、第 2 フィールドに PRINT と指定し、印刷しない場合には NO PRINT の指定を行なう。

印刷する場合には第 3 フィールドに日付をパンチする。

第 4 フィールドにはソートされる索引の種類を指定する。印刷しない場合にも、この指定は欠かせない。印刷する場合は、第 4 フィールドの指定が同時に印刷する索引の種類を兼ねる。

異種の索引を印刷する場合、例えば KWOC Index と Author Index と WADEX を続けて印刷する場合には、第 4 フィールドで KWOC、第 5 フィールドで AUTHOR、第 6 フィールド WADEX を指定すればよい。

表 4. SORPRI コントロール・カード

Col. 1~	13~	25~	37	49~	61~
SORPRI	SKIP PRINT NO PRINT	mm/dd/yy	KWIC KWOC AUTHOR WADEX	AUTHOR WADEX KWOC	WADEX KWOC AUTHOR

同種のを2部または3部印刷する必要があるれば、例えば KWIC を2部印刷する場合には、第4フィールドと第5フィールドとともに KWIC を指定すればよい。

KWIC 索引と KWOC 型索引は入力ファイルが別なので、その両者を混合して指定することはできない。

## B. セグメントの作業

### 1. コントロール・カードの解読

コントロール・カードを読み、第1フィールドのセグメント名をチェックする。第2フィールドに SKIP 指定があれば、直ちに次のセグメント WORDFR をロードする。SKIP 指定がなければ、タイプライターに 'SORPRI LOADED' のメッセージを出力する。

第2フィールドに印刷指定があれば、印刷用スイッチを1にセットし、NO PRINT の指定があれば、スイッ

チを0にセットする。

第4フィールドの索引種別指定の解読にはサブ・ルーチンを使用し、KWIC であればスイッチを-1に、KWOC であれば0に、WADEX ならば1に、AUTHOR であれば2にセットする。

コントロール・カードの指定に誤りがあれば、'CONTROL CARD ERROR. RESET CARDS.' のメッセージをタイプアウトする。正しいコントロール・カードをセットし直し、C/R キーを押せば、解読ルーチンが再び実行される。

### 2. 入力ファイルの分割

入力ファイルとワーク・ファイル2本(MT #1, MT #4)を開く。

入力ファイルを get し、奇数番のレコードはワーク・

## 出力例 3. KWIC Index

MECHANIZED INDEXING SYSTEM BY ANZAI	K W I C I N D E X	12/09/69	PAGE 1
ORY, TESTING, AND MECHANIZATION OF THE LIBRARY AND INFORMATION SCIENCE MULA DERIVED FROM A FACTOR ANALYSIS OF THE EXCHANGE OF	ABC RETRIEVAL SYSTEM * ABSTRACTS: A QUARTERLY PROGRESS REPORT * ACADEMIC DISSEMINATIONS * ALL ALLOCATION FOR ACADEMIC INFORMATION CENTERS * ACADEMIC LIBRARIANS' SALARIES * ACADEMIC LIBRARIANSHIP * ACADEMIC LIBRARIES * USER CIRCULATION SATI ACADEMIC LIBRARIES * ACADEMIC LIBRARIES * ACADEMIC LIBRARIES * ACADEMIC LIBRARY IN URBAN DEVELOPMENT * ACADEMIC LIBRARY WORLD: NOT SO ROUND * ACADEMIC MEDICAL LIBRARY STATISTICS * ACCESS COMPUTER * * FOR RETRIEVAL BY CONFL ACCESS TO INFORMATION: NSF AS AN AID TO TH ACCOUNTING FOR A TECHNICAL INFORMATION CEN ADMINISTRATIVE KNOW-HOW IN LIBRARIES * ADVENT OF HUMANIZATION IN COLLEGE LIBRARY ADVERTISING AND MARKETING RESEARCH * AFRICA * * GANTIZATION OF NATIONAL DOCUMENTA ALABAMA AND MISSISSIPPI: A REPORT * *ND TE ALERTING SYSTEM * ALGORITHM * ALLOCATION FORMULA DERIVED FROM A FACTOR A AMERICA, WITH SPECIAL REFERENCE TO BRAZIL AMERICAN CATALOGING RULES, 1967 * AMERICAN DOCUMENTATION INSTITUTE: A SKETCH AMERICAN HOSPITAL ASSOCIATION * ANALYTICAL SURVEY OF VIROLOGY LITERATURE R ANGLO-AMERICAN CATALOGING RULES, 1967 * ARBITRARY LOGICAL STATEMENTS IN MECHANIZED ARCHIBALD MACLEISH, LIBRARY OF CONGRESS * ARCHITECTURE AND THE COLLEGE LIBRARY * ASIAN COLLEGE STUDENTS * ASLIB * ASPECTS * ASSIST SMALL FIRMS * *N ON MARKET RESEARCH AUTOMATED LIBRARY SYSTEM PROPOSAL * AUTOMATED LITERATURE ALERTING SYSTEM * AUTOMATIC CLASSIFICATION TECHNIQUE FOR DOC AUTOMATIC INDEXING METHODS * AUTOMATIC PROCESSING OF SERIALS RECORD: SR AUTOMATICALLY GENERATED WEIGHTS AND LINKS AUTOMATION * AUTOMATION ON TEN DOLLARS A DAY * BANKS: AN EVALUATION OF ONE LIBRARY'S TRANS BANKING * BENCHMARK CALCULATION FOR RETRIEVAL BY CON BENEFIT ANALYSIS IN LIBRARIES * BIBLIOGRAPHIC DIMENSIONS IN INFORMATION CO BINARY ATTRIBUTES IN A RANDOM ACCESS COMPU BINARY SEARCH ALGORITHM * BIO-MEDICAL LIBRARY SERIALS SYSTEM * BIOLOGIST * *RE: ANALYSIS OF JOURNAL ARTIC	AMDO-20(69)006 ASPR-21(69)1325 CORL-30(69)051 CORL-30(69)237 AMDO-20(69)039 CORL-30(69)101 CORL-30(69)214 CORL-30(69)204 CORL-30(69)265 CORL-30(69)344 LITR-18(69)066 CORL-30(69)126 CORL-30(69)362 BMLA-57(69)233 AMDO-20(69)168 AMDO-20(69)213 LITR-17(69)111 BMLA-57(69)160 LITR-18(69)029 SPLI-60(69)219 LITR-17(69)267 CORL-30(69)369 AMDO-20(69)016 AMDO-20(69)167 CORL-30(69)051 LITR-17(69)245 LIIS-07(69)059 AMDO-20(69)152 BMLA-57(69)177 AMDO-20(69)234 LIIS-07(69)059 AMDO-20(69)253 CORL-30(69)012 LITR-18(69)037 CORL-30(69)267 ASPR-21(69)156 ASPR-21(69)204 ASPR-21(69)064 SPLI-60(69)288 AMDO-20(69)016 LIIS-07(69)117 AMDO-20(69)061 LIIS-07(69)131 AMDO-20(69)198 CORL-30(69)307 SPLI-60(69)215 BMLA-57(69)064 SPLI-60(69)003 SPLI-60(69)017 AMDO-20(69)168 CORL-30(69)156 AMDO-20(69)247 AMDO-20(69)168 AMDO-20(69)167 SPLI-60(69)349 AMDO-20(69)143	ALTMANN, BERTHOLD GILCHRIST, ALAN MCGRATH, W.E. ASH, JOAN LUNIN, L.F. SCHILLER, A.R. SHORES, LOUIS TRUESWELL, R.W. CAIN, S.M. HUGHES, L.W. JOSEY, E.J. KLOTSCH, J.M. RITLER, PAUL PING, V.M. MANACHER, G.K. RENNETT, J.L. HELMKAMP, J.G. SIMON, B.V. KILGOUR, F.G. CHRISTIANSON, E.B. KINGWILL, D.C. PHINAZEE, A.B. BRANNON, P.B. DAVIS, C.H. MCGRATH, W.E. VICENTINI, A.L.C. HORIUCHI, TKUKO SCHULTZ, C.K. YAST, ELLEN CARROLL, K.H. HORIUCHI, TKUKO UHLMANN, WOLFRAM GOLDSCHMIDT, EVA THOMPSON, D.E. LEWIS, M.G. VICKERY, B.C. LLOYD, G.A. LEES, R. COOK, MEYER BRANNON, P.B. HOSONO, KIMIO SALTON, GEPARD OGAWA, HARIKI ARTANDT, SUSAN KILGOUR, F.G. ATWOOD, RUTH LOVE, ERTKA MARON, M.E. DALE, H.L. MANACHER, G.K. KELLER, J.F. LUBETZKY, SEYMOUR MANACHER, G.K. DAVIS, C.H. GROSCH, A.N. LEITH, J.D.

## 索引の自動化システムMISA

ファイル #1 に put し、偶数番のレコードはワーク・ファイル #4 に put する。この振り分けにはスイッチを使用する。スイッチは 0 に initialize し、スイッチが 0 であればワーク・ファイル #1 に put して、スイッチを 1 にセットする。スイッチが 1 であれば、ワーク・ファイル #4 に put してスイッチ 0 をにセットする。

Put するたびに、レコード・トータル・カウンターに 1 を加える。

入力ファイルが EOF 状態になれば、入力ファイル、ワーク・ファイルとともに close する。

### 3. 併 合

ソーティングには種々の方法があるが、このセグメントでは merge-sort と呼ばれる併合分類方式を用いている。

MISA は使用できる磁気テープ装置を 4 台と想定しているが、内 2 台はシステム・テープと MISA のオブジェクト・テープに使用される。仮にシステム・テープを取り外してワーク・テープと取り替えても、ソーティングに使用できるテープ装置は 3 台に限定される。仮に 4 台使用できれば、2-way merge-sort が可能であるが、3 台に限定されるため、SORPRI では 1-way merge-sort を行なう。2-way に較べて能率は落ちるがやむをえない。

ワーク・ファイル #1 と #4 の各々からレコードを get して、その大小を比較し、正順 (小→大) 順に並んだレコードのストリングをワーク・ファイル #2 に作るのであるが、各ファイルから get するレコードの数は、1, 2, 4, 8, 16, … と併合の段階ごとに 2 進的に増えていく。従って併合によって作られる正順ストリング内のレコー

## 出力例 4. KWOC Index

MECHANIZED INDEXING SYSTEM BY AN7AI	K W O C I N D E X	11/29/69	PAGE 1
ABC	ALTMANN, BERTHOLD      RIESSLER, W.A. THEORY, TESTING, AND MECHANIZATION OF THE ABC RETRIEVAL SYSTEM *	AMDO-20(69)006	
ABSTRACTS	GILCHRIST, ALAN LIBRARY AND INFORMATION SCIENCE ABSTRACTS: A BRIEF PROGRESS REPORT *	ASPR-21(69)325	
ACADEMIC	ASH, JOAN THE EXCHANGE OF ACADEMIC DISSERTATIONS *	CORL-30(69)237	
ACADEMIC	BIXLER, PAUL THE ACADEMIC LIBRARY WORLD: NOT SO ROUND *	CORL-30(69)362	
ACADEMIC	CAIN, S.M. SERVICE HOURS IN SELECTED ACADEMIC LIBRARIES *	CORL-30(69)265	
ACADEMIC	HUGHES, L.W. SOUTH OF EXPLOSIVE EXPONENTIALISM IN ACADEMIC LIBRARIES *	CORL-30(69)344	
ACADEMIC	JOSEY, E.J. COMMUNITY USE OF ACADEMIC LIBRARIES *	LITR-18(69)066	
ACADEMIC	KLOTSCHÉ, J.M. THE ROLE OF THE ACADEMIC LIBRARY IN URBAN DEVELOPMENT *	CORL-30(69)126	
ACADEMIC	LUNIN, L.F. ACADEMIC INFORMATION CENTERS *	AMDO-20(69)039	
ACADEMIC	MCGRATH, W.E.      HUNTSINGER, R.C.      BARBER, G.R. ALL ALLOCATION FORMULA DERIVED FROM A FACTOR ANALYSIS OF ACADEMIC DEPARTMENTS *	CORL-30(69)051	
ACADEMIC	PINGS, V.M.      OLSON, E.E.      ORR, R.H. SUMMARY REPORT OF A STUDY OF ACADEMIC MEDICAL LIBRARY STATISTICS *	BMLA-57(69)233	
ACADEMIC	SCHILLER, A.R. ACADEMIC LIBRARIANS' SALARIES *	CORL-30(69)101	
ACADEMIC	SHORES, LOUIS THE JUNIOR COLLEGE IMPACT ON ACADEMIC LIBRARIANSHIP *	CORL-30(69)214	
ACADEMIC	TRUESWELL, H.W. USER CIRCULATION SATISFACTION VS. SIZE OF HOLDINGS AT THREE ACADEMIC LIBRARIES *	CORL-30(69)204	
ACCESS	RENNETT, J.L. ONLINE ACCESS TO INFORMATION: NSF AS AN AID TO THE INDEXER/CATALOGER *	AMDO-20(69)213	
ACCESS	MANACHER, G.K. A BENCHMARK CALCULATION FOR RETRIEVAL BY CONFLUENCE OF BINARY ATTRIBUTES IN A RANDOM ACCESS COMPUTER *	AMDO-20(69)168	
ACCOUNTING	HELMKAMP, J.G. MANAGERIAL COST ACCOUNTING FOR A TECHNICAL INFORMATION CENTER *	AMDO-20(69)111	
ADMINISTRATIVE	SIMON, B.V. THE NEED FOR ADMINISTRATIVE KNOW-HOW IN LIBRARIES *	BMLA-57(69)160	
ADVENT	KILGOUR, F.G.	LITR-18(69)029	



ド数は、その 2 倍、すなわち 2, 4, 8, 16, 32, …と回を重ねるごとに倍増していく。そして、ストリングの大きさが、レコード総数に達すると、ソーティングは終了する。

比較のために get すべきレコード数に制限があるため、比較が終って put してから、次にいずれのファイルからレコードを get すべきか、あるいは get せずに作業エリアに残留しているレコードを put すべきかの判断はかなり複雑なアルゴリズムとなり、誤りを犯しやすい。併合を正しく遂行するには、いくつかのスイッチをうまく使わなければならない。

#### 4. 大小比較

大小比較は併合の過程で行なわれる。KWICとKWOC型とではソートのキーとなるキーワードの位置が異なる

で、スイッチによって分岐し、番地修飾用インデックス・レジスターの初期値を別個にセットしてから、同一の比較ルーチンを使えばよい。KWICの場合はキーワードの先頭から 40 字までを比較し、KWOC 型の場合は標目を比較し、同じであれば第 1 著者名について比較を行なう。

比較は文字単位で行なうが、能率をよくするため、最初の 4 文字についてはワード単位で比較する。

ワードの初字が A (8 進コード 21) から I (8 進コード 31) までの場合は、ワードの先頭ビットに 0 が立つので、プラスの数値となり、J (8 進コード 41) 以降の場合は先頭ビットに 1 が立つのでマイナスの数値となる。この原理を応用して、ワードがプラスかマイナスかで大小を比較する。プラス同志、マイナス同志の場合のみ、文字単位の大小比較を行なう。<sup>5)</sup> この場合、ストリング

#### 出力例 5. Author Index

MECHANIZED INDEXING SYSTEM BY ANZAI	AUTHOR INDEX	11/29/69	PAGE 1
ADLER, I.S.	ADLER, I.S. CURRENT TRENDS IN GOVERNMENT STATISTICS FOR THE CHEMICAL INDUSTRY *	ASPR-21(69)079	
ALTMANN, BERTHOLD	ALTMANN, BERTHOLD      RIESSLER, W.A. THEORY, TESTING, AND MECHANIZATION OF THE ABC RETRIEVAL SYSTEM *	AMDO-20(69)006	
ANZAI, IKUO	ANZAI, IKUO COMPUTER TECHNIQUES FOR PROCESSING STRINGS OF CHARACTERS *	LIIS-07(69)109	
ARTANDI, SUSAN	ARTANDI, SUSAN      WOLF, E.H. THE EFFECTIVENESS OF AUTOMATICALLY GENERATED WEIGHTS AND LINKS IN MECHANICAL INDEXING *	AMDO-20(69)198	
ASH, JOAN	ASH, JOAN THE EXCHANGE OF ACADEMIC DISSERTATIONS *	CORL-30(69)237	
ATWOOD, RUTH	ATWOOD, RUTH      LIVINGSTON, FRANCES AUTOMATION ON TEN DOLLARS A DAY *	SPLI-60(69)215	
AUSTIN, C.J.	AUSTIN, C.J. TIME-SHARED COMPUTING: IMPLICATIONS FOR MEDICAL LIBRARIES *	BMLA-57(69)116	
BACH, HARRY	BACH, HARRY THE SNOWS OF YESTERDAY *	CORL-30(69)301	
BAKER, N.R.	BAKER, N.R. OPTIMAL USER SEARCH SEQUENCES AND IMPLICATIONS FOR INFORMATION SYSTEMS OPERATION *	AMDO-20(69)203	
BARBER, G.R.	MCGRATH, W.E.      HUNTSINGER, R.C.      BARBER, G.R. ALL ALLOCATION FORMULA DERIVED FROM A FACTOR ANALYSIS OF ACADEMIC DEPARTMENTS *	CORL-30(69)051	
BARLUP, JANET	BARLUP, JANET MECHANIZATION OF LIBRARY PROCEDURES IN THE MEDIUM-SIZED MEDICAL LIBRARY: 7. RELEVANCY OF CITED ARTICLES IN CITATION INDEXING *	BMLA-57(69)260	
BATTY, C.D.	BATTY, C.D. PROGRAMMED INSTRUCTION IN CLASSIFICATION AND INDEXING *	ASPR-21(69)127	
BAUMANIS, G.J.	GIFFORD, CAMOLYN      BAUMANIS, G.J. ON UNDERSTANDING USER CHOICES: TEXTUAL CORRELATES OF RELEVANCE JUDGMENTS *	AMDO-20(69)021	
BECKER, JOSEPH	BECKER, JOSEPH INFORMATION NETWORK PROSPECTS IN THE UNITED STATES *	LITR-17(69)306	
BENNETT, J.L.	BENNETT, J.L. ONLINE ACCESS TO INFORMATION: NSF AS AN AID TO THE INDEXER/CATALOGER *	AMDO-20(69)213	
BIRD, WARREN	BIRD, WARREN TWX AND INTERLIBRARY LOANS *	BMLA-57(69)125	
BISHOP, DAVID	BISHOP, DAVID VETERINARY MEDICINE AND THE MEDICAL SCHOOL LIBRARY *	BMLA-57(69)275	
BIXLER, PAUL	BIXLER, PAUL THE ACADEMIC LIBRARY WORLD: NOT SO ROUND *	CORL-30(69)362	
BLODGETT, J.H.	BLODGETT, J.H.      SCHULTZ, C.K.	AMDO-20(69)221	

## 索引の自動化システム M I S A

を1字/ワードの形に拡散せず、必要な文字だけを、インデックスで指定し、サブルーチンでとり出すようにする。

### 5. 併合ファイルの分割

併合の結果はワーク・ファイル#2に書き込まれるが、このワーク・ファイルは、次の併合に備えて、ワーク・ファイル#1と#4に分割される。

分割の原理は「2. 入力ファイルの分割」のそれと同じであるが、分割単位のレコード数は、2, 4, 8, 16, …と2進的に累増する。

### 6. 索引の印刷

印刷行数は、見出しが1行、空白行が2行、detailは56行である。Detailの1行は120字である。

索引種別指定用スイッチの内容を調べて、指定された

索引に必要な見出し用ラベルを見出し印刷エリアに転送し、ページ・カウンタをクリアする。

KWICは1レコードが1行に印刷され、各レコードの間には空白行を置かない。

KWOC型索引は、1レコードが標題の長さによって2行で印刷する場合と、3行になる場合とがある。レコードの間には1行の空白を置く。

レコードの整形はKWICが複雑で、KWOCは簡単であったが、印刷の場合は逆にKWICが簡単で、KWOC型が複雑である。印刷行数の可変に加えて、ワード・エントリーと著者エントリーとを識別し、指定された索引の種類によって取捨する必要があるからである。

レコードの先頭が'@@@@'であれば印刷を終り、ファイルをcloseし、コントロール・カードの次のフォル

## 出力例 6. W A D E X

MECHANIZED INDEXING SYSTEM BY ANZAI		W A D E X	11/29/69	PAGE 2
	THE NEED FOR ADMINISTRATIVE KNOW-HOW IN LIBRARIES *			
ADVENT	KILGOUR, F.G.			LITR-18(69)029
	COMPUTERIZATION: THE ADVENT OF HUMANIZATION IN COLLEGE LIBRARY *			
ADVERTISING	CHRISTIANSON, E.B.			SPLI-60(69)219
	INFORMATION RETRIEVAL FOR ADVERTISING AND MARKETING RESEARCH *			
AFRICA	KINGWILL, D.C.			LITR-17(69)267
	THE ORGANIZATION OF NATIONAL DOCUMENTATION AND INFORMATION SERVICES IN SOUTH AFRICA *			
ALABAMA	PHINAZEE, A.H.	JORDAN, C.L.		CORL-30(69)369
	CENTRALIZED LIBRARY PURCHASING AND TECHNICAL PROCESSING FOR SIX COLLEGES IN ALABAMA AND MISSISSIPPI; A REPORT *			
ALERTING	BRANNON, P.B.	HURNHAM, D.F.	JAMES, R.M.	AMDO-20(69)016
	AUTOMATED LITERATURE ALERTING SYSTEM *			
ALGORITHM	DAVIS, C.H.			AMDO-20(69)167
	THE BINARY SEARCH ALGORITHM *			
ALLOCATION	MCGRATH, W.E.	HUNTSINGER, R.C.	BARBER, G.R.	CORL-30(69)051
	ALL ALLOCATION FORMULA DERIVED FROM A FACTOR ANALYSIS OF ACADEMIC DEPARTMENTS *			
ALTMANN, BERTHOLD	ALTMANN, BERTHOLD	RIESSLER, W.A.		AMDO-20(69)006
	THEORY, TESTING, AND MECHANIZATION OF THE ABC RETRIEVAL SYSTEM *			
AMERICA	VICENTINI, A.L.C.			LITR-17(69)245
	THE ORGANIZATION OF NATIONAL DOCUMENTATION AND INFORMATION SERVICES IN LATIN AMERICA, WITH SPECIAL REFERENCE TO BRAZIL *			
AMERICAN	HORIUCHI, IKIKO			LIIS-07(69)059
	THE ANGLO-AMERICAN CATALOGING RULES, 1967 *			
AMERICAN	SCHULTZ, C.K.	GARWIG, P.L.		AMDO-20(69)152
	HISTORY OF THE AMERICAN DOCUMENTATION INSTITUTE: A SKETCH *			
AMERICAN	YAST, HELLEN			BMLA-57(69)177
	THE PROGRAM OF THE AMERICAN HOSPITAL ASSOCIATION *			
ANALYTICAL	CARROLL, K.M.			AMDO-20(69)234
	AN ANALYTICAL SURVEY OF VIROLOGY LITERATURE REPORTED IN TWO ANNOUNCEMENT JOURNALS *			
ANGLO	HORIUCHI, IKIKO			LIIS-07(69)059
	THE ANGLO-AMERICAN CATALOGING RULES, 1967 *			
ANZAI, IKUO	ANZAI, IKUO			LIIS-07(69)109
	COMPUTER TECHNIQUES FOR PROCESSING STRINGS OF CHARACTERS *			
ARBITRARY	UHLMANN, WOLFRAM			AMDO-20(69)253
	A GENERAL METHOD FOR MATCHING ARBITRARY LOGICAL STATEMENTS IN MECHANIZED RETRIEVAL SYSTEMS *			
ARCHIBALD	GOLDSCHMIDT, EVA			CORL-30(69)012
	ARCHIBALD MACLEISH, LIBRARY OF CONGRESS *			
ARCHITECTURE	THOMPSON, D.E.			LITR-18(69)037

ドの指定を調べる。もしもブランクであれば、‘OUTPUT IN MT 2’ というメッセージをタイプアウトし、次のセグメントである WORDFR をロードする。ブランクでなければ、再度ファイルを open し、スイッチで分岐しながら指定された索引を印刷する。この場合、見出し用ラベルが変更され、ページ番号も更新されることは言うまでもない。

コントロール・カードのスペースの関係で、印刷を指定できる索引の種別は3種（同種のものを含めて）である。カードのフォーマットを変更すれば、それ以上の指定もできるが、実際にその必要があるとは思われない。

(出力例 3, 4, 5, 6 参照)

## VII. WORDFR の機能

ソートずみの KWOC ファイルまたは WADEX ファイル (MT # 2) を入力ファイルとし、いわばその副産物としてキーワードの生起頻度を計算し、キーワードの ABC 順に頻度表を印刷する機能を持っている。

KWIC File や Author File は入力ファイルとして使うことができない。Author File を入力した場合には、著者名の生起頻度を計算し、それを著者名の ABC 順に印刷することになる。

原始データの標題中で使用されているすべての語の生起頻度を特に印刷したい場合には、KWOCED のコント

### 出力例 7. Word Frequency

MECHANIZED INDEXING SYSTEM BY ANZAI	WORD FREQUENCY	PAGE 1
A	32	BIG 1
ABC	1	BINARY 2
ABOUT	1	BIO 1
ABSTRACTS	1	BIOLOGIST 1
ACADEMIC	12	BIOMEDICAL 1
ACCESS	2	BLOCKS 1
ACCOUNTING	1	BOOK 2
ADMINISTRATIVE	1	BOOKS 3
ADVANCES	1	BOON 1
ADVENT	1	BORROWED 1
ADVERTISING	1	BRAZIL 1
AFFECTING	1	BREAKING 1
AFRICA	1	BRIEF 1
AFTER	2	BRITISH 1
AID	1	BS 1
ALABAMA	1	BSC 1
ALERTING	1	BUDGETING 1
ALGORITHM	1	BUILDING 2
ALL	1	BUREAU 1
ALLOCATION	1	BUSINESS 1
AMERICA	1	BY 5
AMERICAN	3	BYSTANDER 1
AMONG	1	CALCULATION 1
AN	11	CANADIAN 1
ANALYSIS	7	CARD 2
ANALYTICAL	1	CATALOG 5
AND	32	CATALOGER 1
ANGLO	1	CATALOGING 5
ANNOUNCEMENT	1	CATALOGS 1
ARBITRARY	1	CATALOGUING 1
ARCHIBALD	1	CELL 1
ARCHITECTURE	1	CENSUS 1
ARTICLES	2	CENTER 1
AS	6	CENTERS 3
ASIAN	1	CENTRALIZED 1
ASLIB	1	CHANGE 1
ASPECTS	1	CHANGING 4
ASSIST	1	CHARACTERISTICS 2
ASSOCIATION	1	CHARACTERS 2
ASSOCIATIONS	1	CHEMICAL 2
AT	3	CHILDREN 1
ATTRIBUTES	1	CHOICES 1
AUTOMATED	2	CIRCULATION 1
AUTOMATIC	3	CITATION 1
AUTOMATICALLY	1	CITATIONS 1
AUTOMATION	2	CITED 1
BACKGROUND	1	CLASSIFICATION 4
BANE	1	CLIPPING 1
BANKS	1	COLLECTED 1
BARRIER	1	COLLECTION 2
BASIC	2	COLLEGE 9
BEING	1	COLLEGES 1
BENCHMARK	1	COLLEGIATE 1
BENEFIT	1	COMCON 1
BETWEEN	1	COMMERCIAL 1
BIBLIOGRAPHIC	1	COMMON 1

ロール・カードの第2フィールドで ALL WORDS, 第3フィールドで NO AUTHOR の指定を行なえばよい。

#### A. コントロール・カードとオプション

コントロール・カードのフォーマットは表5のとおりである。

その第1フィールドには必ず WORDFR の指定がなければならない。

表 5. WORDFR コントロール・カード

Col. 1～	13～
WORDFR	SKIP mm/dd/yy

このセグメントを実行しない場合は、第3フィールドに SKIP と指定すればよい。実行する場合には、第2フィールドに日付をパンチしておく。

#### B. セグメントの作業

##### 1. コントロール・カードの解説

コントロール・カードを読み、第1フィールドのセグメント名を確認し、誤りがあれば、‘CONTROL CARD ERROR. RESET CARD.’ というメッセージをタイプアウトする。

第2フィールドに SKIP の指定があれば直ちにジョブを打ち切る。SKIP の指定がなければ、‘WORDFR LOADED’ のメッセージを出し、第2フィールドの日付を見出し印刷用エリアに転送し、‘SET MT 2 (INP)’ の指示をタイプアウトして pause 状態に入り、C/R によって次のステップに移る。

##### 2. 入力と頻度計算

SORPRI でソートされた KWOC File または WADEX File を MT #2 から読みとり、レコードの先頭のキーワードと、直前に get したレコードのそれとを比較し、一致すれば頻度カウンターに1を加える。一致しなければ、古いキーワードとその頻度を table に登録し、異り語数カウンターに1を加え、頻度を頻度総計カウンターに加える。

Table に112語が登録されると、印刷作業に移る。また、入力データが尽きると、最終ページの印刷作業に移るが、その前に table の残余部分にブランクを埋める。

##### 3. 印刷

印刷作業は3種に分れる。第1は見出しの印刷、第2は detail の印刷、第3は異り語数と頻度総計の印刷である。

第1と第3の作業は簡単であるが、第2の作業は、detail が56行の2段組となるため複雑である。2段組にするためには、table を前半と後半に区分し、前半部と後半部のそれぞれ対応する部分を順次 detail 印刷用エリアに転送して印刷する。

1ページ分の印刷が終ると、再び入力と頻度計算のルーチンに戻る。(出力例7参照)

## VIII. 辞書のリスト処理

リスト処理 (list processing) とは、データの集合を、その物理的關係によってではなく、論理的關係によって処理する手法であるが、ここでいう辞書のリスト処理とは、辞書を木構造によって表現し、処理することを意味している。

木構造辞書の自動割り付け、照合、変更の技法ならびにその経済性については、拙稿 ‘電子計算機による文字列の処理’<sup>5)</sup> で詳細に論じているので、本稿では省略する。

MISA では、リスト構造辞書の効率化を計るために、いくつかの工夫が試みられている。

#### A. 語幹形の採用

今仮に次の7語を辞書に登録するとしよう。

COMPARE  
COMPARED  
COMPARES  
COMPARING  
COMPARISON  
COMPARATIVE  
COMPARATIVELY

これらの7語をリストに割り付けると、22ワードを必要とする。ところが、これらの7語に共通な部分 COMPAR のみを語幹形として辞書に登録し、辞書を引く場合には、COMPARE の6字がマッチすれば、それ以降にどのような文字が続こうと、マッチしたものと見做すことにすれば、占有スペースは6ワードにすぎなくなる。

語幹形の語を通常形 (完全形) の語と区別するため、辞書に登録する際に ‘COMPAR+’ のごとく、その後にプラス記号を付加しておけばよい。

この語幹形は語尾変化の多い語を登録する場合にきわめて経済的であるが、その乱用は慎まねばならない。たとえば、COMPAR+ の場合には、前掲の7語以外に COMPARABLE, COMPARABLY, COMPART, COMPARTMENT の諸語も登録されたことになるからである。従って、語幹形の使用に当っては、登録したい語の

前後の語を辞書でよく調べなければならない。

語幹形の語を読み込んだ場合には、最後の文字を収容する節 (node) の第 7 ビットを ON にする。辞書を引く場合には、マッチした後にその節の第 7 ビットをテストし、それが ON であれば、それ以上の照合は行わず、辞書に登録された語と認定する。

筆者が図書館・情報学関係雑誌記事の処理に当って実験的に使用したストップ・ワードの辞書は通常形と語幹形の両者から成り、筆者が混合形と呼ぶタイプに属している。その実登録語数は 275 語であるが、語幹形を含んでいるので、有効語数は 585 語である。占有メモリーは 746 ワードであり、有効語 1 語の平均スペースは約 1.27 ワードであり、語幹形導入前の平均スペース 2.5 ワードの約半分にすぎない。

MISA では、プログラム・エリアと作業エリアを除くすべての使用可能なメモリーをリスト構造の辞書に使用できるようにしてあり、KWICED の場合、そのスペースには、現時点で 8261 ワードが確保されている。従って、収容可能な有効語数はおよそ 6500 語に達するものと推定される。

#### B. 索引の付加

頭文字 Z の入力ストリングがあるとして、頭文字のアルファベット順に編成されたリスト構造辞書を参照する場合には、Z の節 (第 1 レベル) に達するまでには、26 回の照合が必要である。

一般に、求められる第 1 レベルの節に達するのに必要

な照合回数の平均は 13.5 である。もし特殊な索引を用意して、求める第 1 レベルの節に直ちに到達することができれば、探索効率は飛躍的に上昇するはずである。この索引には、各頭文字の第 1 レベルの節の番地が格納されることになるが、通常の逐次探索方式を索引の照合に適用するならば、その照合回数の平均は同じく 13.5 となり、索引を付加する意義がまったくなくなる。従って、この索引はダイレクト・アクセス方式のものでなければならない。

MISA では、このようなダイレクト・アクセス方式の索引が辞書の割り付けに平行して自動的に作成され、探索効率を高めている。(三田研究・教育情報センター)

- 1) 安西郁夫. “雑誌論文引用形式の簡略化,” *Library science*, no. 4, 1966, p. 173-80.
- 2) Bishop, Charles “An integrated approach to the documentation problem,” *American documentation*, vol. 4, no. 2, April 1953, p. 54-65.
- 3) 橋本昌幸, 中嶋 淳. “電子計算機による機械検索システム I: 自動索引, 検索および情報選択提供, (SDI),” 第 3 回 Documentation 研究集会発表論文集, 1966, p. 207-14.
- 4) 斎藤 孝. 電算機による情報の蓄積と検索システムの設計. 慶応義塾大学修士論文, 1968, 75 l.
- 5) 安西郁夫. “電子計算機による文字列の処理—基本的技法と応用,” *Library and information science*, no. 7, 1969, p. 109-16.