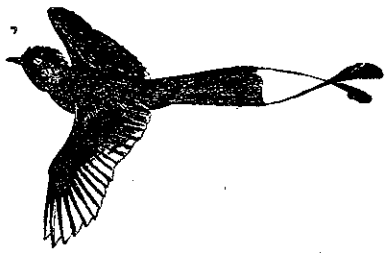


A PRELIMINARY REPORT ON THE COMPUTERIZED LIBRARY INFORMATION RETRIEVAL SYSTEM

R. Gowri
Centre for Theoretical Studies

and

Malati Hegde
Centre for Ecological Sciences
Indian Institute of Science
Bangalore-560012

January 1986

ENVIS-DIC TECHNICAL REPORT NO. 2

Centre for Ecological Sciences
Indian Institute of Science
Bangalore-560 012
INDIA

Introduction

This report deals with the Computerised library information retrieval at the ENVIS Centre at CES, Indian Institute of Science, Bangalore. Science and technology is growing very fast. Research in these fields has generated enormous quantity of out put. This growing magnitude of documents (information) increases the risk of missing a relevant document(information) on a particular topic, when the retrieval is manual. Hence the need for computerising the information retrieval procedure. Dec 1090 installed at I.I.Sc Campus has been used for this work.

The queries about the information are expected to come from any part of the country, hence our system has been designed to work on off-line. Programs in the system have been written in COBAL for its sophisticated file handling features. The queries about the information can be either on Author name or on Title or on Subject or any Combination of these. The search on the collection is done depending on the query and the output of the search consists of the entire information about the documents intended for.

Methodology

Search through Author - name or Title is done by simple string matching in the appropriate files created for this purpose. This is a one level search. Whereas the search through the subjects is divided into 5 non intersecting levels.

a. level 1 says about the Geographical location
b. level 2 says about the Taxon

3rd, 4th and 5th levels are hierachical division of the subject.

The key words in each level are chosen to cover the collection at this Centre. The set of keywords under each level is given below(Appendix.)

APPENDIX

Level 1

GEOGRAPHICAL

Keyword	Code
ASIA	01
EUROPE	02
AFRICA	03
OCEANIA	04
ANTARCTIKA	05
NORTH AMERICA	06
SOUTH AMERICA	07
MID EAST	08
SOUTH ASIA	09
SOUTHEAST ASIA	10
FAR EAST	11
ATLANTIC	12
PACIFIC	13
INDIAN OCEAN	14
ARABIAN SEA	15
BAY OF BENGAL	16

WEST COAST	17
EAST COAST	18
HIMALAYAS	19
WESTERN GHATS	20
EASTERN GHATS	21
VINDHYAS	22
SOUTH INDIA	23
NORTH INDIA	24
INDIAN DESERT	25
NORTH EASTERN INDIA	26
ANDHMAN AND NICOBARS	27
GANGA	28
KAVERI	29
GODAVARI	30
KRISHNA	31
TUNGABHADRA	32
SHARAVATI	33
AGHANASHINI	34
GANGAVALI	35
KALI	36
MAHARASHTRA	37
GOA	38
KARNATAKA	39
TAMIL NADU	40
KERALA	41
GUJARATH	42
SAHYADRI	43
NILGIRI	44
ANNAMALAI	45
HIGH RANGES	46
CARDAMOM HILLS	47
AGASTYAMALAI	48
MALNAD	49
KHARAVALI	50
MAIDAN	51
UTTARA KANNADA	52

SHIMOGA	53
CHIKMAGALUR	54
DAKSHINA KANNADA	55
COORG	56
HASAN	57
INDIA	58
WESTERN INDIA	59
EASTERN INDIA	60
CENTRAL INDIA	61

TAXON

VIRUSES	001
BACTERIA	002
INVERTEBRATES	003
VERTEBRATES	004
PROTOZOA	005
SPONGES	006
COELENTERATES	007
WORMS	008
MOLLUSCS	009
ECHINODERMS	010
ARTHROPODA	011
CRUSTACEA	012
INSECTA	013
HYMENOPTERA	014
TERMITES	015
ANTS	016
WASPS	017
BEES	018
FISHES	019
AMPHIBIA	020
REPTILES	021
SNACES	022
TURTLES	023
CROCODILES	024
LIZARDS	025
BIRDS	026

RUTACEAE	064
SOLANACEAE	065
VERBINACEAE	066
CYPERACEAE	067
POACEAE	068

SUBJECT MATTER - LEVEL I

KEY WORD	CODE
POPULATION ECOLOGY	01
POPULATION GENETICS	02
COMMUNITY ECOLOGY	03
PRODUCTIVITY	04
BIOGEOGRAPHY	05
SOCIOBIOLOGY	06
ETHOLOGY	07
GENETICS	08
EVOLUTIONARY BIOLOGY	09
RESOURCE MANAGEMENT	10
ENVIRONMENTAL EDUCATION	11
ENVIRONMENTAL ACTIVISM	12
BIOMETRY	13
SOCIAL SCIENCES	14
HISTORY	15
PHILOSOPHY	16
SYSTEMATICS	17
ANATOMY	18
PHYSIOLOGY	19

SUBJECT MATTER LEVEL II

KEY WORD	CODE
COMPETITION	001
SYMBIOSIS	002
PREY PREWATOR	003
PARASITISM	004

ALTRUISM	005
CO-OPERATION	006
SOCIAL GROUPS	007
NICHE	008
DIVERSITY	009
TERRITOREALITY	010
SOCIAL HIERARCHY	011
CASTES	012
LIFE HISTORY	013
REPRODUCTIVE STRATEGIES	014
HABITAT PREFERENCE	015
LEARNING	016
COMMUNICATION	017
SOIL	018
WATER RESOURCES	019
FLORA	020
FAUNA	021
MICROBIOTA	022
WASTELANDS	023
FORESTS	024
LAKES	025
RIVERS	026
ESTUARIES	027
SEACOAST	028
OPENSEA	029
ISLANDS	030
AGRICULTURE	031
HUNTING GATHERING	032
POACHING	033
HORTICULTURE	034
ANIMAL HUSBANCRY	035
FISHERIES	036
AGROFORESTRY	037
SYLVIPASTORAL SYSTEMS	038
ENERGY	039
INDUSTRY	040

MINING	041
NATURE RESERVES	042
BIOSPHERE RESERVES	043
TRANSPORT	044
HUMAN HABITAT	045
HEALTH	046
EDUCATION	047
EXTINCTION	048
ETHNOBOTANY	049
PHENOLOGY	050

SUBJECT LEVEL III

EXPERIMENTS	01
OBSERVATION	02
THEORY	03
ECODEVELOPMENT	04
CONSERVATION	05
PLANNING	06

Numerical code that appears against each keyword in the above list is used while searching. The search through keyword is initiated after accepting the keyword under each level. The code equivalent of the keyword is obtained from the KEY-DOC-FILE and search for information is done in KEY-FILE using this code.

FILE Structures

There are three main indexed sequential files. Their file structure's are as follows.

1. BOOK-FILE (it is indexed sequential)

SL No	Author Name	Title	Keywords				
6ch.	30 ch.	120 Ch	2	3	2	3	2

178 Ch

2. REPRINT-FILE (It is indexed sequential)

Sl No	Author name	Title	Keywords					Reference
6 Ch	30 Ch	120 CH	2	3	2	3	2	120 Ch

298 Ch

3. TREPORT-FILE (it is indexed sequential)

Sl No	Author name	Title	Keywords					Reference
6 Ch	30 Ch	120 Ch	2	3	2	3	2	120 Ch

298

The other 12 files required in the system are generated from the above three files. Their file structure is given below-

4. AUTHOR-FILE (Indexed sequential)

Sl No.	Author name	Link
6 ch	30 ch	6 ch

41 ch

Steps involved in obtaining this file.

Step 1 The above three main files are sorted on Author's name

Step 2 The different documents of the same Author are identified

Step 3 Record serial number is used to link the same Author documents in AUTHOR-FILE

5. TITLE (Indexed sequential)

Sl No	Title (Contracted)
6 ch	120cch

126 ch

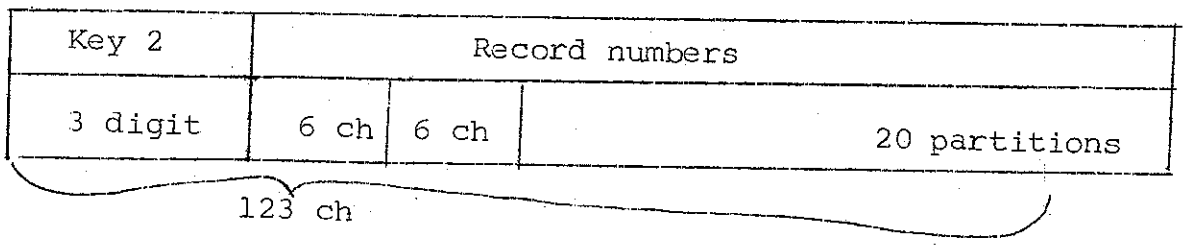
This file is created from the main files by dropping the common words as 'THE', 'OF', 'TO','FROM', 'AND','OR','IN','ON', 'AN','A'. The reason for dropping these common words is that trivial interchange or of missing of these common words may lead to mismatch in searching and hence missing of a relevant document.

6. KEY1-FILE (indexed sequential)

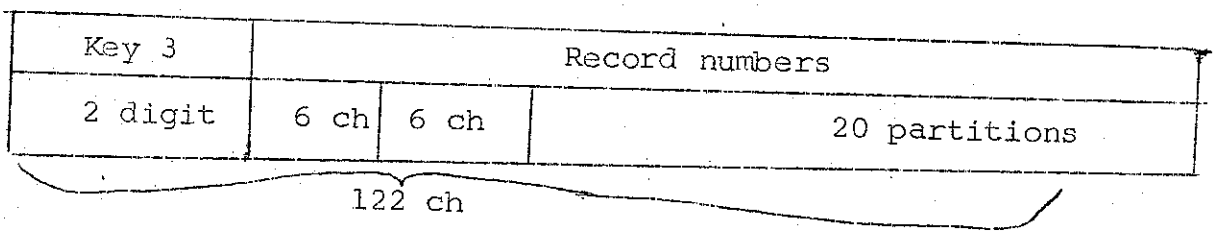
KEY1	Record numbers		
2 digit	6ch	6ch	20 partitions

122 ch

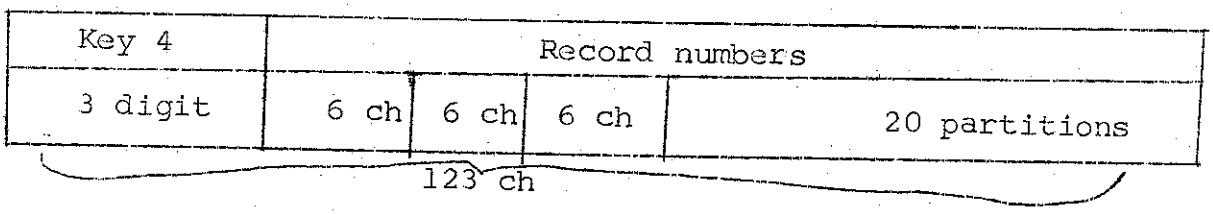
7. KEY 2-FILE (indexed sequential)



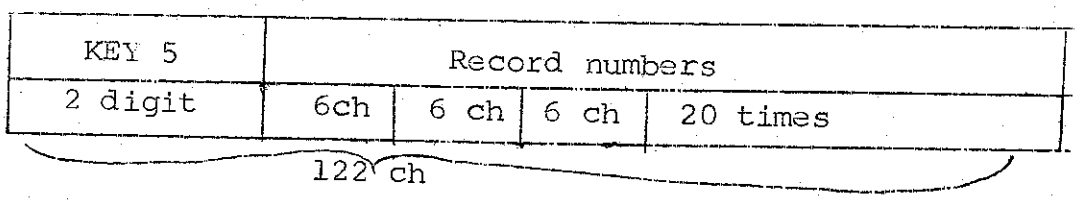
8. KEY 3-FILE (indexed sequential)



9. KEY 4-FILE (indexed sequential)

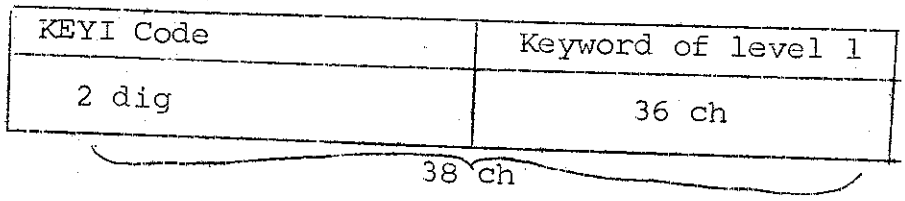


10. KEY 5-FILE (indexed sequential)



These files gives the record serial number of documents for each key word in each levels. These files are of vital importance when search in through subject.

11. KEYI -DOC-FILE



The file structure of other KEY-DOC FILES are similar.

SYSTEM FLOW

The system displays the message 'SEARCH THROUGH?'.

The answer is accepted in the counter MESSAGE which can be

'A'	for AUTHOR
'T'	for TITLE
'K'	for KEYWORD (subject)
'AT'	fpr AUTHOR + TITLE
'AK'	for AUTHOR + KEYWORD
'TK'	for TITLE + KEYWORD
'ATK'	for AUTHOR+TITLE+KEYWORD

These are three different procedures (paths) system to do the search on 'AUTHOR' , 'TITLE' and 'KEYWORD'. For the remaining four combinations the search is done by executing the appropriate combinations of these three procedures.

Logic flow of the procedures is as follows.

Main

Step. 1

IF MESSAGE = 'A' PERFORM PROCEDURE 1
IF MESSAGE = 'T' PERFORM PROCEDURE 2
IF MESSAGE = 'K' PERFORM PROCEDURE 3
IF MESSAGE = 'AT' PERFORM PROCEDURE 1 & 2
IF MESSAGE = 'AK' PERFORM PROCEDURE 1 & 3
IF MESSAGE = 'TK' PERFORM PROCEDURE 2 & 3
IF MESSAGE = 'ATK' PERFORM PROCEDURE 1 & 2 & 3
ELSE DISPLAY 'WRONG SEARCH'

Procedure 1

Step 1. DISPLAY 'AUTHOR'S NAME?'
ACCEPT IN answer AUTHOR

Step 2. DISPLAY 'WANT TO SPECIFY the CATEGORY?'
Accept in answer B,R,T.N.

Step 3. COUNT=0
SEARCH AUTHOR-FILE (indexed sequential)
Sequentially till the author-name matches.

Step 4. If the category is not specified
Access the respective indexed sequential files (Book-File, or Reprint-File or Treport-File) and COUNT-COUNT+1 write the details of the document.

Step 5. If category is specified
Details of the documents is written after matching the category ie if category do not match just ignore the document and go further.

Step 6. Otherwise access the next match with the help of the link till link=0
If COUNT=0 DISPLAY
'Sorry no documents on this AUTHOR'

Procedure 2

Step 1. DISPLAY 'ENTER THE TITLE'

Step 2. DROP the stop words (Common words) from TITLE and form contracted TITLE

Step 3. Search sequentially in TITLE-FILE (indexed sequential) till the match occurs. When match occurs match the category go to the respective file and print the details of the documents.

Step 4. If match is not found and till the end display
'Sorry this document is not in our Library collection!'

Procedure 3

Step 1. Accept the keyword (not the code) at 5 level

```
DISPLAY 'LEVEL 1?'      KEY 1
DISPLAY 'LEVEL 2?'      KEY 2
DISPLAY 'LEVEL 3?'      KEY 3
DISPLAY 'LEVEL 4?'      KEY 4
DISPLAY 'LEVEL 5?'      KEY 5
```

Step 2. Get the equivalent numerical code from corresponding KEY-DOC-FILES

Step 3. With the help of numerical code records are obtained from KEY-FILES

Depending upon the AND and OR on the quiery the conjugation is done.

Step 4. From the record number we can get the category.

Step 5. Go to respective master files from category and by record number matching details about the documents can be printed and COUNT=COUNT+1

Step 6. If Count=0

DISPLAY 'No documents on these keywords'

The printed information about the documents will contain

AUTHOR NAME TITLE REFERENCE

The documents in the ENVIS Centre are broadly classified into 3 groups

a) WESTERN GHATS
b) ECODEVELOPMENT
c) BIOLOGICAL DIVERSITY

The collection under each classification is arranged in the alphabetical order on author's name. Once the information is present in the collection and with the out put locating the documents in the Library is not difficult.