TECHNICAL REPORT No.TR-03

INDIAN INSTITUTE OF TROPICAL METEOROLOGY

PASHAN, PUNE - 411 008

COMPUTER DIVISION

FILE EXTRACT UTILITY

MRS. R.R. JOSHI
MR. P.W. DIXIT

AUGUST 1992

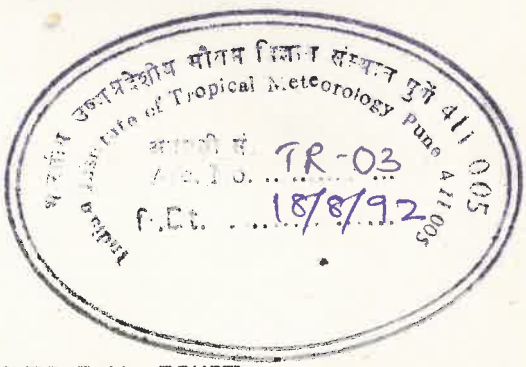
INDIAN INSTITUTE OF TROPICAL METEOROLOGY

PASHAN

PUNE : 411 008

COMPUTER DIVISION

FILE EXTRACT UTILITY



MRS. R.R. JOSHI AND SHRI P.W. DIXIT

INDIAN INSTITUTE OF TROPICAL METEOROLOGY, PUNE 8.

---

01 Security classification: Unclassified 02 Distribution: Unrestricted

---

03 Report status: New 04 Series: CPT 05 Report No.: 03/92

---

06 Report Type: Technical Report-Utility

---

07 Title and subtitle: FILE EXTRACT UTILITY

---

08 Personal author(s): Mrs. R.R. Joshi and Mr. P.W. Dixit

---

09 Affilliation author(s): Indian Institute of Tropical Meteorology,Pnue 8.

---

10 Corporate author(s): ---

---

11 Originating unit: Computer & Data Division

---

12 Sponsor(s): Name: In house

---

13 Date of submission: August 1991 Publication/Issue Date: August 1992

---

14 Form of distribution: Hard copy

---

15 Language of text: English

---

16 No. of references: One

---

17 Abstract: This document describes the procedures for handling File Extract utility of ND-system. The document is self explanatory and describes salient features of this utility. It is also useful to handle large data.

---

18 Keywords: File Extract,Input file,Output file,Extract specification,Append

---

19 Whether Refereed: Yes

TECHNICAL NOTE : FILE EXTRACT

Introduction: File extract is one of the many utilities available on ND-560/CX supermini computer system. In this report authors have tried to present a brief description about it's usage. Details are provided where ever felt necessary. An effort is also made to highlight the points which are not given in concern manual but are crucial in the proper use of this utility and in understanding of the relevant features. Authors have tried to document this report in simple terms by giving as many examples as possible to facilitate it's understanding by the new users. File extract is the utility which can extract records from one file and after suitable modifications write onto another file or output device by giving simple commands.

File extract package has ND - Version 25.

Invocation : File extract can be invoked through SINTRAN command.

@ FILE-EXTRACT CR

There are four parameters required to be given for correct extraction. INPUT FILE, OUTPUT FILE, EXTRACT SPECIFICATION, OUTPUT RECORD LAYOUT SPECIFICATION.

Facilities in File Extract:

1) The extraction of subsets from files based on record numbering: To extract 10 records say from record no. 21 to 30. Example:

@ F-E CR

INPUT FILE: TEST

OUTPUT FILE : "TEST1"

EXREACT SPECIFICATION : 21-30

OUTPUT RECORD LAYOUT SPECIFICATION : CR

The output file contain records from 21 to 30.

2) The extraction from files based on record contents:
For this facility there are several extract specifications which are discussed in extract selection specification on page number 4-7.

3) Rearranging of files:
To interchange the contents of columns 50-60 and 10-20 ,it is possible to do so by giving information in output record layout. Other parameters remain same as above.

OUTPUT RECORD LAYOUT SPECIFICATION : 50-60,10-20

The output file will contain information interchange with columns 10-20 and 50-60 .

4) Appending of files or subset of files to other files:
To append full input file or output of file extract on given output file.

Example:

```
@ F-E CR
INPUT FILE : TEST
OUTPUT FILE : TEST1,A
EXTRACT SPECIFICATION : 20=5
OUTPUT RECORD LAYOUT SPECIFICATION :CR
```

This example will extract the records which have 5 in column 20 from file TEST and append to existing file TEST1. Output file must be a existing file and is not valid for devices like TERMINAL or LINE PRINTER. By giving CR in extract specification whole input file will be appended to the file TEST1.

5) Reformatting of files according to record layout,length and organization:
This is useful to create file of fixed record length.

6) File splitting by one run:
To split input file in two files, one which contains extracted portion and another which contains non extracted portion. It can be done by giving a colon (:) at the end of the OUTPUT FILE.

Example:

```
@ F-E CR
INPUT FILE : TEST
OUTPUT FILE : "TEST1":
SPLIT OPTION OUTPUT FILE : "A"
```

Then file A will contain non extracted portion and TEST1 will contain extracted portion. Other parameters remain same.

7) Generation of readable reports:
It contains heading and page numbering routed to TERMINAL, LINE PRINTER or any disk file.

Example :

```
@ F-E CR
INPUT FILE : TEST
OUTPUT FILE: "TEST1"
EXTRACT SPECIFICATION : CR
```

OUTPUT RECORD LAYOUT SPECIFICATION : Hnn,PAGE

where nn is the number of lines per page. It will write a heading mask to the terminal and wait for user heading. PAGE is the optional parameter ; which gives page number at the top of the page. If output file is TERMINAL then give two carrige returns after heading. After heading give two spaces and then give carriage return.

8) Saving of input parameter in mode files for later automatic processing. ($AUTO)

Details of parameters of File Extract:

1) Input file: Input file should be any randomly accessable SINTRAN text file. There are four additional facilities which can be invoked by different commands in this parameter. $MODE, $AUTO, $KEY and Fnnn. Details are as follows :

i) $MODE : By selecting this option text will appear on the terminal screen.

```
@ F-E CR
INPUT FILE : $MODE
MODE SAVE FILE: "A:SYMB"
INPUT FILE :  TEST
OUTPUT FILE : "TEST1"
```

Whatever information feeded for the commands it is stored in file A. This A file will serve as output file for commands. Contents of A file are as follows:

```
@F-E
TEST
"TEST1"
1-4=4213
20-24,1-20
```

Here F-E invokes the File extract. TEST is the input file. TEST1 is the output file. 1-4=4213 is the extract specification and 20-24, 1-20 is the output layout specification.

ii) $AUTO: By selecting this option commands in input file are invoked.

Example:

```
@ F-E CR
INPUT FILE : $AUTO

AUTO RUN TIME COMMAND FILE : B:SYMB
```

B file is a existing file from which information is taken for required commands automatically. e.g. name of the input file, output file, extract specification etc. Such files are to be created in PED i.e. Program Editor. Contents of the B file are as follows:

```
TEST
"TEST1"
1-4=4213
20-24,1-20
```

Here File extract has already invoked by command F-E. Information for parameters are taken one by one from this file i.e TEST as input file, TEST1 as output file etc.

iii) Fnnn: Fixed record length input file option: To process a fixed record length input file not containing CR and LF this option is used. nnn indicate input file record length (maximum 1024 bytes).

Example:

@ F-E CR

INPUT FILE : TEST,F48

Here TEST is the file which contains data of fixed record length - 48 other parameters remain same.

2. Output file: Output file may be any existing / non existing SINTRAN III disk file or output device viz. LINE PRINTER or TERMINAL. It is not possible to use magnetic tape as output file. APPEND and file split option can be used under this parameter, which are discussed earlier in file extract facility no. 4,6. respectively.

3. Extract selection specification: One input line is allowed for giving specifications. There are four types of selections available.

i. Specification of input file record intervals: Subset of the input file can be selected. Also it is useful for rearranging files. Examples can be seen in facilities in file extract (1) and (3) resp.

Syntax : Start record no. - End record no.

Record numbers are specified with 1 to 9 digits. Blank(^) is the terminator may be followed by parentheses or any other extract selection criteria.

ii. Specification of input record field values :

a) Numeric field evaluation:

Syntax : Start position - End position < operation code >
minimum value - maximum value

In case of one digit field end position is not necessary.

Start position : It is the start byte number of numeric field within input record.

End position : It is the end byte number of numeric field within input record.

Operation code : = equal to
not equal to
> greater than
< less than
Minimum value : It is the minimum numeric value to be compared
Maximum value : It is the maximum numeric value to be compared

Example : 1-5 = 42182

This means it will extract numeric value 42182 which is lying between 1 through 5 bytes from input file. Here there is only one value (minimum value). But for range two values are required.

Example : 1-5 = 42182 - 43063.

b) Text field evaluation :

Syntex :

< start position > - End position < operation code > < "text string" >
Where ;
Start position : It is the start byte number within input record to be evaluated.
End position : It is the end byte number within input record to be evaluated may be omitted for one byte field.
Operation code : = equal to
not equal to
Text string : Text string may contain any character and must be within double quotes. Length of the text string must be same as the field length specified by the start / end position. If text string is shorter, it is acceptable but if it is longer than given limit then the program is terminated with error message.

Example :

1) 45 - 50 = "OSLO5"
2) 45 - 50 = "EOX" - Acceptable
3) 45 - 50 = "RESEARCH" - Not Acceptable

iii) Specification of text string which are occured/ not occured within a record :

A text string search specification will cause the entire input record to be scanned for the existence of the given text strings.

Syntax : Text < Operation code > < "Text string" > where,

Text : specifies within the entire record
Operation code : = equal to
not equal to
Text string : any text enclosed by double quotes

Example :

TEXT = "INDIAN"

iv) Specification of a text string which is to occur/not occur within a specified subset of a record :

A limited text string search will cause the specified subset of the input record to be scanned for the existence of a given text string.
Syntax : < Start position > - < End position > < Operation code > < "Text string " >

Start position/ End position: It is the start/end byte number within an input record where the text search is to be done.

Operation code : = equal to
not equal to
Text string : The search text string may contain any character except double quotes and must be enclosed by double quotes.

Example :

1) 40 - 45 = "INDIA"
2) 40 - 70 = "INDIA"
3) 40 = "I"

Selection criteria may be connected by the logical operands like AND,OR etc,. These can be used together with the parentheses nesting.

Syntax :

< Extract criteria A > < logical operand > <Extract criteria B >

Extract criteria A and B are the criteria discussed above logical operand : These are two operands :

'AND' : Both A and B criteria must be fulfilled
'OR' : Either A or B must be fulfilled

Example :
12 - 15 = 6166-9265 .AND. 1-5 ="INDIA"

Parentheses Nesting : It is used in complex selections. Extract criteria along with logical operand can be combined in parentheses. A Start/End parentheses must be placed before/after extract criteria or another parentheses.

Example :

((1-2 = "T1" .OR. 1-2 = "T2").AND.10=2).AND.(15-22 > 90000.OR.23 = "*")

Meaning of this criteria is select those records of type T1 or T2 which has code 2 in column 10 and also either has amount more than 90,000 or indication of * in column 23.

4) Output layout specification:
There are two types of specifications :

i. Specification of output record layout as one or more of the following elements:

a) A copy of input record:

By giving CR or by giving the length of record input record will be copied with suitable specifications.

Example :

```
@ F-E CR
INPUT FILE  : TEST
OUTPUT FILE : "TEST1"
EXTRACT SPECIFICATION : 1-20
OUTPUT LAYOUT SPECIFICATION : CR
```

b) Subset of input record : Subset of input record can be used as form output record.

Example :

```
@ F-E CR
INPUT FILE : TEST
OUTPUT FILE : "TEST1"
EXTRACT SPECIFICATION : CR
OUTPUT LAYOUT SPECIFICATION : 50-55,1-20
```

This will produce output record containing characters 50 through 55 and then first 20 characters.

c) Imbedded constants: Any extra constant can be added in between two subsets.

Example :

```
@ F-E CR
```

INPUT FILE ,OUTPUT FILE and EXTRACT SPECIFICATION parameters remains same as above but,

```
OUTPUT LAYOUT SPECIFICATION : 50-55,"ABC",1-26
```

d) Input record number: This option will produce an output record containing a successive five digit record numbering as it's first element.

```
OUTPUT LAYOUT SPECIFICATION : L
```

ii. Specification of output environment :

a) SHOW : First input file record will be written on terminal together with position mask line. This is useful to see the position number.
b) Wnn : Wait obtion to be used when the terminal is used as output file. nn - number of lines to be written before wait.
c) Hnn : Heading option to be generated with one line heading. nn - number of lines per page this can be used together with page number option.
d) Enn : It is also heading option for predefined heading.

e) Pnn : It produces a position mask as predefined heading.

Symbols used : @ : IITM >
CR : Carriage Return

**** INFORMATION UNDERLINED SHOULD BE GIVEN BY USER ****

Acknowledgement: The authors wish to express their grateful thanks to Shri. D.R. Sikka and Shri. R. Suryanarayana, for their encourgement and providing facilities. Authors also thank to Shri S.S. Aralikatti his valuable help. Authors wish to thank Shri. R. Vijaykumar for reviewing this technical report and offering constructive suggestions. The inclusion of his suggestions have lead to a significant improvement in the presentation of text.

Reference: SINTRAN III utilities Manual No. 60.151.02