# 360 0. S. FORTRAN IV <br> FREE FIELD INPUT/OUTPUT SUBROUTINE PACKAGE 

by

R. W. Doran

This research was supported in part by the United States Air Force Electronic Systems Division under Contract F19628-C-0035.

## STANFORD UN IVERS ITY COMPUTER SC IENCE DEPARTMENT COMPUTATIONAL LINGUISTICS PROJECT OCTOBER 1967

## Introduction

Programmers dealing with aspects of natural language processing have a difficult task in choosing a computer language which enables them to program easily, produce efficient code and accept as data freely written sentences with words of arbitrary length. List processing languages such as LISP are reasonably easy to program in but do not execute very quickly. Other, formula oriented, languages like FORTRAN are not provided with free field input.

The Computational Linguistics group at Stanford University Computer Science Department is writing a system for testing transformational grammars. As these grammars are generally large and complicated it is important to make the system as efficient as possible, so we are using FORTRAN IV (O.S. on IBM 360-65) as our language. To enable us to handle free field input we have developed a subroutine package which we describe here in the hope that it will be useful to others embarking on natural language tasks.

The package consists of two main programs, free field reader, free field writer, with a number of utility routines and constant COMMON blocks.

## Free Field Reader (FREAD)

FREAD was written to enable us to read free field input including data containing longwords, special characters, and integers and to eventually control the system by the form of its input with the detection of orders.

FREAD is a REAI*8 function with one dummy argument (INIEGER*2). When it is called it returns the next entity it finds in the input stream. It communicates information concerning that entity through labelled common viz.

INTEGER*2 NUM, ISPEC
IめGICAL*I NUMFI, $\emptyset R D F L$

The 'input stream' refers to the sequence of characters from the punched card input, with column 72 of a given card assumed to be followed hard by column 1 of the next. Double quotes " are omitted from the input stream along with any characters between consecutive double quotes. This gives us the facility of allowing comments anywhere in the input.

The different entities are as follows:
1/. Special character.
This is taken to be any character whose EBCDIC 8 bit code has a value of less than 128 (except for blank, ", \$).

2/. Integer.
A sequence of digits whose value is less than $2^{15}$.
An integer may be followed immediately by a word.

3/. Word.
A sequence of alphameric characters (i.e., those with 8 bit EBCDIC codes of value greater than 128) of length less than or equal to 40 (40 is arbitrarily set as 'being of arbitrary length') and not beginning with a \$.

4/. Order
A word of length less than 8 and beginning with a $\$$.

As FREAD encounters each entity it returns a value and sets conditions as follows:

1/. Special character.
FREAD is set equal to the character followed by 7 blanks. ISPEC is set to the value of the character's 8 bit code (see appendix 1). NUMFL and $\emptyset$ RDFL are .FALSE.

2/. Integer.
FREAD is set to the value of the integer. If the integer has a too-large value a system overflow interrupt will occur. ISPEC is zero, NUMFi is .TRUE., $\emptyset$ RDFI is .FAISE. (Here it would make good sense to have ISPEC carry the value of the integer. Our method is left over from the 7090 where $\operatorname{FREAD}$ was itself an integer f-unction, a-user could easily change this.)

## 3/. Worã.

If the word is cf length less than or equal to 8 characters it is itself returned as FREAD (with trailing blanks), if it is of greater length a keyword is returned. ISPEC is zero, NUMFL and $\not \subset$ RDFL are
. FALSE.
The keyword of a longword is obtained by replacing the tail end of the first 8 characters by an integer of value according as to the
position of this longword in the sequence of those encountered. For example, if PRめNøMINALIZATION is the 15 th longword encountered it will always have the keyword PRめNфM15.

If a word of more than 40 characters is encountered FREAD prints an error message and truncates the word.

4/. Order
The order is returned as a word. ISPEC is zero, NUMFL is .FALSE., $\varnothing$ RDFL is .TRUE.

If an order has more than 8 characters an error message is written and the keyword returned (this is because keywords can alter as the input alters, but orders should be definite if they are to be of any use).

Example.
Suppose in a given run FREAD encounters the following input

it will behave in the order of the following table.

| Entity | FREAD | ISPEC | $\phi$ RDFL | NUMFI |
| :---: | :---: | :---: | :---: | :---: |
| \$START | \$START | 0 | T | F |
| ( | ( | 77 | F | F |
| R | R | 0 | F | F |
| $=$ | $=$ | 126 | F | F |
| 12 | value 12 | 0 | F | T |
| NP | NF | 0 | F | F |
| ) |  | 93 | F | F |
| , | , | 107 | F | F |
| PRON¢MINALIIE | PR N $^{\text {¢ }}$ MII | 0 | F | F |
| , | , | 107 | F | F |


| Entity | F'READ | ISPEC | $\emptyset \mathrm{RDFL}$ | NUMFL |
| :---: | :---: | :---: | :---: | :---: |
| \$\$N\$ | \$\$N\% | 0 | T | F |
| , | , | 107 | F | F |
| N1\$1\$M67N | N1.1 1 \$M62 | 0 | F | F |
| , | , | 107 | F | F |
| 622 | value 622 | 0 | F | T |
| N | N | 0 | F | F |
| . | , | 75 | F | F |

Further operational points

I/, As illustrated in the previous example $N 1 \$ 1 \$ \mathrm{M} 67 \mathrm{~N}$ returns NI \$1\$M62 which is an ambiguous keyword. Such words should be avoided as they will cause trouble on output.

2/. FREAD must be initialized by calling the subprogram INTIIZ. This initializes all the tables of longwords so more than one independent run may be made in the same job.

3/. FREAD performs a readout, printing each card image as it is read. NUM in / $\varnothing$ RDCM/ is used if one wishes to number the input, If an integer is placed in NUM it will be written out in the left hand margin of the next card read out,

4/. A listing of the longwords and their keys may be obtained by calling LiNg $\emptyset$ UT.

5/. When interpreting the input from FREAD it is handy to have constants representing the 8 bit values of the various characters, For this purpose a common block/CNSTCM/ is provided (see the example programs in appendix 2 for a listing). /CNSTCM/ is initialized in a block data subprogram which is part of the package.

6/. Lower case letters may be used if the system you are using allows you to print them.

7/. The present capacity for longwords is between 50 and 250 . (Most likely around 200.) This capacity can be readily increased if desired.

FROUT is a subroutine with which one can write out data in a free field format with keywords which have been read by FREAD expanded if desired．When designing a free field writer it is difficult to allow enough flexibility to obtain an aesthetically arranged printout， The options we have included allow one to obtain a reasonable－looking result in most cases．

FROUT gets it data from a linear sump KSUMP in labelied common ／MAINCM／viz．

CめMMめN／MAINCM／CHRTR，KSUMP，ISUMP，NCHRTR

REAL＊8 CHRTR，KSUMP（2000）
INTEGER＊2 ISUMP，NCHRTR

FROUT has 7 arguments；the first，ISTART，is INTEGER＊2，the remainder BI－－－B6 are IめGICAI＊1．

When FROUT is called it writes out the contents of KSUM？，from ISTART＋1 up to ISUMP，in a compacted free field form with the following options：

BI：．TRUE．Print and punch
．FALsE．Print only
B2：．TRUE．Expand keywords
．FALSE．Do not expand keywords

B3：．TRUE．Start a new line on a period
．FALSE．Don＇t
B4：．TRUE．Start a new line on a comma
．FAISE．Don＇t
B5：．TRUE．Start anew line on an＝
．FALSE．Don＇t

B6: .TRUE. Sequence number the output lines .FALSE. Don't

Further Operational Points

1/. To obtain greater generality of use we assume that no subprogram which places data in KSUMP initializes ISUMP. Care should be taken in remembering to initialize ISUMP in the main control routine,

2/. The keywords of longwords encountered by FREAD may be placed in the sump by calling the subroutine KEXSM. This enables you to obtain a punched listing of the longwords in order, which when placed in front of input-to a later job will ensure that the longwords receive the same keywords as before.

3/. For convenience when filling KSUMP a common block/FCSTCM/ is provided which contains the special, symbols in REAL*8 $\uparrow$ OM, For a listing of /FCSTCM/ see the example programs in appendix 2 . The names of the constants are the same as for /CNSTCM/ (see appendix 1) with the initial $N$ replaced by $F$. /FCSTCM/ is initialized by a block data subprogram which is part of the package.

4/. The output of FROUT is compacted, but if blanks are desired they may be inserted by placing FBIANK into KSUMP.

5/. /MAINCM/ includes two constants, PAGE and REC $\varnothing$ RD. When FRDUT encounters PAGE in KSUMP it, skips to the head of a new page when printing (new card when punching); when it encounters REC $\varnothing$ RD it skips to the start of a new line.

6/. FROUT should never fail on its input unless it is asked to output an integer of value greater than $2^{15}-1$, in which case a system overflow interrupt will occur; If FROUT is given a bad keyword it

for ISPEC if needed.

subprograms that use FREAD, e.g. returning an order to your control
7/. CHRTR \& NCHRTR in /MAINCM/ are used for communicating with
will write an error message.

Table of special characters in /CNSTCM/ and their codes

| Name in /CNSTCM/ | Character | Punch codë | Hexadecimal code | Decimal value |
| :---: | :---: | :---: | :---: | :---: |
| NBLANK | blank |  | 40 | 64 |
| NCENT | $\phi$ | 12-8-2 | 4A | 74 |
| NST $\dagger$ P | - | 12-8-3 | 4B | 75 |
| NLESS | $<$ | 12-8-4 | 4 C | 76 |
| NLEFTP | $($ | 12-8-5 | 4 D | 77 |
| NPLUS | + | 12-8-6 | 4 E | 78 |
| NL $\chi_{\text {R }}$ | 1 | 12-8-7 | 4F | 79 |
| NLAND | \& | 12 | 50 | 80 |
| NXCIM | - : | 11-8-2 | 5. | 90 |
| NDOLLR | \$ | 11-8-3 | 5B | 91 |
| NSTAR | * | 11-8-4 | 5 C | 92 |
| NRITEP | ) | 11-8-5 | 5 D | 93 |
| NSC $\varnothing$ IN | ; | 11-8-6 | 5E | 94 |
| NLIN $\phi$ T | $\square$ | 11-8-7 | 5F | 98 |
| NMINUS |  | 11 | 60 | 96 |
| NSLASH | 1 | 0-1 | 61 | 97 |
| N1211 |  | 12-11 | 6 A | 106 |
| NC $¢$ MMA | , | 0-8-3 | 6 B | 107 |
| NPERC | \% | 0-8-4 | 6 C | 108 |
| NLINE | - | 0-8-5 | 6 D | 109 |
| NGREAT | > | 0-8-6 | 6 E | 110 |
| NQUERY | ? | 0-8-7 | 6.5 | 111 |
| NC $\chi_{\text {L }}$, $\dagger \mathrm{N}$ | : | 8-2 | 7 A | 122 |
| NB $\varnothing$ UND | \# | 8-3 | 7 B | 123 |
| NAT | @ | 8-4 | 7 C | 124 |
| NQUU¢TE | , | 8-5 | 7 D | 125 |
| NEQUAL | = | 8-6 | 7 E | 126 |
| NDQU $\chi^{\text {T }}$ | " | 8-7 | 7 F | 127 |

## Appendix 2

This section contains two examples of programs using our free field package. The first illustrates how subprograms using $\operatorname{FREAD}$ and FROUT are controlled, using our routines for reading, expanding, and storing a phrase structure grammar (PSGINN) and then writing it out in an expanded form (PSGSMP). The second shows how to construct a program which itself uses FREAD.
$-1$ $\square$

$r$

COMOILFK GPTITNS - NAME = MAIN,OPT=00,LINFCNT=57, SOURCE,EACOIC,NDLIST,DECK,LOAD,NOMAD, NUEUII, IVUID


ISN 0003
I SN 0004 ISN 0005

ISN 0006
$\stackrel{\rightharpoonup}{N} \quad$ ISN nent
ISN nen7
ISN 0003
ISN yon ${ }^{\text {IS }}$
ISN 3019
ISN onil
ISN 0012
ISN OOIT
ISN 0014
ISN 0015
ISN 0016
ISN 0017
ISN 0019
ISN 0019
ISN 002?
, I SUMP, NCHRTR
INTEGER*2 I SUMP, NCHRTR
FBLANK, FLAND, FMIMUS,FSLASH,FCENT FSTOF FLESS FFLEFTP,FPLUS 2 FLOR , FXCLM, FDOLLR,FSTAR, FRITEP,F,COLN,FLNOT ,F1211. FCOMMA, 3 FPERC, FLINE, FGREAT,FQUERY,FCOLON,FBCUND,FAT ,FQUOTE,FEQUAL, 4 FOQUOT, PAGE RECORT
REAL*R
FRLANK,FLAND ,FMINUS,FSLASH,FCENT FSTOP FFLESS FLEFTP,FOLUS FLOR ,FXCLM FDOLLR,FSTAR ,FRITFP,FSCOLN,FLNOT ,FI211, FCDMMA,
FDERC ,FLINE FGREAT,FQUERY,FCOLON,FBCUND,FAT ,FQUOTF,FEOUAL
LOGICAL*1 T/.TRUF./.F/.FALSF./
ISTART =
1 SUMP $=1$
CALI INETLZ
CALL PSGENN
cal l dSgsmp
FROAT ISTAPT,F,
CALL PSGSMP
LNGOUT
FNT
***** END OF COMPILATION ******

```
    "AFTEST PS-RULFS OLASOPE OYELARAN.A U G . 2 2:1967."
SS = # S #.
S = NPVP.
VP = (PRE) V (((NP) (PP) (AGNT), SS, AP) (ADV)).
v = AU X(VB(ADV),CUP).
AUX = ((DO, (HAVEEN)(BFING)))AUXA.
AUXA= (MOD)(PRES,P A S T )(ASP).
ADV = (SS, ADVB (SS), PP).
AP = ((PRE) ADJ (SS), SS).
PP = PRE NP.
NP = (NPSS, (D)NN U,SS).
D = (PRE) (ART(ADJ) (SS), (D) ADJ).
PRE = (NEG) (PRT).
ART = (WH)(INDEF,O F F,.
v = LOVE,PROCRASTINATE,TEMPORIZE,ALTERNATE,LOSE.
ADVH = QUICKLY,MAGNIFICENTLY,FANTASTICALLY,SORROWFULLY,SADLY.
CCP = BE.
ADJ = STUPENDOUS,BIG,GIANT,TURQUOISE,CATASTROPHIC.
N = ELEPHANT,HIPPOPCTAMUS,GNU,UNICORN,AARDVARK,RURU,NIGHTHAWK.
INDEF = A.
DFF = THE .
$END
```




```
MOD PRFS ASO,
MID PRES
MODPASTASO,
M!])PAST.
ADV =
SS.
pp,
ADVB SS,
ADVR.
AP =
SS,
PRF ADJ SS,
DRE ADJ,
ADJ SS,
AOJ.
p\rho=
PRFNP.
VP =
S5,
NP SS,
    NNU,
    n N NU.
    D = -
    PRFARTSS,
    PRFA R TADJSS,
    PRE ART ADJ,
    PRF ART,
    ORE ADJ,
    D AOJ,
    ARTSS,
    ART ADJ SS,
    ART ADJ,
    ART,
    ADJ.
    PRE =
    PRT.
    NEG PRT,
    NEG.
    ART =
    WH INDEF,
    WH DEF,
    INOEF,
    DEF.
    v =
    TEMPORIZE,
    PROCRASTINATE,
    LOVE,
    LOSF,
    4LTFRNATE.
    ADVB=
    SORROWFULLY ,
    SADLY,
    QUICKLY
    MAGNIFICENTLY,
    FANTASTICALLY.
    COP =
    HE.
    ADJ =
    TURQUOISE,
    STUPENDOUS,
```




TABLEOF LON G W O R D

K E Y WORD

PROCRAS 1 TEMPORI 2 ALTERNA3 MAGNIFI 4 FANTAST5 SORROWF6 STUPEND7 TURDUOIS CATASTRG HIPPOP 10 NIGHTHII

EXPANSION

PROCRASTINATE TEMPORIZE
A LTERNATE MAGNIFICENTLY FANTASTICALLY SORROWFULLY STUPENDOUS TURQUNI SE CATASTROPHIC HIPPOPOTAMUS NIGHTHAWK

COAPILFR TPTIOMS－NNF＝MAIN，JPT＝JO，LINESNT＝57，SOURCE，FREDIC，NOLIST，NJDECK，LDAD，MAP，NOEOIT，NJID

ISN 0002
I SN 0003
ISN 0004
ISN 0005
ISN 0036
I SN 0007
ISN 0009
ISN 0009

ISN 0010

ISN0011
IMPLICIT INTEGER＊2（A－Z）
COMMON MMAENCM／CHRTR，KSUAP，ISUMP，NCHRTP MAINEY
REAL＊R CHRTR，KSJMP 2 IJJJI MAINCM
INTEGER＊？ISUMP，NCHRTR
COMMON／GRDCA／NUM，I SPFFEORDFL ，NUMFI
INTEGER＊2NUM，ISPEC
LOGICAL＊1 ORDFL，NUMFL
COMMON／FCSTCM／
1 FBLANK，FLANT，FMINUS，FSLASH，FCENT，FSTOP，FLESS，FLEFTP，FPLUS
2 FLOR FXCLM，FDOLLR，FSTAR．FRITEP，FSCOLN，FLVJT，FI211，FCOMMA，
3 FPFRC，FLINE，FGREAT，FOUERY，FCCLJN，FROJND，FAT ，FQUOTF，FEQUAL，
4 FDOUחT，PAGE，RECORO
RE AL＊！3
1 FBLANK，FLAND，FMINUS，FSLASH，FCENT，FSTOP，FLFSS，FLEFTP，FPLISS
2 FLIR ，FXCLM，FDOLLR，FSTAR，FRITEP，FSCJLN，FLNJY，FI211，FCDMMA，
3 FPERC，FIINF，FGREAT，FQUERY，FCOLDN，FBOJND，FAT ，FQUDTE，FEQUAL，
4 FDQUDT，PAGF，RECDRT
COMMON／CNSTCY／VBLANK，VLAND，NMINUS，NSLASH，NCFNT，VSTOP，VLESS，
1 NLSFTP，NPLIIS，NLOR ，NXCLM，NOOLLR，NSTAR，NRITFP，NSCBLV，VLVIT，
2 N1211，NCOMMA，NPERC，NLINE，NGREAT，NQUFRY，NCDLDN，NADUND，NAT
3 NQ！JOTE，NEQUAL，NDQUOT
INTEGFR＊2 NBLANK，NLAND，NMINUS，NSLASH，NCEVT，VSTOP，VLESS
1 NLEFTP，NOLUS，NLIR ，NXCLM，NDMLLQ，NSTAR，NRITEP，NSCMI N，VLNJT
2 N1211，NCOMMA，NPERC，NLINE，NGKEAT，NQJERY：VGJL JN，NBOUNH，VAT
3 VQUOTF，NEQUAL，NDQUOT
LOGICAL＊1 T／，TRUF．1，F／．FALSE．／
PEAL＊8 FREAC
CALLIYITLZ
ISTART $=0$
ISUMP $=1$
ISUMP $=1$
KSUMPIISUMP）$=$ PAGE
CALLFROUTI ISTART，F，F，F，F，F，FI
$\mathrm{J}=1$
NUM
100CHRTR＝FREAD（DUMMY）
I F（ORDFL）GOTO ern
ISUAP $=I$ SUM，
SSUMDI +1
KSUMPI ISUMP）－CHRTR
IF（ISPEC．NE．NSTOP）GOTOION
$\mathbf{J}=\mathbf{J}+1$
NUM $=1$ J
1 SUMP $=1$ SUMP +1
3 SUMP 1 SUMP）＝RECDRD
0010100
200 CALL FROUT（ISTART，F，T，T，F，F，T）
RETURN
END

MAINCM
MaINC＊
02うこと
DRDCM
ORDCM
FESTCM
FESTCM
FCSTCY
FC STCM
FESTCM
FE STCY
FCSTCM
FCSTCM
FCSTCM
FESTCM
CNSTCY
CNSTCM
CNSTCM
custcm
CNSTCM
CNSTCM
CVSTEM
CNSTCM

ISN 0013
ISN 0014
ISN 0016
ISN 0016
ISN 0018
ISN 0019
I SN 0020
I SN 0021
ISN 0022
I SN GO23
ISN 002 S
ISN 0026
ISN 0027
ISN 0029
ISN 0030
ISN 0031
ISN 0032
ISN 0033
ISN 0034
ISNOO3 S
ISN 0036
1 ANTIDISESTABLISHMENTARIANISM.
2 BIOGRAPHY.
3 CORPULENT
4 DIDACTIC.
5 ELEMENTARY.
6 FLOCCINAUCINIHILIPILIFILIPICATION
7 GRAND • HORRIBLE. INCREDIBLE. JUXTAPOSED.
11 KANGAROO. LAMPOON.MINUET.NEPOTISM.
15 ORANGATANG• PETTY.QUESTIONABLE.RHETORIC.
19 STAMINA.
20 TAMATATAWHAKATANGIHANGIKOAUAUTAMATEA21 UNDULATION. VIBRAPHONE.WISHYWASHY.
24 XANADU. YPKES. LEN.$\$$

ANTIDISESTABLISHMENTARIANISM.
BIOSRAPHY.
COPPULENT.
OIDACTIC.
ELEAENTARY.
FLOCCINAUCINIHILIPILIFILIPICATION.
GRA ND.
HIRRIGLE.
INC REOIBLE.
JUXTAPOSED.
KAN:AQDO.
LAMPOON.
MINJET.
NEPITISM.
ORAVGATANG.
DETTY.
OUESTIONABLE.
SHETOiils.
STA1IVA.
rATATATAWHAKATANGIHANGIKIJAUAUTAMATEA.
UNはノII ATION.
VIGRAPHUNE .
WISHYNASHY.
xAVADU.
YPRES.
ZEN.

