IV. SPECIFICATION INPUT TO THE REVISED SMART SYSTEM

Michael Lesk

"Cheshire-Puss," she began, rather timidly ... "Would you tell me, please, which way I ought to go from here?"
"That depends a good deal on where you want to get to," said the cat.
"I don't much care where --" said Alice.
"Then it doesn't matter which way you go."

Alice in Wonderland, Ch. 6

1. Introduction

The revised SMART system contains a wide variety of processing methods offering great flexibility in retrieval schemes to the prospective user. The selection of the individual schemes to be used in any given run is determined by the programmer through a set of specifications submitted at the beginning of the data cards for a SMART run. These cards are interpreted by the program SETFLX and used to set up a table of common locations; these are used to tell the SMART system which methods are to be used, which weights are to be attached to their results, and what output is to be printed.

In general, three classes of parameters are used. Some parameters are flow-control parameters; these decide which processing methods are used in the given run. Other parameters set weights and cutoffs needed by the processing programs. Finally, another set of parameters controls mechanical functions such as printing.

In Part 2, the individual parameters are presented with their meanings and spellings. In Part 3, the program that reads the parameter cards is explained.

2. Parameter Names Recognized by the SMART System

A. Parameters Which Affect the Flow of Control

Parameter Name	Value	Meaning	Initial Value
STATPR		Statistical phrases are found during the lookup of English-language input	OFF
CLUSPR		Clustered phrases are found during lookup	OFF
SYNTAX		Syntactic phrases are found during lookup	OFF
CONCON	n	Concept-concept correlation is performed and iterated n times.	OFF
DOCDOC		Document-document correlation is performed and used to expand answers to requests	OFF
HIER	v	Hierarchical expansion is performed. Value is either EXPAND or SHRINK, indicating respectively expansion or replacement of vectors being changed.	OFF
SCORES		An evaluation is performed.	OFF
AUTHOR		The author's name is used for retrieval purposes.	Not Yet Implemented (NYI)
JOURNL		The journal in which the document appears is used for retrieval purposes.	NYI
CITES		The citations in the bibliography of the document are used for retrieval purposes.	NYI

B. Parameters which Set Weights, Cutoffs, etc.

STEMWT	w	Weight assigned to concepts derived from word stems via thesaurus lookup.	1.0
STATWT	w	Weight assigned to concepts from statistical phrases.	0.0
SYNWT	w	Weight assigned to concepts from syntactic phrases (criterion trees)	0.0
CLSWT	w	Weight assigned to concepts from clustering procedures.	0.0
BODYWT	w	Weight assigned to concepts coming from body of text.	1.0
TITLWT	w	Weight assigned to concept coming from title of text.	1.0
ROOTWT	w	In a hierarchy expansion, weight assigned to parent nodes (stems).	0.0
BRANWT	w	In an hierarchy expansion, weight assigned to brother nodes.	0.0
LEAFWT	w	In a hierarchy expansion, weight assigned to sons of nodes.	0.0
CROSWT	w	In a hierarchy expansion, weight assigned to cross-references.	0.0
EXPAND	v	Specifies expansions to take place; controls CONCON, HIER expansion methods. Possible values are REQS (expand requests only), DOCS (expand documents only), ALL (expand everything).	REQS
LOGVEC		Weight all concepts with a weight of 1, regardless of source.	OFF
CUTRD	c	Cutoff for request-document correlation.	0.35
CUTDD	c	Cutoff for document-document correlation.	0.50

CUTCC	c	Cutoff for concept-concept correlation, first iteration.	0.60
CUTC2	c	Cutoff for concept-concept correlation, second and later iterations.	0.60
MODERD	v	Mode of request-document correlation. May be COS or OVLAP.	COS
MODEDD	v	Mode of document-document correlation.	COS
MODECC	v	Mode of concept-concept correlation, first iteration.	COS
MODEC2	v	Mode of concept-concept correlation, after first iteration.	COS
DOCTAP		Read documents from a document collection tape which has been mounted on A6.	OFF
AUTHWT	w	Weight assigned to concepts from author's name.	NYI
JOURWT	w	Weight assigned to journal.	NYI
CITEWT	w	Weight assigned to citation.	NYI

C. Parameters Controlling Mechanical Functions

A1	n	Assign the SMART logical tape function A1 to FORTRAN logical tape number n, corresponding to a physical tape unit as determined by (IOU). A1 is the FORTRAN monitor system tape.	1
A2	n	Same for A2 - system input tape.	5
A3	n	Same for A3 - system print tape	6
B4	n	Same for B4 - system punch tape.	7
A4,...,A8 B1,...,B3,B6		SMART scratch tapes (For example, to change the punch tape to B6, the specification B4 = 16 would be used. This indicates that SMART tape B4 is to be assigned to FORTRAN logical tape 16 which with the Harvard IOU table corresponds to physical real machine tape drive B6.)	Assigned as in FORTRAN (IOU) table

PAGE	n	Set initial page number to n (of doubtful utility).	1
ANSWER	v	Print answers in format as determined by v, which may be SHORT, MEDIUM, or LONG.	OFF
ENGTXT		Print English texts of documents read.	OFF
NOTFND		Print words not found in dictionary.	OFF
PUNCH		Punch document concept-vectors.	
NODECO		Print node correspondences for syntactic phrase matches.	OFF
SYNANA		Print syntactic analysis for each sentence analyzed.	OFF
PREQCO		Print request-document correlations	NYI
PDOCCO		Print document-document correlations.	NYI
PCOCOR		Print concept-concept correlations.	NYI
PCORDD		Print concordance.	NYI
THES	n	Thesaurus used is version n.	1
RANNUM	n	Initial random-number is n.	time
MAXCON	n	Largest concept number is n.	32000
MAXTIM	t	Maximum job time is t minutes.	infinity
STOP		End of parameter list.	
X		End of parameter list.	

3. Description of SETFLX

The parameter list for SETFLX is punched on any number of cards and in any place on these cards. Names of parameters must be punched without imbedded blanks and separated from other parameter names by

blanks. A numerical or logical value for a parameter must follow the parameter, separated by a blank. Floating point numeric values must be punched with a decimal point. The parameter scan will stop if a STOP or X specification is detected, or before a card with a * in column 1. The following is a sample parameter list:

```
ENGTXT  NOTFIND  STATWT  1.5  PRNVEC    STATPR
MODERD  COS  SCORES  X
```

This looks up English texts in the dictionary, prints the text and the words not found, finds the statistical phrases and weights them 1.5, and then evaluates the run using cosine correlation for the request-document correlation.

The parameter list is read by a two-program system, CALSET and SETFLX. CALSET initializes all parameter values and sets up a table of parameter names. SETFLX reads and interprets the parameter cards, using the table of CALSET. This table consists of two word items. The first word of each item is a six-character BCD parameter name. The second word is interpreted as follows:

1. If positive, its decrement is a subscript indicating the common location affected by the parameter. If the address is nonzero, this location is set to nonzero; if the address is zero, the programmer is expected to provide a number to be read into this location.

2. If negative, the program expects a logical value following the parameter. The decrement of the second table word gives the number of possible values of the logical parameter, and the actual spelling of the values follow in successive two-word items.

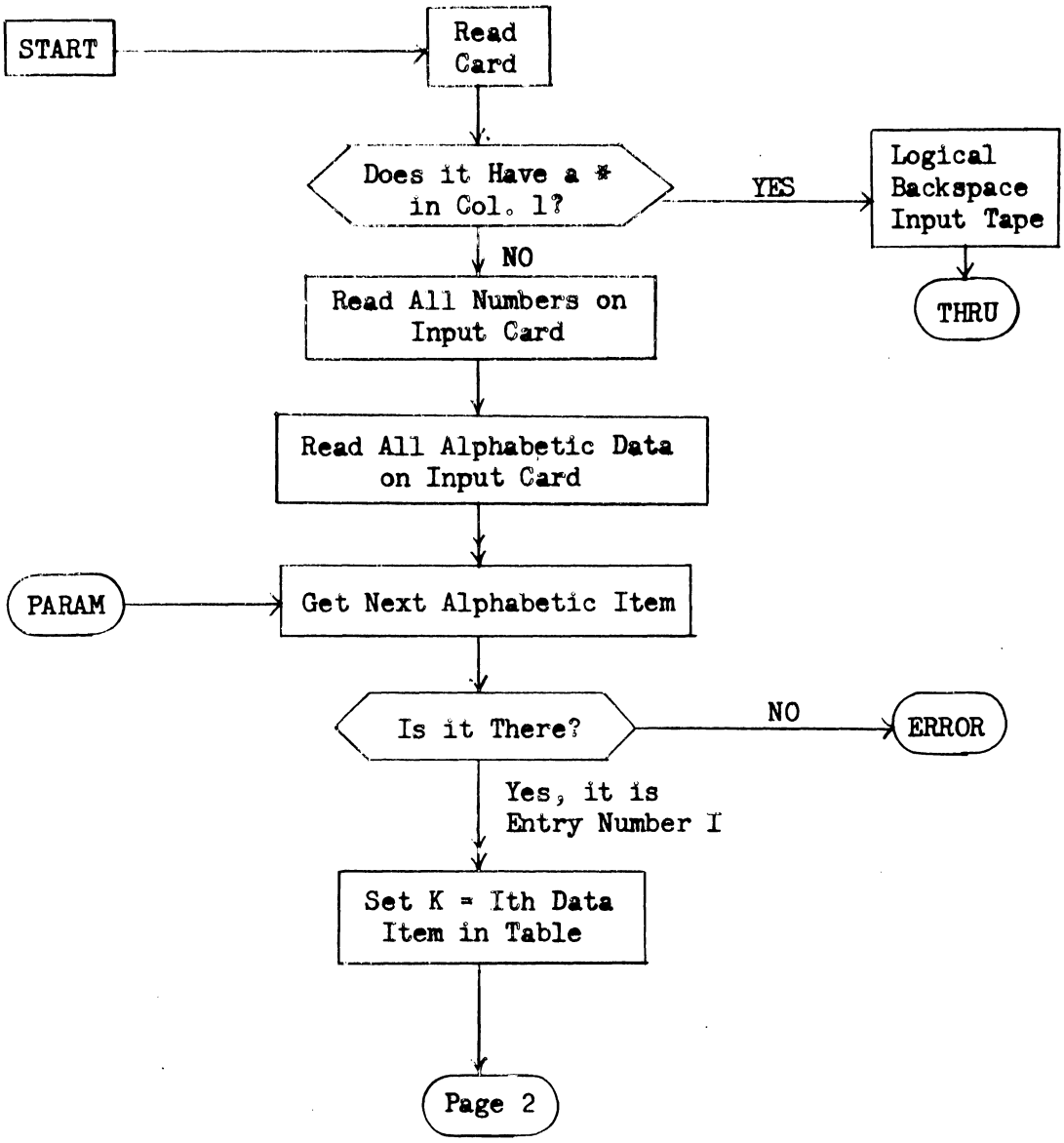
3. If zero, this indicates the end of the parameter list.

The actual card scans are performed by Smithsonian Astrophysical Observatory subroutines (TSH), REREAD, and (IOH).

CALSET contains a short program to copy the initial value table into the proper locations and then call SETFLX. Most of CALSET, however, is devoted to the tables needed by SETFLX. The macro-operations MURPHY, DSAPIO, TWEED, and CROKER are used to make up these tables. DSAPIO PARAM defines a table entry for an "on-off" parameter named PARAM. MURPHY PARAM, INIT,(F) defines a parameter PARAM with initial value INIT (if the F is given, in floating point), with a value allowed to be supplied by the programmer. TWEED PARAM,(VAL1,VAL2,...,VALn),VALk defines a parameter PARAM with possible logical values VAL1,... of which VALk is the initial value. CROKER INIT,(F) defines an initial value item INIT (if the F is supplied, in floating point). A flowchart is appended to detail the SETFLX operations.

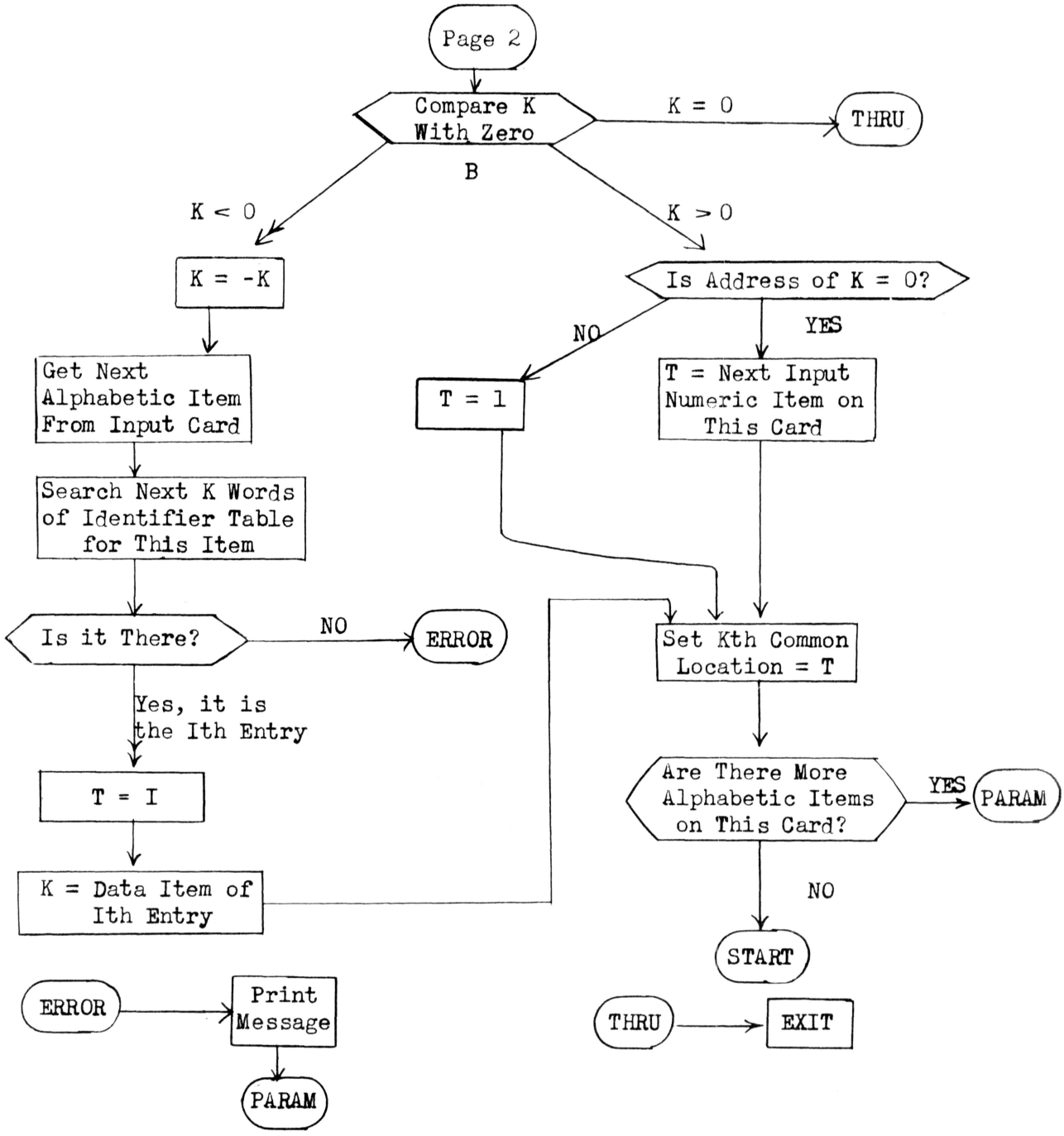
APPENDIX A

SETFLX



Program for SETFLX

Flowchart 1



Flowchart 1 (continued)