The University of Wisconsin
Computer Sciences Department
1210 West Dayton Street
Madison, Wisconsin 53706

BIBCON

An 1108 Program For Producing Concordances To
Prose, Poetry and Bibliographic References

by

Richard L. Venezky

Computer Sciences Technical Report #113

revised

October 1971

CONTENTS

INTRODUCTION

BIBCON is a computer program which generates concordances to prose, poetry, bibliographic references, and other strings of symbols.[1] It runs on the 1108 and is especially adapted for use by non-programmers. The program description which follows is divided into two sections; the first, in a relaxed, semi-personal style, is for the computational novitiate, the person who has had little or no contact with computers, computerese, and computniks, and anticipates blessings without complete baptism. The second section, in cold, passive, multi-adjectival style, is a terse technical description of the program and its use. For the reader who experiences difficulty in choosing between the two, a short quiz is included on the next page, complete with instructions for scoring and routing.

---

[1] BIBCON is based loosely on Control Data Corporation's KWIC. The 1108 version was developed by Mr. William Holman of the Madison Academic Computing Center, University of Wisconsin. A new version of this system, containing basically the same features as described here, but with a highly simplified control card structure, has been designed and will be implemented soon.

QUIZ

Instructions: Read this passage and then answer the questions below.

For a quantitative study of psychoanatomical scatology (sponsored by an Air Force grant), Last Exit to Brooklyn has been punched onto cards in a fixed field format with each 80-character record containing a phrase plus various phrase descriptors. The cards have been blocked (10 to 1 blocking factor) and written onto magnetic tape at low density with binary parity. A concordance to the phrases is to be made, using the first 100 words in the Thorndike-Lorge list as stopwords.

1. How large (in characters) is each physical block on tape?

2. How large (in characters) is each logical record on tape?

3. Is tape parity odd or even?

4. Is the tape density 200 bpi, or 556 bpi? What is bpi?

5. Who wrote Last Exist to Brooklyn?

6. If "the" occurs in the encoded phrases, will it occur in the concordance?

GRADING INSTRUCTIONS

1. If you gave up before the end of the quiz, skip the answers and start reading on page 4.

2. If you finished, compare your answers to those shown below.
    With at least four correct answers, you can probably skip
    section 1 and wade through section 2. Otherwise, start on the
    next page.

ANSWERS

1. 800

2. 80

3. odd

4. 200

5. Hubert Selby, Jr.

6. no

SECTION I

Guide For The Perplexed

<u>What it does</u>

1.1 BIBCON allows you to take materials like those in figure 1 and produce materials like those shown in figure 2.

<u>insert figures 1 and 2 here</u>

The materials in figure 1 will hereafter be called <u>input</u> and those in figure 2, <u>output</u>. To use BIBCON, you must prepare input according to certain conventions, to be described shortly, and specify, <u>inter alia</u>, how you want the output to appear (hereafter called the <u>output format</u>). Basic to the operation of this program is the selection (by the program) of <u>words</u> (to be defined shortly) from the input, along with the contexts in which they occur and the context indentifiers (e.g., the page and line numbers). These <u>words</u>, with their accompanying materials (contexts, identifiers) are then sorted in alphabetical order and printed. This, by definition, is a concordance.

1.2 A <u>word</u>, as BIBCON sees it, is a string of characters--any characters--that occurs between spaces (called <u>blanks</u>). Thus, in the sequence

TO BE OR NOT TO BE,THAT IS THE XYPHCGS-HCS 3.

all of the following are words:

Cap. VI. 1. behaldeþ þ ge eowre soþfestnisse ne doan fore monnum þ ge sie geseanæ from heom from him (*sic*) elles ⊦ elcur ge ne habbaþ lean ⊦ mearde mid eower fæder þæne þe in heofunum is   2. *forþon þonne* þu wirce ælmisse ne blau þu beman for þe swa liceteras doan in heora somnungum ꝺ in tunum þ hie sie weorþade from monnum soþ ic sæcge eow hie onfengun heora lean   3. ðe þonne wircendum ælmesse nyte se winstrae hónd þin hwæt þin sio swiþre dóa   4. þæt þin ælmes sie in degulnisse ꝺ þin fæder se þe gesið in degulnisse geldeþ ðe

Figure 1

Chapter VI, verses 1-4 of the <u>Rushworth Matthew</u>, edited by W.W. Skeat. (<u>The Holy Gospels</u>, Cambridge: University of Cambridge Press, 1871-1887, p. 53).

CL$NSADE
  BLINDE GESEE* HALTE GANGA* HREOFE SINDUN CL$NSADE 7 DEAFE GEHERA*   7   MATT 11 5
  DEADE ARISA* *URFENDE GODSPELL SECGA*

CL$NSIGA*
  WA EOW BOKERES 7 FARISSEAS LICETERAS FOR*ON=23 +E GE CL$NSIGA* (. . MATT 23 25
  UTAN IS C$LCES 7 .... EINNE *ONNE=45 FULLE SINDUN NEDNIMENDE 7
  UN-CLENNISSE

CL$NSIG$*
  UNTRYMNISSE=6 H$LE* DEA+E *$CCE* HREOFE CL$NSIG$* DEOFUL-SOECE'IO, . MATT 10 8
  UT-WEURPA* ARWUNGA GE ONFENGON ARWUNGE GESELLA*

CNEHT
  HENU CNEHT MIN *UNE IC GECEAS SE LEOFA MIN IN *$M WEL GE-LICADE . . MATT 12 18
  SAULE MINE IC SETTE GAST MINNE OFER HINE 7 HE DOEME+ *EODUM S$GE*

  7 +REATADE HINE SE H$LEND=456 7 EODE FRCM=4 HIM (DEOFUL 7 GEH$LED W$* MATT 17 18
  SE CNEHT

  FOR*ON SWA HWA EADMEDA* HINE SWA CNEHT *IOS *E IS MARE IN RICE . . MATT 18 4
  HEGFUNAS

  7 INGANGENDE (HUS GEMCETTUN *UNE CNEHT MID MARIA MODER HIS 7 . MATT 2 11
  FOR*FALLENDE GEBEDUN TO HIM 7 ONTYNDEN HEORA GOLD-HORD BROHTUN HIM LAC
  GOLD RECILS 7 MURRA (IS SMERENNIS

  *A HIE WERON GEWITENS HENU ENGEL DRIHTNES $TEAWDE IN SWEFNE IOSEP . MATT 2 13
  CW=*ENDE ARIS 7 GENIM *UNE CNEHT 7 HIS MODER 7 FLEUH IN $GYPTI 7 W$S *$R
  O**$T IC S$CGE +E FOR*CN=23 +E TOWARD IS SO*LICE (TE HERODES SOECA*
  *CNE CNEHT TO OFSL$ANNE HINE

Figure 2

Part of the concordance to the Rushworth Matthew.

TO	TO	XYPHCGS-HCS
BE	BE,THAT	3
OR	IS	
NOT	THE	

Notice that by not leaving a space after the comma which follows the second be, be,that becomes a word. The number 3 is followed by punctuation plus a space. Unless instructed otherwise, the program will ignore terminal punctuation. Thus, the following are all the same word (∧ means blank):

∧JOE∧ ∧JOE: ∧

∧JOE,∧   JOE-∧

∧ JOE.∧

In the alphabetical listing, text words longer than 30 characters will have only the first 30 characters listed.

1.3 This is how the program defines words, but not all words need be included in the alphabetical list in the output.[1] You can specify that certain words (called stop words) not be included, or that only certain

---

[1] The program and the computer will be used interchangeably and anthropomorphically. While technically certain functions and restrictions are attributable to the program (BIBCON) and others to the computer, there is no need for the non-technical user to burden himself with such distinctions.

words be included (all included words are called <u>key words</u>). Further-
more, you can either list your own stop words, or specify that the
standard stop word list in Appendix A be employed. These restrictions
apply only to words in the alphabetized list. The contexts for all
words which are included in this list will be complete. The prepara-
tion of stop word and key word lists is described in section 4.7. All stop
words used for a concordance with their frequencies of occurrence are
listed before the concordance is printed.

Input Preparation

2.1 Input is divided into units called <u>records</u>, each record being a
line of poetry, or a bibliographic citation, or something similar.
Records, in turn, are divided into either two or three components
(this is your option)--<u>Id</u> (identifier) and <u>Title</u>, if only two components,
and these plus <u>Notes</u>, if three. Each record is punched into cards,
with the contents of each component (hereafter called <u>fields</u>) starting
in the same relative position in each record. Furthermore, all records
and all fields of all records must be the same length, and no record
can be longer than 400 characters (five punched cards)[1]. What this
means is that each field size is selected to accommodate the largest

---

[1] A character is equivalent to a position (column) on a card, no matter
what it contains (letter, number, blank, punctuation, special symbol).
A MAN, BILL, CAME. for example, contains 18 characters,
counting from the first A to the final period.

element (character string) that will be placed in that field. Titles

and Notes that are smaller than this maximum length are positioned

at the beginning of the field and are followed by enough blanks to

fill out the field. To set up a record for input is to designate the start-

ing and ending character positions for each field. The fields do not

have to be in any particular order, nor must they be contiguous.

(The form for specifying these bounds for BIBCON is explained in

section 4.3.) An example is shown in figure 3. The record described

here will later be called a logical record (in contrast to a physical

block).

<hr>

insert figure 3 here

Words from the Title field only are included in the alphabetized

list. However, the Id and, at the user's option, the Notes are printed

with each key word derived from the record in which they occur.

2.2   Since input is generally prepared on punched cards, you can

use only those symbols found on the standard keypunch. These

symbols, in the sequence in which BIBCON alphabetizes them, are

shown in Appendix B. The only restriction on symbols is that the

$\frac{7}{8}$ punch ($\boxminus$) must not appear in col. I of any card except a

Computing Center control card. Words are alphabetized by BIBCON by

first placing each in a fixed length field (you may specify the length up

3600 COMPASS SYSTEM CODING FORM

PROGRAM

ROUTINE

THE UNIVERSITY OF WISCONSIN
COMPUTER SCIENCES DEPARTMENT
311 NORTH PARK STREET
MADISON, WISCONSIN 53706

NAME	RLV
PAGE	12
DATE	10 July 1968

LOCN	OPERATION,MODIFIERS	ADDRESS FIELD	COMMENTS	IDENT
ATT	1.	BEHALDE* GE EOWRE SØ* FESTINISSE NE DØAN FØRE MØNNUM (GE SIE GESEANS		
FRØM=4	HEØM=4	FRØM=4 HIM ELLES Q ELCUR GE NE HABBA* LEAN Q MEARDE MID EØWER FØD		
R *ØNE	*E	IN HEØFUNUM=8 IS		
ATT	2.	FØR*ØN=23 *ØNNE=45 *U WIRCE SLMISSE NE BLAU *U BEMAN FØR *E SWA LICET		
RAS DØAN	IN HEØRA	SAMNUNGUM=2 7 IN TUNUM (HIE SIE WEØR*ADE FRØM=4 MØNNUM SØ* I		
SØCGE EØW	HIE ØNFENGUN HEØRA LEAN			

Figure 3

Coding for <u>Rushworth Matthew</u> (see Figure 1, section 1.1).

to 30 characters--see section 4.4), cutting off excess characters on the right if too many characters exist, or filling blanks on the right if too few characters exist. (The blank has a lower value than $\underline{A}$.) Then the words are alphabetized.

BIBCON will handle uppercase characters only (letters, numbers, special symbols) and these must be in a linear sequence. Therefore, if capitalization or superscripts or subscripts are to be marked, it must be accomplished through additional characters. If these characters are placed anywhere except at the end of a word, they will have a pronounced effect upon the alphabetization of the word in which they occur. An encoding scheme for Old English texts, including techniques for marking capitalization, restorations, expansions, and superscripts, is described in Appendix C.

Some Other Input Features

Associated Forms

3.1   BIBCON will, if you desire, not only detect stopwords or keywords, but will also detect certain suffixed forms of these words (called associated forms). The suffixes used are:

e	or	al	ing	ment
y	ied	er	ity	ions
s	ily	ie	ible	als
ed	ion	ly	ness	ations
es	able	est	action	ances
is	iest	ier	ance	ings
				ments

If associated forms are requested (request procedure is described in section 4.6), and no match is found for an input word in the stopword (or keyword) list, the stop words which (a) start with the same letter as the input word and (b) are marked with an asterisk are compared to the first part of the input word, comparing as many letters as are contained in the stopword. If a match is made, the remainder of the input word is compared to the suffix list. If a match is made here, a stop word or key word has been found, depending upon the specification chosen for 'keys' on the General Information Card (see section 4.2).

You can also supply your own suffixes, rather than using the ones shown above (see section 4.5).

Normal Forms

3.2   Once again by option, normal forms of up to 300 different words can be designated. Each keyword found by BIBCON will be compared with this list; if found, the normal form, not the input form, will be included in the alphabetized output list. For example, if debt were

specified as the normal form for <u>dette</u>, an output listing might appear

as:

DEBT

WE INCUR'D A LARGE DETTE   A6   21

<u>Keyword List</u>

3.3   After the last entry of the concordance, a list of all keywords

and their frequencies of occurrence will be printed.[1]

<u>Specification of Input</u>

4.1   The complete input to BIBCON includes the following:

1.   Computing Center control cards (see Appendix E).

2.   BIBCON control cards

3.   Input data (records containing prose, or poetry, or whatever).

4.   An EOF card (see Appendix E).

5.   Another Computing Center Control card, called the FIN card (see

     Appendix E).

Input data preparation was discussed in the preceding sections.

The computing center control cards are discussed in Appendix E;

this section deals with the BIBCON control cards. (The control cards

for the input in figure 3, section 2.1 are shown at the end of this

section.) For all entries on all BIBCON control cards, data are to be

punched as far to the right as the field will allow (unless specified otherwise).

---

[1]You can obtain this list on a magnetic tape, also. To do so, see
section 7.5 and Appendix D.

4.2   General Information Card

The first card specifies the general options desired. All
options on this card are specified by punching the appropriate
character in a specified column. The columns, their titles and
contents are described below.

6	12	18	24	30	36	42	48	49-80
keys	stops	norms	checkpt	datesort	restart	colseq	omit	bypass characters

Title	Column	Contents
keys	6	If you want the concordance list to contain only words which you specify, punch a K here. Otherwise, leave this column blank. If you punch a K here, then the words you want in the list must be specified as stop words (see section 4.7).
stops	12	If you want to use the standard list of stop words plus their associated forms (see Appendix A and Section 3.1), punch an S here. Otherwise, leave this column blank.
norms	18	If you are specifying normal forms, punch an N here (specification of the forms themselves is discussed in section 4.8). Otherwise, leave this column blank.

Title	Column	Contents

checkpt   24   Ignore this column for now (see section 8.2)

datesort   28-30   If two or more keywords are identical, characters
at the beginning of each logical record will be
compared to decide the ordering of the entries.
The number of characters compared will be equal
to 6 times the number of computer words required
to hold the keyword (at 6 characters per computer
word). Thus, if keyword length (maxkey on the Out-
put Format Card--Section 4.4) is 15 characters, 18
characters of the first field of each logical record
will be compared (the minimum number of whole
words to hold 15 characters is 3; 3 times 6 is 18).

Datesort allows dates of the form year month
day to be sorted by date (rather than alphabetically),
assuming that the field in which they occur is sorted.
(Otherwise, datesort is meaningless.)

The number punched in datesort is the first
character location for the month.

If the first three characters starting from the
datesort position are in the list below, then they are
replaced by an integer string, according to the ordinal
position of the month in the calendar year.

Title	Column	Contents

JAN   APR   JUL   OCT

FEB   MAY   AUG   NOV

MAR   JUN   SEP   DEC

After sorting, the integers are replaced by the letter strings they replaced.

Remember that if the month starts in column 60 of the third card of a record, its character position is 220, not 60.

restart   36 — Ignore this for now. (It is described in Section 8.2.)

colseq   42 — If you want keywords alphabetized in the sequence shown in Appendix B, leave this column blank. If you want the Old English collating sequence shown in Appendix C, punch a 1 here. And if you want to arrange your own sequence, punch a 2 here and punch on a following card the sequence you want, starting with the highest order character in column 1. To avoid ambiguities, all 64 characters should be included. (See Appendix B for punching instructions.) In general, blank should be the highest order character so that forms like BOY will precede forms like BOYS, BOYISH, etc. Under no

Title	Column	Contents
		circumstance should a $\frac{7}{8}$ punch (∅) appear in column 1 of any card other than Computing Center control cards.
omit	48	Punch a 1 here if you want words which start with selected characters to be omitted. (If you are supplying keywords--k in column 6 of the General Information Card--the omit option becomes a select option, designating words to include in the concordance.)
bypass characters	49-80	Characters for the omit (or select) option are punched here, beginning in column 49 and continuing without intervening blanks.

4.3   Input Format Card

	6	12	18	24	30	36	42	48	54	60	66
logical length	phy- sical length	id1	id2	title1	title2	notes1	notes2	average	unit	parity	
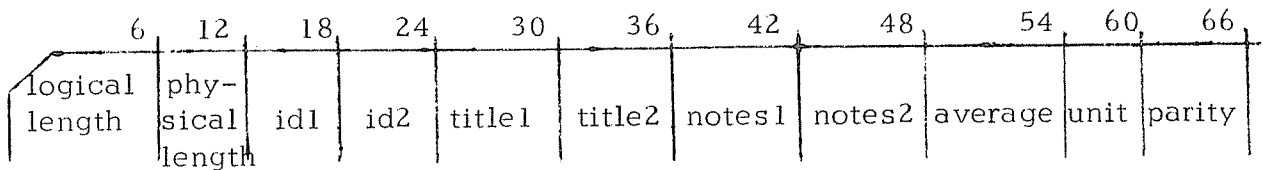
Title	Column	Contents
logical length	4-6	The length in characters of each logical record (see section 2.1). This must be 400 or less.

Title	Column	Contents
physical length	10-12	The length, in characters of each physical block. The record described in section 2.1 is a logical record--that is, a unit consisting of an Id, Title, and, optionally, Notes. The number punched in columns 4-6 gives the length of each logical record. A punched card, or a single card image on magnetic tape is called a physical block. If data input is from cards or card images on magnetic tape, each physical block will be 80 characters long. If records are blocked on tape, the length of a physical block will be 80 times the number of cards placed in each block.[1] If card images are blocked, logical record length must either be a multiple of 80, or the unused characters at the end of the last card of each record deleted. In this case, physical block length will be logical record length times the number of logical records per physical block.

---

[1] A physical tape block is a sequence of data, succeeded and followed by a space of a particular length (called a record gap). A card image on tape will have a record gap, then the card image (80 characters), and then another record gap. Two cards in one physical block will be: record gap, two cards (160 characters), record gap. The maximum allowable length of blocked records for BIBCON is 10,000 characters.

Title	Column	Contents
id1	16-18	The first (id1) and the last (id2) character positions of the Id field. (For figure 3,
id2	22-24	section 2.1, id1 is 1 and id2 is 9). The id field must be at least two characters long.[1]
title1	28-30	The first (title1) and last (title2) character positions of the Title field. (For figure 3,
title2	34-36	section 2.1, title1 is 11 and title2 is 160).
notes1	40-42	The first (notes1) and last (notes2) character positions of the Notes field. If Notes are not
notes2	46-48	used, notes1 and notes2 are left blank. (For figure 3, section 2.1, no Notes are used.) (Id, Title, and Notes positions are positions within the Logical record, numbered from 1 through whatever is punched in the length field.)
average	49-54	Average number of keywords in a title (if you don't want to figure it, punch 6 here).
unit	59-60	This is the input unit number. If cards are used, leave this blank (BIBCON will assign it to the appropriate unit). If card images

---

[1] As a general policy, the id field should be the first field in a record.

Title	Column	Contents

on tape are used, punch 10 here, and add the ASG Computing Center control card shown in Appendix E.

| parity | 66 | If cards are used, leave this blank (BIBCON will assign the appropriate parity).[1] If tape input is used, this must be 1 for odd (binary) parity and blank for even (BCD) parity. Which parity to use will depend on the parity selected for transferring cards to tape. (If you have not transferred the cards to tape yet, request odd parity when you do and punch 1 here.) |

4.4   Output Format Card

	6	12	18	24	30	36	42	48	54	60
lines	width	maxkey	spaces	segs	list	notes	unit	print	dots	

Title	Column	Contents
lines	1-6	Number of lines allowed on each output page, (Unless you have reasons for doing otherwise, punch 56 here.)

[1] Parity refers to the error-checking procedure used in reading or writing with magnetic tape. By historical oppression there are two techniques--even and odd--although the earth would revolve more smoothly with only one.

Title	Column	Contents
width	10-12	Number of characters across width of output page. (Unless you have reasons for doing otherwise, punch 131 here.)[1]
maxkey	17-18	Maximum number of characters in each keyword listing (this can be from 1 to 30). Keywords greater than the number punched here will have the rightmost characters chopped off. (18 is a safe number to punch here.)
space	23-24	Number of blank lines between successive entries for a keyword. A blank implies zero.
segs	25-30	The number in this field indicates the number of blank lines which will appear between different alphabetical sections on the output listing. A number larger than lines (cols. 1-6) forces each alphabetic section (A,B,C, etc.) to start at the top of a page.
list	36	Punch a 1 if you want stop words listed before the concordance is printed. Leave blank otherwise. (Even if this option is selected, the stop words and their frequences

---

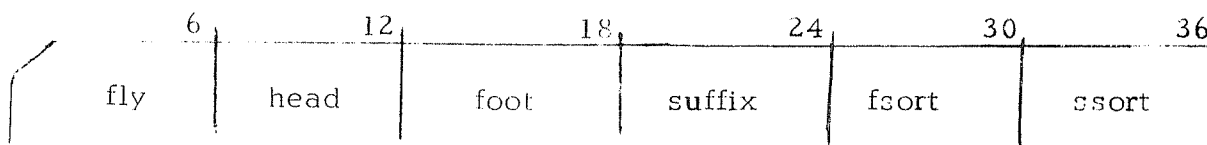[1] If not 131, then it must be greater than the number in cols. 17-18, plus 3, plus id2 minus id1; but not greater than 131.

Title	Column	Contents
		of occurrence will be listed before the concordance thus giving potentially two separate lists of stop words.)
notes	42	Punch a 1 if you want the Notes printed with each concordance entry. If there are no Notes, or if you don't want them listed, leave blank.
unit	47-48	Unit number for writing output file. If you want a single printing of the output, leave this blank. If you are supplying your own magnetic tape for the output, punch a number between 8 and 27 here and add the Computing Center control cards for tapes shown in Appendix E. (By supplying your own tape, you can re-list the output at a significantly lower price than you can by re-running the program.)
print	54	If blank, the keyword will be overprinted to make it darker.[1] If this is not desired, punch a 1 to suppress the overprinting.

---

[1] The overprinting increases by a small amount the cost of running the concordance.

Title	Column	Contents
dots	60	A character in this column requests that whatever character is punched here be repeated from the end of the Title to the beginning of the Id (see figure 2, section 1.1). A blank indicates that this feature is not desired.

4.5 Titles Card

BIBCON will print, if you desire, a fly leaf, a headline (repeated at the top of each concordance page), and a footline (repeated at the bottom of each concordance page). These are called titles. The numbers punched in this card indicate the number of 80-character cards you will supply for each title. These cards must then be placed after the title card, in the order: fly leaf, headline, footline (omitting those not desired). If none of these titles is desired, card 4 must still be included.

	6	12	18	24	30	36
	fly	head	foot	suffix	fsort	ssort

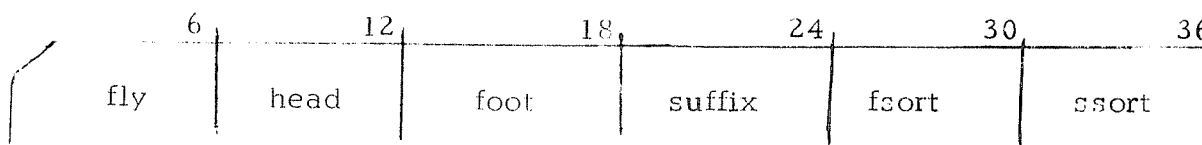| Title	Column	Contents
fly	6	Punch the number of cards included for each entity (0,1,2,etc.) in the appropriate column. Fly should be no

Title	Column	Contents
		of occurrence will be listed before the concordance thus giving potentially two separate lists of stop words.)
notes	42	Punch a 1 if you want the Notes printed with each concordance entry. If there are no Notes, or if you don't want them listed, leave blank.
unit	47-48	Unit number for writing output file. If you want a single printing of the output, leave this blank. If you are supplying your own magnetic tape for the output, punch a number between 8 and 27 here and add the Computing Center control cards for tapes shown in Appendix E. (By supplying your own tape, you can re-list the output at a significantly lower price than you can by re-running the program.)
print	54	If blank, the keyword will be overprinted to make it darker.[1] If this is not desired, punch a 1 to suppress the overprinting.

---

[1] The overprinting increases by a small amount the cost of running the concordance.

Title	Column	Contents
dots	60	A character in this column requests that whatever character is punched here be repeated from the end of the Title to the beginning of the Id (see figure 2, section 1.1). A blank indicates that this feature is not desired.

4.5   Titles Card

BIBCON will print, if you desire, a fly leaf, a headline (repeated at the top of each concordance page), and a footline (repeated at the bottom of each concordance page). These are called titles. The numbers punched in this card indicate the number of 80-character cards you will supply for each title. These cards must then be placed after the title card, in the order: fly leaf, headline, footline (omitting those not desired). If none of these titles is desired, card 4 must still be included.

	6	12	18	24	30	36
	fly	head	foot	suffix	fsort	ssort

Title	Column	Contents
fly	6	Punch the number of cards included for each entity (0,1,2,etc.) in the appropriate column. Fly should be no

Title	Column	Contents
head	12	greater than lines on Card 3. The sum
foot	18	of head and foot should be 10 or less. If either of these limits is exceeded, an esoteric error message will be printed and the concordance not run.
suffix	24	A 1 here indicates that you will supply your own suffixes rather than using those listed in sections 3.1 and 7.3. For this option to have any effect, however, you must also select either the associative forms feature on the Stopword Control Card or the standard stopwords feature on the General Information Card. The suffixes themselves are included after the stopword cards (see section 4.7a).
fsort	30	To have the keyword list which occurs at the end of the concordance listing sorted by frequency punch a 1 here.
ssort	36	To have the keywords sorted from back to front (i.e., going is sorted as if it were spelled gniog), punch a 1 here.

4.6 Stopword Control Card

6	12	18	24,25	48,49	54,55	78
alphas	assoc	sings	endstrip	ends	fstrip	fronts

Title	Column	Contents
alphas	6	A one in this column indicates that any word that starts with a non-alphabetic character is to be treated as a stop word (or keyword if the submitted word list is to be treated as keywords). This feature takes effect after front stripping, and applies only if one or more stopwords are specified.
assoc	12	If the associated forms feature (see section 3.1) is desired, punch a one here. Otherwise, leave blank. (If you specified standard stop words by punching an "S" in column 12 of the general information card (section 4.2) forms will be checked regard-less of what you punch here.)
sings	18	If you want all one-character words to be treated as stop words, leave this column blank; otherwise punch a one here. (This feature also requires at least one stopword to be specified on the Stopword Specification Card.)

Title	Column	Contents
endstrip	19-24	If you want characters other than 0-9 and A-Z to be stripped from the end of key words, leave these columns blank. If you want to specify characters for this process, punch the total number of these characters (24 is the maximum allowable) in these columns (if this number is less than ten, punch the single digit in col. 24).
ends	25-48	Then, punch the characters themselves in consecutive columns, beginning in col. 25. If you want no end stripping of any kind, punch a one in col. 24 and leave columns 25 through 48 blank.
fstrip	49-54	This option is similar to the endstripping just described, except that it applies to the
fronts	55-78	beginning of each word. fstrip specifies the number of characters listed as fronts (max = 24). If fstrip is 0 or blank, however, no characters are stripped.

This card must be included even if it contains all zeros or blanks.

4.7 Stop Word Specification Cards

If the General Information Card (Section 4.2) has a K in col. 6, then the words that are to appear in the concordance list must be punched as stop words (they're not really stop words, but rather key words in this situation; just ignore the devious nomenclature).

If the General Information Card has an 'S' in column 12, indicating standard stop words, do not include any stop word specification cards. If this card does not have an 'S' in column 12, you must include at least one stop word card. (If no stop words are to be included on the stop word card, punch six asterisks in columns 1-6, in place of stop words.)

Stop words are punched with at least one blank between successive words, and must be alphabetized by first letter (deeper alphabetizing is not necessary), proceeding from A-Z regardless of which collating sequence (section 4.2) is used. Stop words beginning with non-alphabetic characters must follow the alphabetic stop words and may appear in any order. Following the blank after the last stop word must be six contiguous asterisks. This signals the end of the stopword list. Neither stopwords nor the terminating asterisks can be broken across cards. The maximum length for stopwords is 30 characte

To indicate stopwords to be checked for associated forms, the character sequence ˏ*ˏ (blank asterisk blank) must follow the stop word and be on the same card as the stopword. Associated forms will not be checked, however, unless the associated forms option is selected on the Stop Word Control Card.

If one stopword forms the first segment of another, as for example. CRACK and CRACKLING, the longest form should precede the shorter one.

Up to 600 stopwords (or keywords) can be specified. If more than 600 are included, the concordance will not be formed and an error message will be printed.

4.7a   Suffix Specification Cards

Up to 50 suffixes may be specified, with each suffix being from one to six characters. (Suffixes longer than six characters will be truncated on the right.) Suffixes are punched in a free field format, using at least one blank between suffixes, and with no suffixes split across cards. At the end of the suffix cards must be a card containing six consecutive asterisks. (For convenience, punch these in cols. 1-6.)

4.8   Normal Forms Cards

If the Normal Form option was specified on the General Information Card (col. 18), then the normal form cards follow immediately after the stop word card (or cards). These cards should be punched the same

way as stop word specification cards, except that on each card, the first word is read as the text form of a word, and the second as the corresponding normal form, and so on, pair for pair. However, the same alphabetizing scheme as described for stop words must be used. To be sure that the words are paired properly, punch just one pair per card--the input (text) form in the first field and the normal form in the second field. Pairs can not be broken across cards. Up to 300 pairs may be specified, but if more are included, the concordance will not be formed.

After the last word pair, a field of at least six asterisks must be punched to signal the end of normalized forms. The maximum length for input (text) or normal words is 30 characters.

If the normalized form option was <u>not</u> selected on the General Information Card, no normalized form cards can be included, not even one containing asterisks only.

4.9 Examples

1. The control cards for the data in figure 3, section 2.1 are shown in figure 4.

_____

<u>insert figure 4 here</u>

2. For the poem shown in figure 5 below, 2 punched cards will be used for each line. These cards will contain the line of poetry plus the line number. Therefore, each logical record will be 160 characters in length and contain an Id and Title field only. The Id field will be punched in columns 1-9, with WW-TCP in columns 1-6 and

```
    S                                            1
  320       80       1      10     12    320                    12
  56       100      16      1       4     1        0             0          .
    1        0        J
    CONCORDANCE  TO  RUSHWORTH  1---MATTHEW
    0        0        0      14='/),123456789
 EOW   GE  HE  HiM  HINE  HIT  IC  IN  IS  ME  MIN  SE  SEO  SWA  TO  W$S  *A  *ONE  *$R  *$T  +E
 *****
```

Figure 4

BIBCON control cards for producing concordance (see figure 2, section 1.1) to materials shown in figure 3, section 2.1.

the line number in col. 9. The Title (line of poetry) will go in

cols. 11-160.

A flyleaf and headline, but no footlines will be specified.

The coded input for this concordance, plus the control cards are

shown in figure 6. Figure 7 shows part of the resulting concordance.

---

insert figures 5, 6, 7 here

Be composed--be at ease with me--I am Walt Whitman, liberal
    and lusty as Nature,
Not till the sun excludes you do I exclude you,

Not till the waters refuse to glisten for you and the leaves to rustle
    for you, do my words refuse to glisten and rustle for you.

My girl I appoint with you an appointment, and I charge you that
    you make preparation to be worthy to meet me,
And I charge you that you be patient and perfect till I come.

Till then I salute you with a significant look that you do not forget
    me.

Figure 5

Walt Whitman, "To a Common Prostitute," in The American
Tradition in Literature 3rd. ed. Ed. by Sculley Bradley, Richmond C.
Beatty and E. Hudson Long. New York: W.W. Norton & Co., 1956,
pp. 115-116.

```
S
160    80      1       9      11     160
56    120     16      0      0      0
        1       1       1       1       1
A CONCORDANCE TO WALT WHITMAN'S 'TO A COMMON PROSTITUTE'
WALT WHITMAN 'TO A COMMON PROSTITUTE'
        1       0       1       1
******
WW-TCP  1 BE COMPOSED -- BE AT EASE WITH ME -- I AM WALT WHITMAN, LIBERAL AND LU
STY AS NATURE,
WW-TCP  2 NOT TILL THE SUN EXCLUDES YOU DO I EXCLUDE YOU,

WW-TCP  3 NOT TILL THE WATERS REFUSE TO GLISTEN FOR YOU AND THE LEAVES TO RUSTLE
FOR YOU, DO MY WORDS REFUSE TO GLISTEN AND RUSTLE FOR YOU.
WW-TCP  4 MY GIRL I APPOINT WITH YOU AN APPOINTMENT, AND I CHARGE YOU THAT YOU M
AKE PREPARATION TO BE WORTHY TO MEET ME,
WW-TCP  5 AND I CHARGE YOU THAT YOU BE PATIENT AND PERFECT TILL I COME.
```

Figure 6

BIBCON deck for Walt Whitman's "To a Common Prostitute."

WALT WHITMAN--TO A COMMON PROSTITUTE

I

BE COMPOSED -- BE AT EASE WITH ME -- I AM WALT WHITMAN, LIBERAL AND LUSTY AS NATURE,	WW-TCP 1
NOT TILL THE SUN EXCLUDES YOU DO I EXCLUDE YOU,	WW-TCP 2
MY GIRL I APPOINT WITH YOU AN APPOINTMENT, AND I CHARGE YOU THAT YOU MAKE PREPARATION TO BE WORTHY TO MEET ME.	WW-TCP 4
MY GIRL I APPOINT WITH YOU AN APPOINTMENT, AND I CHARGE YOU THAT YOU MAKE PREPARATION TO BE WORTHY TO MEET ME.	WW-TCP 4
AND I CHARGE YOU THAT YOU BE PATIENT AND PERFECT TILL I COME.	WW-TCP 5
AND I CHARGE YOU THAT YOU BE PATIENT AND PERFECT TILL I COME.	WW-TCP 5
TILL THEN I SALUTE YOU WITH A SIGNIFICANT LOOK THAT YOU DO NOT FORGET ME.	WW-TCP 6

LEAVES

NOT TILL THE WATERS REFUSE TO GLISTEN FOR YOU AND THE LEAVES TO RUSTLE FOR YOU, DO MY WORDS REFUSE TO GLISTEN AND RUSTLE FOR YOU.	WW-TCP 3

LIBERAL

BE COMPOSED -- BE AT EASE WITH ME -- I AM WALT WHITMAN, LIBERAL AND LUSTY AS NATURE.	WW-TCP 1

LOOK

TILL THEN I SALUTE YOU WITH A SIGNIFICANT LOOK THAT YOU DO NOT FORGET ME.	WW-TCP 6

LUSTY

BE COMPOSED -- BE AT EASE WITH ME -- I AM WALT WHITMAN, LIBERAL AND LUSTY AS NATURE.	WW-TCP 1

Figure 7

Part of the concordance to Walt Whitman's "To a Common Prostitute."

SECTION II

Reference Manual for Programmers

INTRODUCTION

6.1　BIBCON is a 1108 system for producing concordances to prose,
poetry, and other alphanumeric strings. Input is a user-specified,
fixed field format, with logical records of up to 400 characters. Within
each logical record two fields must be designated--title and identifier,
and a third field, notes, is optional. Concordances are formed for
character strings in the title field only. With each concordance entry
the title, identifier, and at the user's option, the notes fields are
listed.

　　The 1108 version of BIBCON includes the following features.

1.　Blocked or unblocked input.

2.　Specification of stop words or keywords, or use of program-
supplied stop words.

3.　Normalization of up to 300 different input words.

4.　Front stripping, end stripping, and testing for suffixed forms.

5.　Separate listing of all keywords and stop words with frequency
of occurrence counts.

6.　Suppression of successive listings of the same keyword.

7.　Variable page length and page width on output.

8.　Flyleaf, headline, and footline options.

6.2 <u>Input</u>

The input to BIBCON consists of the following five items:

1. Computing Center control cards (see Appendix E)

2. BIBCON control cards (see section 8)

3. Data records (see section 7)

4. EOF card (see Appendix E)

5. FIN card (see Appendix E)

6.3 <u>Output</u>

The output from BIBCON consists of the following:

1. List of control instructions

2. Stop words (optional)

3. Suffixes (optional)

4. List of stop words with frequency of occurrence of each.

5. Concordance

6. Totals for:

Titles read in

Stop words used

Key words found

Index entries

Pages

Time used

7. List of key words with frequency of occurrence of each

Data Records

7.1   A logical record for input, which may describe a line of poetry,
or a bibliographic reference, or any other string of characters, con-
tains in fixed positions within the record a title field, an identifier
field, and, optionally, a notes field. The first and last character
positions for each of these fields is specified on the input format card
(see below). The only restriction on symbol selection is that ⌐
(7-8 punch) must not appear in col. 1 of any card except a Computing
Center control card.

Stopwords and Keywords

7.2   Concordances are formed to words in the title field only.[1]

If a word is over 30 characters in length, the word is truncated
on the right to 30 characters for listing as a keyword.

Words not to be included in the alphabetized portion of the con-
cordance (stopwords) or words to be included (keywords) can be specified
by the user. Or the user can specify that the words listed in Appendix
I be used as stop words. The user can also specify that all words
beginning with a non-alphabetic character be treated as stop words.

---

[1] A word is any character string which occurs between blanks.

Associated Forms

7.3   If stop words are used, a BIBCON option allows checking for suffixed forms of stop words. The suffixes checked for are:

e	or	al	ing	ment
y	ied	er	ity	ions
s	ily	ie	ible	als
ed	ion	ly	ness	ations
es	able	est	ation	ances
is	iest	ier	ance	ings
				ments

If associated forms are requested (request procedure is described in section 4.6), and no match is found for an input word in the stopword (or keyword) list, the stop words which (a) start with the same letter as the input word and (b) are marked with an asterisk are compared to the first part of the input word, comparing as many letters as are contained in the stopword. If a match is made, the remainder of the input word is compared to the suffix list. If a match is made here, a stopword or keyword has been found, depending upon the specification chosen for 'keys' on the General Information Card (see section 8.2).

Normal Forms

7.4 At the user's option, up to 300 different keywords can be entered in the alphabetized list of the concordance under a form supplied by the user rather than under the input form.

If, for example, the user specifies that DETTE should be listed under DEBT, then all lines containing either DEBT or DETTE will be listed together.

Keyword List

7.5 Following the concordance and BIBCON-generated job statistics will be a list of all keywords with their frequencies of occurrence in the concordance. This list is output to logical unit 7 during printing of the concordance and later re-read for printing. To save this list, assign unit 7 to a user supplied tape. (Tape format for unit 7 is described in Appendix D.)

Standard Format

8.1 The following cards, in the order shown, compose the control deck. The cards in the left column are mandatory; those in the right column, optional. (Examples of control cards can be found in section I and at the end of Appendix E.)

	Necessary		Optional
1.	Concordance type	1a.	Collating sequence
2.	Input format		
3.	Output format		
4.	Titles	4a.	Flyleaf
		4b.	Headlines
		4c.	Footlines
5.	Stop word usage	5a.	Stop words
			(terminated by a field of six asterisks)
6.	If no stop words are included, a card with six asterisks must be present.	6 .	Normalized Forms
			(terminated by a field of six asterisks)

8.2   General Information Card

(All fields are right justified.)

6	12	18	24	30	36	42	48	49-80
keys	stops	norms	checkpt	datesort	restart	colseq	omit	bypass char-acters

Title	Column		Contents
keys	6	K	key words supplied by user
		S	stop words supplied by user

a blank indicates that stop words are to be provided,
i.e., it is treated as an "S".

Title	Column	Contents
stops	12	S standard stop word list plus associated forms (see Appendix A and Section 7.3) desired; any other character in this column indicates that the user will provide either key words or stop words (depending on "keys" above).
norms	18	indicates whether the user will provide a list of normalized words; N a list of normal words will be provided by the user following the stop word cards.[1] Any other character in this column indicates that no normalized forms are desired by the user.
checkpt	23-24	When this field is zero or blank, BIBCON forms a complete concordance. The processes involve: a. Read input, form records for sorting. b. Do core sort on each batch of records and output to scratch files (tapes or serial drum files). c. Tape sort; yielding sorted file on one file (tape or serial drum file). d. Read in sorted records, format, and print. However, starting or stopping at any phase break can be specified by punching one of the numbers shown below. Number Interpretation 1. Stop before tape sort phase. 2. Stop after tape sort phase. 3. Start at tape sort phase and finish concordance.

---

[1]Since normal forms are checked before stopwords, the normal form list can be used to increase the length of the stopword list. This is done by normalizing additional stopwords to a word listed in the stopword list.

Title	Column	Contents
		4. Do tape sort phase only.
		5. Start after tape sort phase and finish concordance.

For options 1-2, an output message will give tape unit and format.

For options 3-5, the <u>restart</u> field must also be punched (see below).

The numbers 1-5 are punched in column 30. For numbers 0, 1 and 2, a 1 in column 29 will produce a dump of each entry that BIBCON produces for sorting during the input phrase. These entries are listed between the core sort and tape sort phases.

datesort 28-30

If two or more keywords are identical, characters at the beginning of each logical record will be compared to decide the ordering of the entries. The number of characters compared will be equal to 6 times the number of computer words required to hold the keyword (at 6 characters per computer word). Thus, if keyword length (maxkey on the Output Format Card--Section 8.4) is 15 characters, 18 characters of the first field of each logical record will be compared (the minimum number of whole words to hold 15 characters is 3; 3 times 6 is 18).

Datesort allows dates of the form <u>year month day</u> to be sorted by date (rather than alphabetically), assuming that the field in which they occur is sorted. (Otherwise, datesort is meaningless.)

The number punched in datesort is the first character location for the <u>month</u>.

If the first three characters starting from the <u>datesort</u> position are in the list below, then they are replaced by an integer string, according to the ordinal position of the month in the calendar year.[1]

---

[1] These forms will be searched for as shown here, regardless of the collating sequence specified.

Title	Column	Contents
		JAN APR JUL OCT FEB MAY AUG NOV MAR JUN SEP DEC
		After sorting the integers are replaced by the letter strings they replaced. Remember that if the month starts in column 60 of the third card of a record, its character position is 220, not 60.
restart	34-36	If options 3-5 are used for checkpt, restart contains the length of the logical records on the input tape. If this tape was generated by BIBCON, the logical record length will have been included in a message at the end of the job listing (THE LOGICAL RECORDS ARE XXX CHARACTERS LONG). The tape must be assigned as unit 2.
colseq	42	If you want keywords alphabetized in the sequence shown in Appendix A, leave this column blank. If you want the Old English collating sequence shown in Appendix C, punch a 1 here. And if you want to arrange your own sequence, punch a 2 here and punch on a following card the sequence you want, starting with the highest order character in column 1. To avoid ambiguities, all 64 characters should be included (see Appendix B for punching instructions.) In general, blank should be the highest order character so that forms like BOY will precede forms like BOYS, BOYISH, etc. The $\frac{7}{8}$ punch (Ɇ) must not be in col 1 of any card other than computing center control cards.
omit	66	Punch a 1 here if you want words which start with selected characters to be omitted. (If you are supplying keywords --k in column 6 of the General Information Card--the omit option becomes a select option, designating words to include in the concordance.)

Title	Column	Contents
bypass character	49-80	Characters for the <u>omit</u> (or <u>select</u>) option are punched here, beginning in column 49 and continuing without intervening blanks.

8.3 <u>Input Format Card</u>

This card describes the input file. All fields are right justified.

| logical length	physical length	idl	id2	title1	title2	notes1	notes2	average	unit	parity
6	12	18	24	30	36	42	48	54	60	66

Title	Column	Contents
logical length	4-6	number of characters per logical record (maximum is 400).
physical length	10-12	length of physical blocks in characters. (Must not be greater than 10,000; but for efficiency should be less than 2000).
idl	16-18	position of first identifier character in logical record.
id2	22-24	position of last identifier character in logical record (must be greater than idl).
title1	28-30	position of first title field character in logical record.
title2	34-36	position of last title field character in logical record (must be greater than title 1).
notes1	40-42	position of first notes character in logical record (zero indicates no notes).
notes2	46-48	position of last notes character in logical record (zero indicates no notes; otherwise, must be greater than notes1).

Title	Column	Contents
average	52-54	average number of key words per logical record. If zero or blank, a value of 5 is assumed.
unit	59-60	logical unit number of input file (8-27); if blank or zero, 5 is assumed. Logical units 2-4 are used by BIBCON for sorting. Logical unit 7 is used for the list of key-words and frequencies.
parity	66	parity of unit; blank or zero implies even, any other number implies binary.

All values must be right justified within their particular 6-character fields. If the data values do not fulfill the conditions stated above, error diagnostics and program termination result. Identifier, title, and notes fields may overlap.

8.4   Output Format Card

Parameters in the third control card define output. All fields are right justified.

6	12	18	24	30	36	42	48	54	60
lines	width	maxkey	spaces	segs	list	notes	unit	print	dots
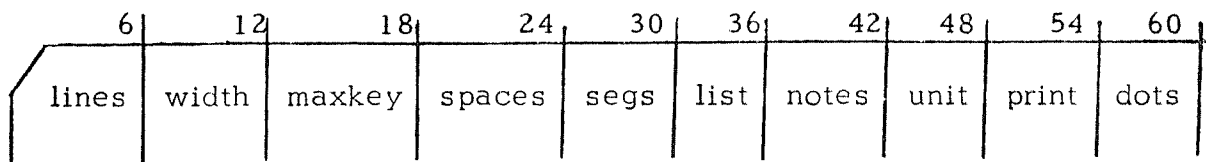
Title	Column	Contents
lines	1-6	number of lines per page.
width	7-12	number of characters across width of page (maximum is 131).

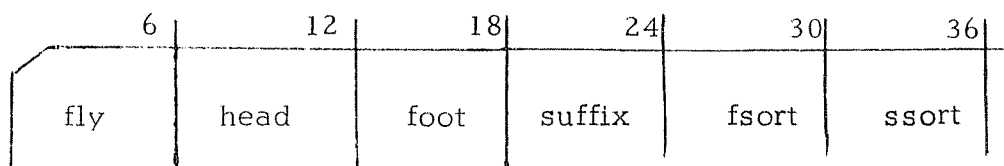| Title	Column	Contents
maxkey	13-18	maximum number of characters output for a keyword; excessive characters are truncated on the right (maximum is 30).[1]
spaces	19-24	number of blank lines between successive entries for a keyword (maximum is 10).
segs	25-30	number of blanks lines between different alphabetical sections (A,B,C, etc.). This is distinct from spaces above. Any number larger than lines forces each new first letter section to the top of a page.
list	31-36	1 list stop words on a separate page 0 suppress stop word listing
notes	37-42	1 print notes 0 suppress notes
unit	43-48	logical unit number of output file (8-28). If zero or blank, logical number 6 is used. Output records consist of 132 characters, the first being the carriage control inserted by BIBCON.
print	54	0 print keywords 3 times to darken the print 1 suppress this feature
dots	60	a character (any character) in column 60 requests that character be repeated on the output between the end of each title and the beginning of the id. A blank here indicates that the option is not desired.

---

[1] This number will be adjusted upwards to fit in the smallest field which is divisible by six. (see Appendix D).

The values on the card must be right justified within their particular 6-character fields. The parameter _width_ must be greater than maxkey + 3 + (id2-id1) to allow room for the title field.[1]

8.5   Titles Card

Cards defining flyleaf, headline, and footline information follow the Output Description card.

Parameters on the first card indicate the number of 80-character lines for each specification: flyleaf, headline, footline. All fields are right justified.

6	12	18	24	30	36
fly	head	foot	suffix	fsort	ssort

Title	Column	Contents
fly	1-6	Punch the number of cards included for each entry (0,1,2,etc.) right justified in the appropriate column. _Fly_ should be no
head	7-12	greater than _lines_ on the Output Format
foot	13-18	Card. The sum of _head_ and _foot_ should be 10 or less. If either of these limits is exceeded, an esoteric error message will be printed and the concordance not run.

---

[1] If _width_--(maxkey+3+(id2-id1)) is not large--say, 12 or greater--the output format may be garbled.

Title	Column	Contents
suffix	24	A 1 here indicates that you will supply your own suffixes rather than using those listed in sections 3.1 and 7.3. For this option to have any effect, however, you must also select either the associative forms feature on the Stopword Control Card or the standard stopwords feature on the General Information Card. The suffixes themselves are included after the stopword cards (see section 8.7a).
fsort	30	To have the keyword list which occurs at the end of the concordance listing sorted by frequency, punch a 1 here.
ssort	36	To have the keywords sorted from back to front (i.e., going is sorted as if it were spelled gniog), punch a 1 here.

Succeeding cards contain text for the flyleaf, headline, and footline in that order. The number of cards for each type of information must correspond to the number of lines specified for each in the parameter card.

If the width of the index page as specified by the width parameter in the Output Description Card is 80 or more characters, the data on each card is centered in the page, and reproduced as punched. If the width of the line is less than 80 characters, only the first width columns of each card appear.

8.6 Stop Word Control

Stop word input immediately follows the flyleaf, headline, and footline cards. The first stop word card specifies the selection of the optional facilities and must be present.

Title	Column	Contents
alphas	6	A one in this column indicates that any word that starts with a non-alphabetic character is to be treated as a stop word (or, keywords, if the submitted word list is to be treated as keywords). This feature will apply only if one or more stop words are supplied. If it is desired without other stopwords, include a single dummy stopword (e.g., ZZZZZZ). If front stripping is specified, this option is checked after front stripping is completed.
assoc	12	If the associated forms feature (see section 3.1) is desired, punch a one here. Otherwise, punch a zero. (If you specified standard stop words by punching an "S" in column 12 of Card 1, associated forms will be checked for regardless of what you punch here.)
sings	18	If you want all one-character words to be treated as stop words, leave this column blank; otherwise punch a one here.
endstrip	19-24	If you want characters other than 0-9 and A-Z to be stripped off of the end of key words, leave these columns blank. If you want to specify characters for this process, punch the total number of these characters (24 is the maximum allowable) in these columns (if this number is less than ten, punch the single digit in col. 24).

ends	25-48	Then, punch the characters themselves in consecutive columns, beginning in col. 25. If you want no end stripping of any kind, punch a one in col. 24 and leave all columns after 24 blank.
fstrip	49-54	This option is similar to the endstripping
fronts	55-78	just described, except that it applied to the beginning of each word. fstrip, specifies the number of characters listed as fronts (max = 24). If fstrip is zero (or blank), however, no front stripping will be done.

This card must be included, even if it contains zeros.

8.7   Stop Word Specification Cards

If the General Information Card has a K in col. 6, then the words

that are to appear in the concordance list must be punched as stop words

(they're not really stop words, but rather key words in this situation; just

ignore the devious nomenclature).

If the General Information Card has an 'S' in column 12, indicating

standard stop words, do not include any stop word specification cards. If

the General Information Card does not have an 'S' in column 12, you must

include at least one stop word card. (If no stop words are to be included

on this card, punch six asterisks in columns 1-6.)

Stop words are punched with at least one blank between succes-

sive words, and must be alphabetized by first letter (deeper

alphabetizing is not necessary), proceeding from A-Z regardless of

which collating sequence (section 8.2) is used. Stop words beginning

with non-alphabetic characters must follow the alphabetic stop words

and may appear in any order. Following the blank after the last stop word must be six contiguous asterisks. This signals the end of the stopword list. Neither stopwords nor the terminating asterisks can be broken across cards. The maximum length for stopwords is 30 characters.

To indicate stopwords to be checked for associated forms, the character sequence _*_ (blank asterisk blank) must follow the stopword and be on the same card as the stopword. Associated forms will not be checked, however, unless the associated forms option is selected on the Stop Word Control Card.

If one stopword forms the first segment of another, as for example, CRACK and CRACKLING, the longer form should precede the shorter one.

Up to 600 stopwords (or keywords) can be specified. If more than 600 are included, the concordance will not be formed and an error message will be printed.

8.7a   Suffix Specification Cards

Up to 50 suffixes may be specified, with each suffix being from one to six characters. (Suffixes longer than six characters will be truncated on the right.) Suffixes are punched in a free field format, using at least one blank between suffixes, and with no suffixes split across cards. At the end of the suffix cards must be a card containing six consecutive asterisks. (For convenience, punch these in cols. 1-6.)

8.8 Normal Forms Cards

If the Normal Form option was specified on the General Information Card (col. 18), then the normal form cards follow immediately after the stop word card (or cards). These cards should be punched the same way as stop word specification cards, except that on each card, the first word is read as the text form of a word, and the second as the corresponding normal form, and so on, pair for pair. However, the same alphabetizing scheme as described for stop words must be used. To be sure that the words are paired properly, punch just one pair per card--the input (text) form in the first field and the normal form in the second field. Pairs can not be broken across cards. Up to 300 pairs may be specified, but if more are included, the concordance will not be formed.

APPENDIX A

Standard Stop Word List[1]

A	Big	Gone
About	Both	Good
Above	But	Got
Across	By	Great
After	Call *	Had
Again	Came	Half
All	Can	Has
Almost	Come *	Have
Alone	Coming	He
Along	Could	Her
Already	Did	Here
Also	Do *	Herself
Always	Does	High
Am	Done	Him
Among	Don't	Himself
An	Dont	His
And	Down	Hold *
Another	During	How
Any	Each	However
Anything	Eight	Hundred *
Are	Either	I
Around	End *	If
As	Even	In
At	Ever	Into
Away	Every	Is
Back *	Everything	It
Be	Far	Its
Became	Few	Itself
Because	First	Just
Become *	Five	Large
Been	For	Last *
Before	Four	Late
Began	From	Least
Begin *	Gave	Less
Being *	Get *	Let *
Best	Give *	Like *
Better	Given	Little
Between	Go *	Live *

[1]Where an asterisk appears after a stop word, that word plus any of the suffixes shown in Section 3.1 will be a stop word.

Word		Word		Word	
Long	*	Out		Ten	
Look	*	Over		Than	
Lost		Own	*	The	
Low		Perhaps		Their	
Made		Put	*	Them	
Make	*	Rather		Themselves	
Man		Real		Then	
Many		Right	*	There	
May		Round		These	
Me		Run	*	They	
Might		Said		Thing	*
More		Same		This	
Most		Sat		Those	
Much		Say	*	Though	
My		Second	*	Three	
Must		See	*	Through	
Myself		Seem	*	Thus	
Near	*	Seen		To	
Need	*	Set	*	Together	
Never		Seven		Too	
New	*	Shall		Toward	*
Next		She		Tried	
Nine		Short	*	Try	*
No		Should		Turn	*
Nor		Show	*	Twenty	
Not		Side	*	Two	
Note	*	Since		Under	
Nothing		Six		Until	
Now		Small		Up	
Number	*	So		Upon	
Of		Some		Us	*
Off		Something		Very	
Often		Soon		Wait	*
Old		Sort	*	Want	*
On		Sound	*	Was	
Once		Still		Way	*
One		Stood		We	
Only		Stop	*	Well	
Open	*	Such		Went	
Or		Sure		Were	
Order	*	Take	*	What	
Other	*	Taken		When	
Our	*	Tell	*	Where	

Whether	Why	Write
Which	Will	Wrong
While	With	Yet
Who	Within	You
Whom	Without	Your
Whose	Would	

APPENDIX B

Keypunch Symbols Arranged in
the Standard Collating Sequence

Symbol as Printed on 1108 Printer	Symbol as Keypunched Model on 026 Keypunch	Symbol as Keypunched on 1108 model 029 Keypunch
@	(7-8)	@
[(| on 9200/9300)	(12-5-8)	[
] (¬ on 9200/9300)	(11-5-8)	]
#	(12-7-8)	#
Δ (▨ on 9200/9300)	(11-7-8)	Δ
blank	blank	blank
A	A	A
B	B	B
.	°	°
.	°	°
.	°	°
Z	Z	Z
)	)	)
‑	‑	‑
+	+	+
<	(12-6-8)	<
=	=	=
>	(6-8)	>
&	(2-8)	&
$	$	$
*	*	*
(	(	(
%	(0-5-8)	%
:	(5-8)	:
?	(12-0)	(12-0)
!	(11-0)	(11-0)
,	,	,
\ (__ on 9200/9300)	(0-6-8)	

The group A through Z is labeled *letters* with a brace.

Symbol as Printed on 1108 Printer	Symbol as Keypunched Model on 026 Keypunch	Symbol as Keypunched on 1108 model 029 Keypunch
0	0	0
1	1	1
.	.	.
numerals .	.	.
.	.	.
9	9	9
'	'	'
;	(11-6-8)	;
/	/	/
.	.	.
X (> on 9200/9300)	(0-7-8)	X
blank (> on 9200/9300)	(0-2-8)	≠

Explanation of Table

This is the list of symbols available on the 1108 printer, arranged in the order in which they alphabetize. E.g., the word NE'ER would appear after the word NEVERMØRE in the alphabetized keyword list, since the symbol ' is later in the collating sequence than the letter V.

There are two models of keypunch available for keypunching input for the 1108. The older Model 026 lacks the ability to keypunch many special characters easily. The newer Model 029 lacks the ability to keypunch only two: the question mark and exclamation point. If you use the 029 keypunch be sure not to use the one marked "B5500 character set."

The table gives instructions for punching the special characters which do not appear on the keypunch keyboard. Depress the multiple-punch key, which keeps the card from advancing more than one column,

and then press the two or three keys indicated in parentheses before releasing the multiple-punch key. The numerals 11 and 12 within the parentheses indicate the minus sign and plus sign keys respectively. The numerals 0-9 indicate the corresponding numeric keys.

If the special symbol denoted variously as ($\neq$) or (") is to be used, then the Bibcon output should be routed to the high speed 9300 printer (see appendix D, example 8).

APPENDIX C

Encoding Scheme for Old English Texts

1. Alphabetic Substitutions

OE symbol	Keypunch
ash	$
thorn	*
eth	+
wynn	W

(All other alphabetics are punched with their Roman equivalents.)

2. Abbreviations and Punctuation

OE symbol	Keypunch
7 (and)	7 (numeric seven)
high point	, (comma)
low point	. (period)

(Other symbols available are percentage sign and numerics other than seven.)

3. Initial Capitalization (encoded at end of word without intervening space)

large cap	)	(right paren)
small cap	))	(two right parens)

(For non-initial caps, the numeric indicators described under Expansions are used. A letter A after a paren indicates a word of all caps.)

4. Expansions (encoded at end of word without intervening space).

$$=n_i.,,,.n_m$$

where n_i is the letter position of an expanded letter in the resulting form.

For example, a͂m would be encoded AMEN=3,4. Inclusive numbers are indicated either by complete enumeration or the hyphenated form $n_a - n_b$.

5. Restorations (encoded at the end of a word without intervening space)

$$/n_i,\ldots,n_m$$

where $n_i,\ldots,n_m$ is as described in Expansions.

6. Superscripts (encoded at the end of a word without intervening space)

$$\neq n_i,\ldots,n_m$$

where $n_i,\ldots,n_m$ is as described in Expansions.

Examples

OE: BVTAN	tweon	lar		
Punched: BVTAN)A	TWEON	LAR		
OE: lufan	ge	gewani͂g F	ꝥ	
Punched: LUFAN	GEWANIGE+=8,9	7	*$T/2,3	

OE: soðlice	sio	lar
Punched: SO+LICE	SIO	LAR
OE: poñ	sio	fylnes
Punched: *ONNE=4,5	SIO	FYLNES

7. Collating Sequence

blank

A

$

B-T

+

*

U-Z

0-9

remaining special characters

APPENDIX D

Tape Formats

Keywords

Keywords are written on logical unit 7 during the concordance output phase and then read and formatted for printing. This unit may be saved by equating it to a user supplied tape.

Records are written at 800 bpi (cpi) in odd parity unless otherwise specified on the @ASG card, 960 words (= 5760 characters) per physical block. (Writing is done with a FORTRAN unformatted WRITE statement.) Unused words in the last physical block are filled with zeros. A single end-of-file mark terminates the file. Logical records will be from 12 to 36 characters in length, depending upon the maxkey specification (Section 8.4). The following chart shows logical record size for all legal maxkey specifications. (The last six characters are a binary integer.)

maxkey	Logical Record Size (No. of Words)
1- 6	2
7-12	3
13-18	4
19-24	5
25-30	6

Concordance Listing

If the output unit for the concordance is not unit 6, the concordance is blocked in SDF format and written on the designated unit with FORTRAN formatted write statements. Each logical record is 132 characters in length, the first character being a carriage control character (blank, 0,1, or +). Tapes are written in odd parity at 800 bpi unless specified otherwise. File one will contain a listing of stopwords with their frequencies of occurrence.

File 2 will contain the concordance records and file 3, the keywords as described above. Each file will be terminated by a single end of file; the last file will have a second end of file after it. If no stop words are specified, file 1 will contain the concordance and file 2, the keywords.

Frequency sort and suffix sort, if selected, create separate files which follow the three files described above (frequency file first, if both are selected).

Sort Records

The sort phase records are written by FORTRAN unformatted write statements on units 2, 3 and 4 in 960 word physical blocks (6 characters/word). Logical length of each record is listed on the output.

The first few logical records in each file will contain control records ; these have keywords composed entirely of zeros and should be ignored. Similarly, the last few logical records in the file are

control records, identified by the pattern XXXX∧∧ in the first six characters of each. (This pattern occurs regardless of the collating sequence chosen.)

Each logical record begins with its keyword field as described in keywords above. This is followed immediately by the id and then the title and notes fields (or notes and title, depending on the ordering of fields in the input record). If datesort is specified, the months which begin with the strings shown in Section 8.2 (JAN, FEB, etc.) are converted to numeric values (1 for JAN, 2 for FEB, etc.) Blanks are converted to zeros. If a non-standard collating sequence is selected, the keyword, but no other part of the record will be encoded for sorting in that sequence.

Merging of Files

Different outputs from the BIBCON sort phase can be merged, using the MACC tape sort package and the resulting file input to BIBCON, phase 3, for formatting and printing. To save the sort phase output, the card sequences shown below should be used (see examples 2 and 6).

Note on Sort Procedures

The sort field used is [Keyword + Ident] or ten words, whichever is least. If notes are not used the Ident is moved to the position between the keyword and the title. If Notes are used the Ident is not moved. Thus if the Ident is to be included in the sort field the input data must have the Ident preceding the Title and notes.

Note on Drum Storage

If your concordance consists of more than about 2000 cards the pre-assigned files may not be large enough!!! Files 2., 3., and 4. should be assigned explicitly by placing the following cards after the @ RUN card:

@ ASG,T   2., F2/128/TRK/4000

@ ASG,T   3., F2/128/TRK/4000

@ ASG,T   4., F2/128/TRK/4000

For more than 1000 cards you should reserve 128 tracks per thousand cards; for more than 20,000 cards you should consider making two or more runs with subsets of the deck and merging the resulting concordances. (A merge routine will be available from MACC consultants.)

Example 1   Standard run

@ RUN...

@ XQT∧BIBCON*X.BIBCON

⟨ BIBCON deck ⟩
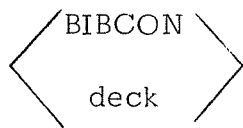
@ EOF

@ FIN

Example 2   Save the sorted file

@ RUN......

@ ASG,TH∧TAPE.,T,$<external tape label>

@ XQT∧BIBCON*X.BIBCON
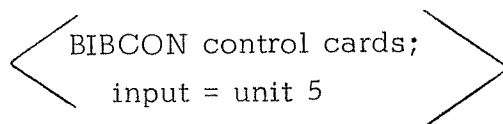
<BIBCON deck>

@ EOF

@ ADD∧29.

@ FIN

The ADD card can also be used to merge two or more files of input data or to read data from a drum file or any other SDF file.

Example 3   Read data from SDF file

@ RUN.

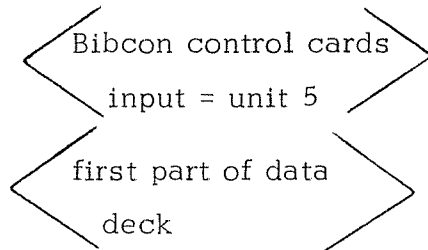@ ASG....(file description; for non-cataloged files only)

@ XQT∧BIBCON*X.BIBCON

<BIBCON control cards; input = unit 5>

@ ADD∧<file name>
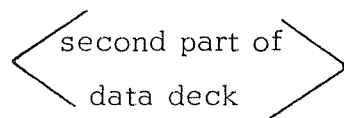
@ EOF

@ FIN

68

Example 4 Insert data from SDF file

@ RUN

@ ASG... (file description; for non cataloged files only)

@ XQT∧BIBCON*X.BIBCON

> Bibcon control cards
>
> input = unit 5

> first part of data
>
> deck

@ ADD∧ <file name>

> second part of
>
> data deck

@ EOF

@ FIN

Example 5   Save Keywords

@ RUN...

@ ASG,TH   SEVEN.,T,$<external tape label>

@ XQT BIBCON*X.BIBCON
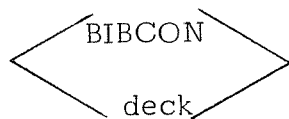
> BIBCON
>
> deck

@ EØF

@ COPY,GM   7.,SEVEN.

@ FIN

Example 6   Save the unsorted file

@ RUN...

@ ASG,TH   TAPE.,T,$<external tape label>

@ XQT BIBCON*X.BIBCON

⟨   BIBCON   checkpoint = 1   ⟩
            decl
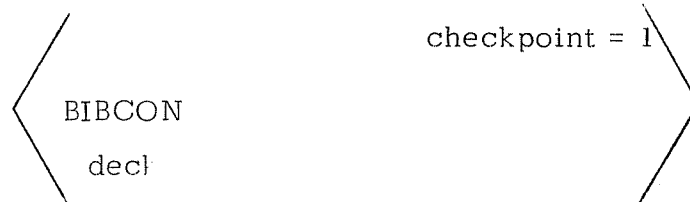
@ EOF

@ ADD 29.

@ FIN

Example 7: Restart, assume the unsorted or sorted files have been saved
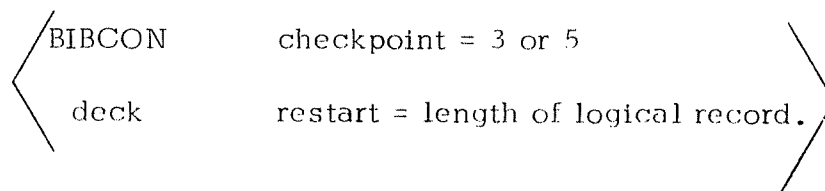
as in example 2 or 6 above

@ RUN...

@ ASG,T   Z.,F2///4000

@ ASG,TH   TAPE.,T,   <external tape label>

@ COPY,G   TAPE.,2.

@ XQT   BIBCON*X.BIBCON

⟨ BIBCON   checkpoint = 3 or 5   ⟩
     deck   restart = length of logical record.

@ EOF

@ FIN

Example 8   To force printed output to the 9300 printer

   7/8 RUN

   7/8 SYM,S PRINT$,,HSPR93

   7/8 XQT BIBCON*X.BIBCON

          <BIBCON deck>

   7/8 EOF

   7/8 FIℓ

The 7/8 SYM card may be used for any of the examples above and should immediately follow the RUN card.

APPENDIX E

MACC Control Cards[1]

All Control Cards in this section require a 8 multiple punch ("8 " on 1108 keypunches) in Col. 1. The other information must begin in Col. 2.

RUN Card

RUN name,proj#,user#,time,page

name   is the last name of the user, or any other name

              that does not contain a comma.

proj#   is the project number assigned by the MACC; in

              other words, the source of funds for running the

              concordance.

user#   is the user's MACC number, also assigned by

              the MACC.

time   time is the time limit in minutes for the job,

              given as nM or nS or nMqS, where n and q

              are integers and S implies seconds, M,

              minutes. If the job is not finished by the time

              it has run the specified limit, it will be termi-

              nated at whatever stage it is at. The maximum

---

[1] As of October 1, 1971; MACC stands for Madison Academic Computing Center. For variations on control cards and for situations not described here, see the 1108 Computing Handbook, The Madison Academic Computing Center.

number of pages allowed. Since the likelihood

of BIBCON printing excessive numbers of 'garbage'

pages is remote, punch 4000 here, which is the

maximum allowed.

Execute Card

$\frac{7}{8}$ XQT∧BIBCØN*X.BIBCON

EOF Card

$\frac{7}{8}$ EØF

FIN Card

$\frac{7}{8}$ FIN

ASG Card[1]

$\frac{7}{8}$ ASG,TH∧<unit number>.,T,$external label

$\frac{7}{8}$ ASG,T ∧<unit number>.,F///4000

$\frac{7}{8}$ ADD∧<unit number>.

$\frac{7}{8}$ SYM,S PRINT$,,HSPR93

---

[1] For variations on these parameters, see the MACC Handbook.