28

BIBCON

A 3600 Program For Producing Concordances To Prose, Poetry and Bibliographic References

by

Richard L. Venezky

Computer Sciences Technical Report #34

March 1969

CONTENTS

APPENDICES

INTRODUCTION

BIBCON is a computer program which generates concordances to prose, poetry, bibliographic references, and other strings of symbols.[1] It runs on the 3600 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 but adequate technical description of the program and its use. For the reader who experiences difficulty in choosing between pleonasm (section 1) and laconism (section 2), a short quiz is included on the next page, complete with instructions for scoring and routing.

---

[1] BIBCON is a variation on Control Data Corporation's KWIC and was developed by Mr. Nathan Relles and the present author.

QUIZ

(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 record 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 Exit 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 the next page.
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

Cap. VI. 1. behaldeþ ꝑ ge eowre soþfestnisse ne doan fore monnum ꝑ ge sie geseonæ from heom from him (*sic*) elles ꝛ elcur ge ne habbaþ lean ꝛ 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 ⁊ in tunum ꝑ hie sie weorþade from monnum soþ ic sæcge eow hie onfengun heora lean 3. ðe þonne wircendum ælmesse nyte se winstræ hond þin hwæt þin sio swiþre dóa 4. þæt þin ælmes sie in degulnisse ⁊ þin fæder se þe gesið in degulnisse geldeþ ðe

Figure 1

Chapter VI, verses 1-4 of the Rushworth Matthew, edited by W.W. Skeat. (The Holy Gospels, 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$NSIGA*
    WA EOW BUKERES 7 FARISSEAS LICETERAS FOR*ON=23 +E GE CL$NSIGA* (   .    .  MATT 23 25
    UTAN IS C$LCES 7 .... BINNE *ONNE=45 FULLE SINDUN NEONIMENDE 7
    UN-CLENNISSE

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

CNEHT
    HENU CNEHT MIN *ONE 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$LEND=456 7 EODE FROM=4 HIM ( DEOFUL 7 GEH$LED   W$*  MATT 17 18
    SE CNEHT

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

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

    *A HIE WERON GEWITEN$ HENU ENGEL DRIHTNES $TEAWDE IN SWEFNE IOSEP .    .  MATT  2 13
    CWE*ENDE ARIS 7 GENIM *ONE CNEHT 7 HIS MODER 7 FLEOH 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. In such cases, the terminal punctuation is ignored. Thus, the following are all the same word (∧ means blank):

∧JOE ∧ ∧JOE:∧

∧JOE,∧ ∧JOE-∧

∧JOE.∧

No word can be more than 80 characters in length. If so, the record in which it occurs will not be included in the concordance and an error message will be printed.

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

---

[1]The program and the computer will be used interchangeably and anthropromorphically. 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.

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 preparation of stop word and key word lists is described in section 4.7.

Input preparation

2.1 Input is divided into units called records, 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) -- Id (identifier) and Title, if only two components, and these plus Notes, if three. Each record is punched into cards, with the contents of each component (hereafter called fields) 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 element (character string) that will be placed in that field. Elements 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 starting and ending

---

[1] A character is equivalent to a space (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.

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 record).

insert figure 3 here

Words from the Title field only are included in the alphabetized list. However, the other fields 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. Words are alphabetized by BIBCON by first placing each in a fixed length field (eight or sixteen 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 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.

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 6	1 BEHALDE* (GE EØWRE SØ*FESTNISSE NE DØAN FØRE MØNNUM (GE SIE	GESEAN$
FRØM=4	HEØM=4 FRØM=4 HIM ELLES Q ELCUR GE NE HABBA* LEAN Q MEARDE MID E	ØWER F$D
ER *$NE	*E IN HEØFUNUM=8 IS	
ATT 6	2 FØR*ØN=23 *ØNNE=45 *U WIRCE $LMISSE NE BLAU *U BEMAN FØR *E S	WA LICET
ERAS DØA	N IN HEØRA SØMNUNGUM=9 7 IN TUNUM (HIE SIE WEØR*ADE FRØM=4 MØNN	UM SØ* I
C S$CGE	EØW HIE ØNFENGUN HEØRA LEAN	

Figure 3

Coding for Rushworth Matthew (see Figure 1, section 1.1).

A coding 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 employ particular stopwords or keywords, but also employ 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	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 stop-word (or keyword) list, the end of the word is compared with these suffixes, starting with the longest suffix and searching for the first match. If a match is found, the suffix is stripped from the word, and the remaining characters compared with the stop word (or keyword) list.

Normalized forms

3.2 Once again by option, normal forms of up to 100 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 dette, an output listing might appear as:

DEBT WE INCUR'D A LARGE DETTE A6 21

Keyword list

3.3 After the last entry of the concordance, a list of all keywords and their frequencies of occurrence will be printed.[1]

Specification of Input

4.1 The complete input to BIBCON includes the following:

1. Computing Center control cards
2. BIBCON control cards
3. Input data (records containing prose, or poetry, or whatever) No end-of-data card is needed.

Input data preparation was discussed in the preceding sections. The computing center control cards are discussed in Appendix D; this

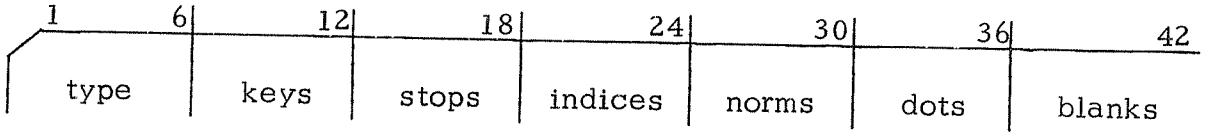
---

[1] You can obtain this list on a magnetic tape, also. To do so, see section 7.4 and Appendix D.

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.)

4.2 Card 1: Concordance Type

The first card specifies the type of concordance desired and certain other options. For the present, only the standard index (concordance) will be designated. A special bibliographic index, described in section 5.1, is also available, but will be ignored for the present. 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.



Title	Column	Contents
type	6	Punch an I here to designate the standard index (concordance) as the one desired.
keys	12	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).

Title	Column	Contents
stops	18	If you want to use the standard list of stop words (Appendix A), punch an S here (keys must, therefore, be blank). Otherwise, leave this column blank.
indices	24	Ignore this column.
norms	30	If you are specifying normalized forms, punch an N here (specification of the forms themselves is discussed in section 4.8). Otherwise, leave this column blank.
dots	36	A period in this column requests that periods be printed 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.
blanks	42	The number in this column (0-9) indicates the number of blank lines which will appear between different keyword listings on the output.

4.3 Card 2: Input Format

In all of these fields, the numbers must be punched as far to the right as the field will allow (right justified).

1 6	12	18	24	30	36	42	48	54	60	66
length	ratio	id1	id2	title1	title2	notes1	notes2	average	unit	parity

Title	Column	Contents
length	4-6	The length in characters of each (logical) record (see section 2.1). This must be 400 or less, and be a multiple of 8.
ratio	10-12	This is the ratio of logical records to physical records, or vice-verse. 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 record. If data input is from cards or card images on magnetic tape, then Ratio is the number of cards (or card images) needed to make one logical record. If this number is greater than one, then it must be preceded by a minus sign. Ratios for various lengths (assuming card input) are shown below.

Length of Logical Record	Ratio (Cols. 11-12)
80	1
160	-2
240	-3
320	-4
400	-5

Title	Column	Contents
		If input records are blocked on tape--that is, if more than one card image is contained in each physical tape record[1] and there is at least one logical record in each physical record, then ratio is the ratio of logical records to physical records and is not preceded by a minus sign.
id1 id2	16-18 22-24	The first (id1) and the last (id2) character positions of the Id field. (For figure 3, section 2.1, id1 is 1 and id2 is 7).
title1 title2	28-30 34-36	The first (title1) and last (title2) character positions of the Title field. (For figure 3, section 2.1, title1 is 10 and title2 is 160).
notes1 notes2	40-42 46-48	The first (notes1) and last (notes2) character positions of the Notes field. If Notes are not 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.)

---

[1] A physical tape record 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 record will be: record gap, two cards (160 characters), record gap. The maximum allowable length of blocked records for BIBCON is such that length times ratio is 10,000.

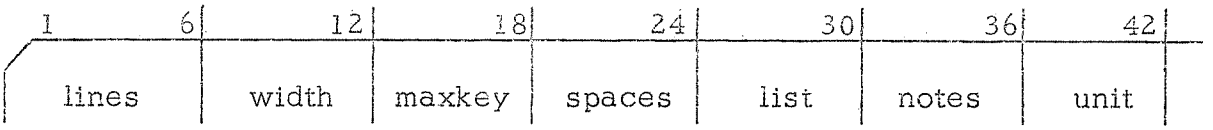
Title	Column	Contents
average	49-54	Average number of keywords in a title (if you don't want to figure it, punch 6 here).
unit	59-60	This is the input unit number. If cards are used, leave this blank (BIBCON will assign it to the appropriate unit). If card images on tape are used, punch 10 here, and add the TAPE and EQUIP Computing Center control cards shown in Appendix D.
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 even parity when you do and don't punch anything 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.

4.4 Card 3: Output Format

All values must be right justified.



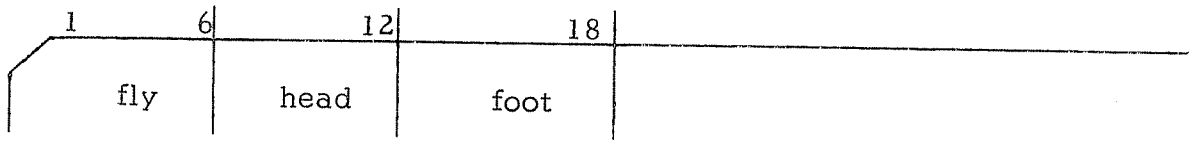
Title	Column	Contents
lines	1-6	Number of lines allowed on each output page, (Unless you have reasons for doing otherwise, punch 60 here.)
width	10-12	Number of characters across width of output page. (Unless you have reasons for doing otherwise, punch 120 here.)[1]
maxkey	17-18	Maximum number of characters in each keyword listing (this can be from 1 to 16). Keywords greater than the number punched here will have the rightmost characters chopped off. (16 is a safe number to punch here).
spaces	23-24	Number of blank lines between successive entries for a keyword. A blank implies zero.
list	30	Punch a 1 if you want stop words listed before the concordance is printed. Leave blank otherwise.

[1] If not 120, then it must be greater than the number in cols. 17-18, plus 3, plus id2 minus id1.

Title	Column	Contents
notes	36	Punch a 1 if you want the Notes printed with each concordance entry. If there are no Notes, or you don't want them listed, leave blank.
unit	41-42	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 an 11 here and add the Computing Center control cards (TAPE and EQUIP) shown in Appendix D. (By supplying your own tape, you can re-list the output at a significantly lower price than you can by re-running the program.)

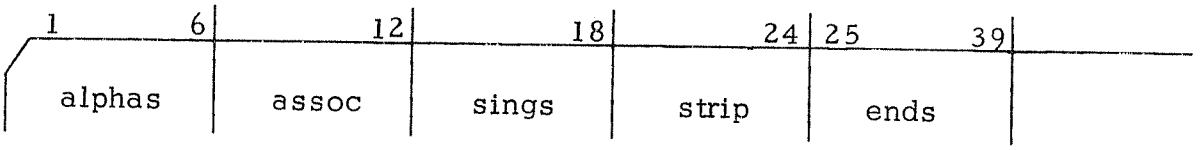
4.5 Card 4: Titles

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.



Title	Column	Contents
fly	6	Punch zero if no flyleaf title cards are present, a 1 if one title card follows, etc. The same procedure is followed for headlines and footlines. The total number of title cards must not be greater than 8.
head	12	
foot	18	

4.6 Card 5: Stopword Control



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,) If this feature is not desired, punch a zero in this column.
assoc	12	If the associated forms feature (see section 3.1) is desired, punch a one here. Otherwise, punch a zero.

Title	Column	Contents
sings	18	If you want all one-character words to be treated as stop words, leave this column blank; otherwise punch a one here.
strip	23-24	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 (15 is the maximum allowable) in these columns (if this number is less than ten, punch the single digit in col. 24).
ends	25-	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.

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

4.7 Card 6: Stop word specification

If card 1 has a K in col. 12, 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 no stop words are included, punch asterisks in Cols. 1-8. Otherwise, punch the stop words in the first 8 columns of each 10 column field, starting in column 1. Thus, the first stop word goes in cols. 1-8, the second in 11-18, and the eighth in cols. 71-78. Stop words must be alphabetized by first letter (deeper alphabetizing is not necessary). A field of eight asterisks must follow the last stop word (and be punched where a stop word could be punched.) Fields before the asterisks must not be left blank.

If associated forms was selected on Card 5, then punch an asterisk in the 8th column of each stop word field which is to be checked for associated forms. Fields without an asterisk in column 8 (of that field) will not be checked for associated forms, even if column 12 on card 5 contains a 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.

4.8 Card 7: Normalized Forms

If the Normalized Form option was specified on card 1 (col. 30), then the normalized form cards follow immediately after the stop word card (or cards). The input form of a word is punched in cols. 1-8 (starting in column 1), the normalized form in cols. 11-18 (starting in column 11). This is repeated for each pair in cols. 21-40, 41-60, and

61-80. Successive cards should be used if necessary (up to 100 pairs can be included. If more are specified, the concordance will not be formed). After the last word pair, a field of 8 asterisks must be punched to signal the end of normalized forms.

If the normalized form option was not selected on card 1, no normalized form cards can be included, not even one containing eight asterisks only.

4.9 Examples

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

insert figure 4 here

2. For the poetry shown in figure 5 below, one punched card will be used for each line. This card will contain the line of poetry plus the line number. Therefore, each logical record will be 80 characters in length and contain an Id and Title field only. The Id field will be punched in columns 1-9, with WW-TCP (for Walt Whitman-To a Common Prostitute) in columns 1-6 and the line number in cols. 8-9. The Title (line of poetry) will go in cols. 10-80.

A flyleaf, but no headlines or 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

```
     I       S
   320      10       1      10      12     320
    60     100      16       1       0       0      17
     1       0       0
    CONCORDANCE TO RUSHWORTH 1--MATTHEW      R. VENEZKY 26 JULY 1968
     0       0       1      14='/),123456789
**********
```

Figure 4

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

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.

```
      I      S
    160     -2      1      9     11    160
     60    120     16             0      0
      1      1      0
A CONCORDANCE TO WALT WHITMAN S, TO A COMMON PROSTITUTE
WALT WHITMAN--TO A COMMON PROSTITUTE
      0      0      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."

CONCORDANCES TO BIBLIOGRAPHIC REFERENCES

Introduction

5.1 Concordances to titles, authors, or both can be produced by BIBCON, using either of two input forms. Form 1 (called Effortless) requires that you punch one BIBCON control card (which has only one character punched in it) and then the bibliographic references in the format shown below. These plus the Computing Center control cards (see Appendix D) complete your part of the job. The rest, while hardly magic, requires no intervention on your part.

Under Form 2 (called Effortful) you specify a large number of options, but to do so you must first read sections 1-4, especially the material on stopwords, keywords, normalized forms, and the output format card.

Format for Bibliographic References

5.2 This format is used for both Form 1 and Form 2. Each citation consists of from three to five punched cards, containing author, title, and source for each reference. The type of card is indicated by a character punched in column one: A for author, T for title, and S for source. Only one type of information (author, title, source) can be placed on a single card.

There can be only one author card, but either one or two title cards, and either one or two source cards. Column one, as mentioned above, contains A, T, or S. Columns 2-8 can either be left blank or punched

with a number, selected by you, which is identical for all cards within a single reference. (Only the numbers 0-9 or blanks can appear in cols. 2-8.) If all A, T, and S cards for a single reference do not contain the same quantity in cols. 2-8 (blank or an actual number), BIBCON will not include that reference in the concordance, but rather will print it as an error. This number can also be your own reference number, and in addition, is a handy aid for reassembling dropped decks. Cards for a reference must be in the order: A T(T) S(S). The complete concordance will include all words which occur on T cards, except those listed in Appendix A (called stop words). An example of Form 1 input is shown in figure 8. (See figure 9 for a concordance to this input.)

figure 8

Form 1

Form 1 requires a single BIBCON control card and this has a B punched in column 6. (The remainder of the card is left blank). This card goes before the bibliographic references, but after the Computing Center control cards shown in Appendix A. For form 1, only punched cards can be used as input. If you want to place the cards on magnetic tape first, then you must use Form 2.

The concordance will contain in the alphabetized list all words[1] in reference titles except:

[1] For a definition of word as used by this system, see section 1.1.

```
        B
A         1WEPMAN, JOSEPH M.
T         1AUDITORY DISCRIMINATION, SPEECH AND READING
S         1ELEM SCH J, 60, (MARCH 1960), 325-33.
A         2FAIRBANKS, GRANT
T         2VOICE AND ARTICULATION DRILLBOOK
S         22ND ED., (1960), HARPER AND ROW, NY.
A         3ELKONIN, D. B.
T         3THE PSYCHOLOGICAL MASTERING OF THE ELEMENTS OF READING.
S         3IN, EDUCATIONAL PSYCHOLOGY IN THE USSR, ED.BRIAN AND SIMON, STANFORD UNI
S         3VERSITY PRESS,
A         4BUROS, O. K. (ED.)
T         4THE 6TH MENTAL MEASUREMENTS YEARBOOK
S         4(1965), HIGHLAND PARK, NJ, GRYPHON PRESS.
A         5SAMPSON, O. C.
T         5THE SPEECH AND LANGUAGE DEVELOPMENT OF 5 YEAR OLD CHILDREN.
S         5BRIT JEP, 29, (1959), 217-22.
A         6BRUCE, D. J.
T         6ANALYSIS OF WORD SOUNDS BY CHILDREN.
S         6BRIT JEP, 34, (1964), 158-69.
A         7WINTZ, HARRIS AND BELLROSE, BETTY
T         7PHONEME-CLUSTER LEARNING AS A FUNCTION
S         7JVLVB, 4, (1965), 98-102.
A         8SAMPSON, O. C.
T         8READING SKILL AT 8 YEARS
S         8BRIT JEP, 32, (1962), 12-17.
A         9MCDONALD, EUGENE T.
T         9A DEEP TEST OF ARTICULATION.
S         9(1964), STANWIX HOUSE, PITTSBURGH, PA.
A        10HICKOK, CRAIG
T        10PRELIMINARY BIBLIOGRAPHY ON VERBAL CONDITIONING AND THE FIRST AND SECOND
T        10SIGNAL SYSTEMS.
S        10R AND D, (1960).
```

Figure 8

Bibliographic input, form 1

1. those listed in Appendix A

2. any single character word

Part of a concordance to a bibliography is shown in figure 9.

insert figure 9 here

Form 2

Form 2 uses the same reference format as Form 1, but requires more control cards.

5.3 Control Card 1 is the Concordance Type Card shown in section 4.2, with a Q punched in column 6. (The contents of the other columns are explained in section 4.2.)

Control Card 2 is the Bibliography File Card. Columns 5-6 indicate where the input references are. If they are on cards immediately following these control cards, then leave columns 5-6 blank. If the card images have been transferred to magnetic tape, then punch 10 here, and insert the EQUIP and TAPE cards shown in Appendix D in the control card sequence.

Column 12 is used only if the data input is from magnetic tape. If that is the case, then the parity of the tape goes here: blank or zero for even (BCD) parity, 1 for odd (binary) parity. (See section 4.3, for a definition of parity).

ANALYSIS
ANALYSIS OF WORD SOUNDS BY CHILDREN. BRUCE, D. J.
BRIT JEP, 34, (1964), 158-69.

ARTICULATION
VOICE AND ARTICULATION DRILLBOOK FAIRBANKS, GRANT
2ND ED., (1960), HARPER AND ROW, NY.

A DEEP TEST OF ARTICULATION. MCDONALD, EUGENE T.
(1964), STANWIX HOUSE, PITTSBURGH, PA.

AUDITORY
AUDITORY DISCRIMINATION, SPEECH AND READING WEPMAN, JOSEPH M.
ELEM SCH J, 60, (MARCH 1960), 325-33.

BIBLIOGRAPHY
PRELIMINARY BIBLIOGRAPHY ON VERBAL CONDITIONING AND THE FIRST AND SECONDSIGNAL SYSTEMS. HICKOK, CRAIG
R AND D, (1960).

DEEP
A DEEP TEST OF ARTICULATION. MCDONALD, EUGENE T.
(1964), STANWIX HOUSE, PITTSBURGH, PA.

DEVELOPMENT
THE SPEECH AND LANGUAGE DEVELOPMENT OF 5 YEAR OLD CHILDREN. SAMPSON, O. C.
BRIT JEP, 29, (1959), 217-22.

Figure 9

Part of the concordance to the input shown in figure 8.

The remaining control cards are those shown in sections 4.4 through 4.8, starting with the Output Format Card. A sample deck structure is shown in figure 10.

figure 10

```
     Q            S                          3

    60    120     16      1      1      1
     1      1      0
               CONCORDANCE TO REFERENCES ON READING
    READING BIBLIOGRAPHY--SEPT. 1968
                    1
A          AN          IN          OF          THE         ********
A       11FOWLER, W.
T       11COGNITIVE LEARNING IN INFANCY AND EARLY CHILDHOOD.
S       11P BULL, 59, (1962), 116-52.
A       12ROAB, S., DEUTSCH, M. AND FREEDMAN, A.
T       12PERCEPTUAL SHIFTING AND SET IN NORMAL SCHOOL CHILDREN OF DIFFERENT READI
T       12NG ACHIEVEMENT LEVELS.
S       12PERCEPT AND MOTOR SKILLS, 10, (1960), 187-92
A       13PICKETT, J. M.
T       13PERCEPTION OF COMPOUND CONSONANTS.
S       13LANG AND SP, 1, (1958), 288-304.
A       14SALZINGER, SUZANNE
T       14MEMORY FOR VERBAL SEQUENCES AS A FUNCTION OF THEIR SYNTACTICAL STRUCTURE
T       14AND THE AGE OF THE RECALLING CHILD.
S       14JP, 64(1), (1966), 79-90.
A       15TEMPLIN, MILDRED D.
T       15CERTAIN LANGUAGE SKILLS IN CHILDREN, THEIR DEVELOPMENT AND RELATIONSHIPS
S       15INST CHILD WELF MONOG, SER-26, MINNEAPOLIS, UNIVERSITY OF MINN. PRESS
A       16BRACKBILL, YVONNE
T       16AROUSAL=  IMPLICATIONS FOR LEARNING TO READ PARADIGM
S       16PROJ LIT REP, NO.1, JULY 1964, 3-7.
A       17WRIGHT, JOHN C.
T       17ACQUIRED RELEVANCE OF CUES IN READING=  THE LEARNING OF SELECTIVE OBSERV
T       17ING RESPONSES
S       17PROJ LIT REP, NO.1, JULY 1964, 7-10
A       18WILLIAMS, JOANNA
```

Figure 10

Deck structure for bibliographic input, form 2.

SECTION II

Reference Manual for Programmers

INTRODUCTION

6.1 BIBCON is a 3600 system for producing concordances to two basic forms of input, standard and bibliographic. Standard 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.

Bibliographic input utilizes column one punches for designating author, title, and source entries. In addition, certain field lengths are variable; the maximum logical record size, however, is 400 characters. While concordances to bibliographic data can be generated from either input form, the bibliographic form is easier to use than the standard form for this type of input.

For both input forms, the following features are available:

1. Blocked or unblocked input on any accessible unit.
2. Specification of stop words or keywords, or use of program-supplied stop words.
3. Normalization of up to 100 different input words.
4. Separate listing of all keywords with frequency of occurrence counts.
5. Selection of output unit.

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 Input

The input to BIBCON consists of the following three items:

1. Computing Center control cards (see Appendix D)

2. BIBCON control cards (see section 8)

3. Data records (see section 7)

6.3 Output

The output from BIBCON consists of the following:

1. List of control instructions

2. Stop words (optional)

3. Suffixes (optional)

4. Concordance

5. Totals for:

   Titles read in

   Stop words used

   Key words found

   Index entries

   Pages

   Time used

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

Data Records

7.1 A logical record for standard input, which may describe a line of poetry, or a bibliographic reference, or any string of characters, contains 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). A logical record for bibliographic input consists of author, source, and title cards, as described in section

Stopwords and Keywords

7.2 Concordances are formed to words in the title field only for standard input, and title, or author plus title for bibliographic input.[1] If a word is over 80 characters in length, the record in which it occurs will not be included in the concordance and an error message will be printed.

Words not to be included in the alphabetized portion of the concordance (stopwords) or words to be included (keywords) can be specified by the user. Or the user can specify that the words listed in Appendix A 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

Each input word is first checked against the stop word list (assuming that stop words are used). If a match is found the word is rejected and the next word located. If no match is found, the ending of the word is compared to the suffixes shown above, starting with the longest suffix. If a suffix match is found, the remaining portion of the word (the stem) is matched against those stop words which have an asterisk in character position 8. If no match is found, the suffix-stem procedure is repeated until a match is found or all possible suffixes have been tested.

Normalized forms

7.4 At the user's option, up to 100 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. This option is restricted to forms of eight characters or less. 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, equate unit 7 to a user supplied tape. (Tape format for unit 7 is described in Appendix E.)

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 D.)

Necessary	Optional
1. Concordance Type	
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

8 asterisks)

(5b. End stop words--

used only if no stop

words are included.)

6. Normalized Forms

(terminated by a field of

8 asterisks)

8.2 Concordance Type

The first card describes the type of index as well as certain user-specified options. All fields are right justified.

1 6	12	18	24	30	36	42	
type	keys	stops	indices	norms	dots	blanks	

type type of index desired;

I standard index; in this case, this card is followed by the input format card, output format card, flyleaf card and stop word cards and normalized form cards if the user desires them.

B standard bibliographic index; no additional control cards are necessary; this card is followed immediately by the reference sets as described in section 9. All parameters are set internally; the standard list is used as a stop word list; no stop words, normalized forms, or flyleaf, headlines, or footlines may be provided by the user. Indexing is performed on the titles of the reference sets.

Q bibliographic index with options; while the input format of the reference sets remains the same as for the standard bibliographic index, the user may specify certain options on the following cards: bibliography file, output format card, stop word card(s), flyleaf, headline and footline card(s), normalized form cards (optional). A blank in column 6 indicates the standard index and is therefore treated as 'I' above.

keys — 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" .

stops — S standard stop word list desired; any other character in this column indicates that the user will provide either key words or stop words (depending on key-stop.)

indices — if type is "Q", the user may specify whether he desires indexing to be performed on the title, author, or both. Otherwise, this column remains blank.

A indexing is to be performed on the author field

B indexing is to be performed on both the author field and the title field

any other character in this column indicates that indexing is to be performed on the title field only.

norms — indicates whether the user will provide a list of normalized words;

N a list of normalized words will be proved by the user following the stop word cards.

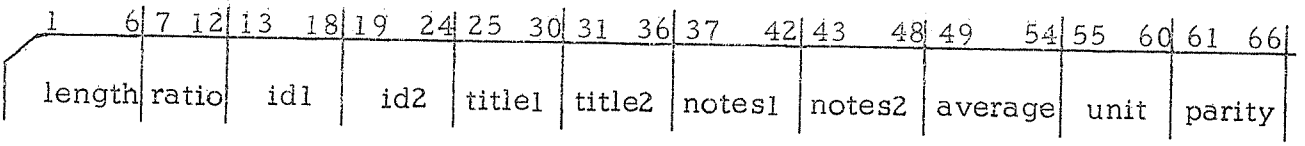
any other character in this column indicates that no normalized forms are desired by the user.

dots — a period in column 36 requests that periods be placed on the output between the end of each title to the beginning of the id. a blank here indicates no periods.

blanks — number of blank lines between different alphabetical sections (A,B,C, etc.). This is distinct from spaces on the Output Format card (section 8.4).

8.3 Input Format Card

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



length — number of characters per logical record (must be a nutiple of 8; maximum is 400).

ratio — number of logical records (LR) per physical record (PR) if LR $\geq$ PR. If PR > LR, then ratio is the number of physical records per logical record and must be preceded by a minus sign. If ratio is positive, length times ratio cannot be greater than 10,000.[1]

id1 — position of first identifier character in logical record.

id2 — position of last identifier character in logical record (must be greater than id1).

title1 — position of first title field character in logical record.

title2 — position of last title field character in logical record (must be greater than title1).

notes1 — position of first floater character in logical record (zero indicates no floater).

notes2 — position of last floater character in logical record (zero indicates no floater; otherwise, must be greater than notes1).

average — average number of key words per logical record. If zero or blank, a value of 5 is assumed.

unit — logical unit number of input file (8-48); if blank or zero, 60 is assumed. Logical units 2-4 are used by KWIC for sorting. Logical unit 7 is used for the list of keywords and frequencies. Units 1, 5, and 6 are exorcised for programming convenience.

---

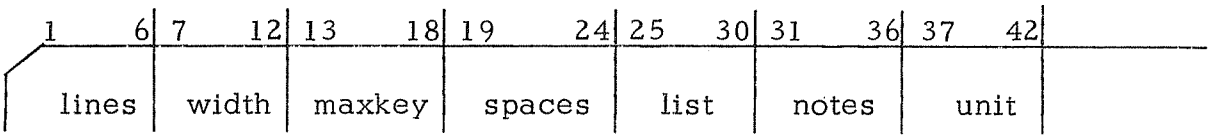
[1] If so, an error message will be printed and the run aborted.

parity — parity of unit; blank or zero implies BCD, 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 floater fields may overlap.

8.4 Output Format Card

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



lines — number of lines per page.

width — number of characters across width of page (maximum is 135).

maxkey — maximum number of characters output for a keyword; excessive characters are truncated on the right (maximum is 16).

spaces — number of blank lines between successive entries (maximum is 10).

list
- 1 list stop words on a separate page
- 0 suppress stop word listing

notes
- 1 print notes
- 0 suppress notes

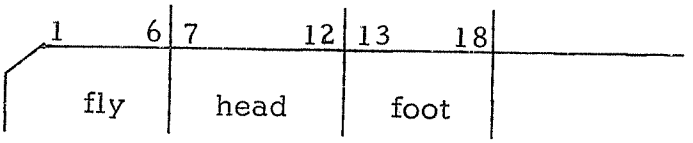
unit — logical unit number of output file; (8-48). If zero or blank, logical number 61 is used. If unit is not zero, blank or 61, output is blocked, 30 lines to the physical record. Output lines are restricted to 119 characters or less under this condition. Each print line consists of 120 characters, the last being the end of record character (32_8).

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.

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.



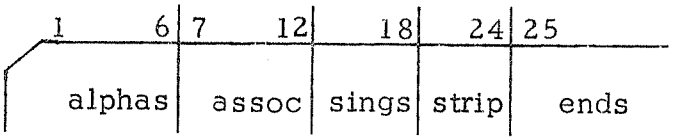
The values on the card must be right justified within their particular 6-character field. The number of headlines and footlines (head + foot) must not exceed 8; otherwise, an error diagnostic is given, and the run is terminated.

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 Cards

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



alphas 1 selects words beginning with a non-alphabetic character as stop words

0 this facility not desired

assoc 1 checks for associated forms

0 this facility not desired

sings 1 do not select one-character words as stop words.

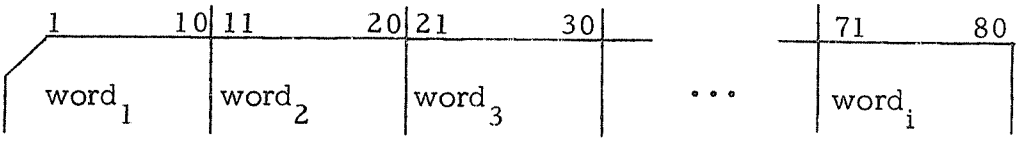
0 select one-character words as stop words.

strip $n > 0$ the n characters ($n \leq 15$) starting in col. 25 are to be stripped from the end of key words.

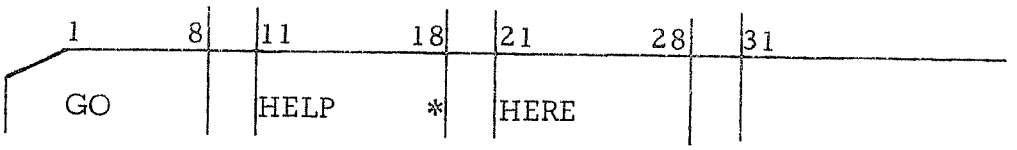
0 standard end-stripping (0-9 and A-Z are not stripped; all other characters are). In this situation ends is left blank.

ends If strip is greater than zero, enter characters to be stripped here in consecutive positions.

Succeeding cards contain the stop words, packed up to 8 per card:



Simple stop words are specified left justified in each 10-character field. The ninth and tenth characters of each field are ignored. Stop word stems from which associated forms are to be checked are specified left justified in the field with an asterisk in the eighth column of the field. For example:

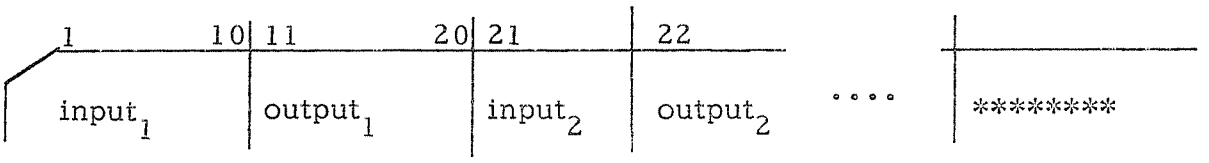


A field containing all blanks signifies the end of stop words on a card; a field containing all asterisks signifies the end of stop word input.

Stop words must be listed so that all words beginning with A precede all words beginning with B, and so forth. All words whose first character is not alphabetic must follow the words beginning with Z. Within a given letter class, the stop words may occur in any order. If a stop word is out of order, a diagnostic is given and the job terminated. Even if no stop words are used, a field of 8 asterisks must occur.

8.7 Normalized Forms

If the Normalized Form option was selected on card 1 (col. 30), then the normalized form cards follow immediately after the stop word card(s). If this option was not selected, either the input data (for card input) or nothing (for input other than on cards) follows the stop word card(s). (An end of file card is supplied by the Computing Center.) The input form of a word is punched in cols. 1-8, left justified; the normal form in cols. 11-18 (left justified). This is repeated in cols. 21-40, ..., 61-80. Up to 100 pairs may be included. If more are submitted, the run will be terminated abnormally. A field of 8 asterisks signals the end of the normalized forms. No blank fields should occur any where before the asterisks field.



9.1 Bibliography Control Cards

Card 1: Concordance Type (see section 8.2)

If the standard bibliographic index is selected, Card 1 is the only control card needed. It is followed immediately by the input data. Otherwise, the follow control cards must follow card 1:

Card 2: Bibliography File (see below)

Card 3: Output Format (see section 8.4)

Card 4: Titles (see section 8.5) and title cards, if titles are selected.

Card 5: Stop Word Control (see section 3.6) and either a card with asterisks in cols. 1-8 or Stop Word Cards.

and, if selected, Normalized Forms Cards.

9.2 Bibliography File Card

(All fields are right justified)

Column		
5-6	unit	Unit number for input file (8-48, 60). If blank or zero, unit 60 is assumed.
12	parity	Parity of input unit; blank or zero indicates even (BCD) parity, any other number indicates odd (binary) parity.

Bibliography Data Cards

9.3 Input data for the bibliographic index consists of reference sets containing one author card, one or two title cards, and one or two source cards. Following is the arrangement of the reference sets:

A punched in column 1 for the author card, followed by a maximum of seven digits for the reference set number (RSN). The data begin in column 9.

T punched in column 1 for the title card, followed by a maximum of seven digits (RSN). The data begin in column 9.

S punched in column 1 for the source card, followed by a maximum of seven digits (RSN). The data begin in column 9.

For any single reference set, a single author card is allowed and either one or two cards for each for the title and source. The reference number, which is optional, must be the same for each card within a single reference set; no check is made on the ordering of the numbers among different sets. The cards must be in the order: A T (T) S (S). There must be at least one of each card type. Optional cards which are not included (T or S) are replaced by BIBCON with blanks; extra cards are ignored. Errors in reference set numbers within a set are listed and those sets are ignored, i.e., not processed. In all of these cases a message giving the reference set number is printed and processing continued.

If a card does not contain an A, T, or S in Column 1, a message is printed and that card is ignored. It is therefore possible to insert any number of blank cards or commend cards in the deck without effecting the index compilation. The reference set number may be numeric or blank.

APPENDIX A

Standard Stop Word List[1]

A
ABOUT
ABOVE
ACROSS
AFTER
AGAIN
AGAINST
AGO
ALL
ALMOST
ALONE
ALONG
ALREADY
ALSO
ALWAYS
AM
AMONG
AN
AND
ANOTHER
ANY
ANYTHING
APPEAR *
ARE
AROUND
AS
ASK *
AT
AWAY
BACK *
BE
BECAME
BECAUSE
BECOME *
BEEN
BEFORE
BEGAN
BEGIN *
BEING *
BELIEVE*
BEST
BETTER
BETWEEN
BIG
BOTH
BRING *
BROUGHT
BUILD *
BUILT
BUT
BY
CALL *
CAME
CAN
CARRY
CASE *
CAUSE *
CHANCE *
CHANGE *
CLOSE *
COME *
COMING
CONSIDER
CONTINUE
COULD
CUT *
DAY *
DEAR *
DEMAND *
DID
DIE *
DIFFEREN
DO
DOES
DONE
DON≠T
DONT
DOWN
DROP *
DURING
EACH
EARLY
EITHER
END *
ENOUGH
EVEN
EVENING*
EVER
EVERY
EVERYTHI
FAR
FEEL *
FEET
FELT
FEW
FILL *
FIND *
FINE *
FIVE
FOLLOW *
FOOD
FOR
FOUND
FOUR
FROM
FRONT
FULL
GAVE
GET *
GIVE *
GIVEN
GO *
GONE
GOOD
GOT
GREAT
HAD
HALF
HAND *
HAPPEN *
HARD
HAS
HAVE
HE
HEAD *
HELD
HER
HERE
HERSELF
HIM
HIMSELF
HIS
HOLD *
HOME *
HOW
HOWEVER
I
IDEA *
IF
ILL
IMPORTAN
IN
INCREASE
INTO
IS
IT
ITS
ITSELF
KEEP *
KEPT
KIND *
KNEW
KNOW *
LARGE
LAST *
LATE
LEFT
LESS
LET *
LIKE *
LINE *
LITTLE
LIVE *
LONG *
LOOK *
LOST
LOW
MADE
MAKE *
MANY
MAY
ME
MEAN *
MIGHT
MISS
MOMENT *
MORE
MOST
MR
MRS
MUCH
MY
MUST
MYSELF
NAME *
NEAR *
NEED *
NEVER
NEW *
NEXT
NO
NOR
NOT
NOTHING
NOW
OF
OFF
OFTEN
ON
ONCE
ONE
ONLY
OPEN *
OR
OTHER *
OUR *
OUT
OVER
OWN *
PART *
PASS *
PERHAPS
PERSON *
PLACE *
PLAN *
POINT *
POSSIBLE
PRESENT*
PRICE *
PRODUCE*
PUT *
QUITE
RATHER
REACH *
READ *
READY
REAL
RECEIVE*
REMAIN *
REST *
RESULT *
RETURN *
RIGHT *
ROUND
RUN *
SAID
SAME
SAT
SAW
SAY *
SEE *
SEEM *
SEEN
SERVE *
SERVE *
SET *
SEVERAL
SHALL
SHE
SHORT *
SHOULD
SHOW *
SIDE *
SINCE
SMALL
SO
SOME
SOMETHIN
SOON
STAND *
START *
STEP *
STILL
STOOD
STOP *
SUCH
SUPPOSE*
SURE
TAKE *
TAKEN
TELL *
TEN
THAN
THE
THEIR
THEM
THEMSELV
THEN
THERE
THESE
THEY
THING *
THIS
THOSE
THOUGH
THOUSAND
THREE
THROUGH
THUS
TIME
TO
TODAY
TOGETHER
TOLD
TOO
TOWARD *
TRIED
TRY *
TURN *
TWENTY
TWO
UNDER
UNTIL
UP
UPON
US *
VERY
WAIT *
WALK *
WALL *
WANT *
WAS
WAY *
WE
WELL
WENT
WERE
WHAT
WHEN
WHERE
WHETHER
WHICH
WHILE
WHO
WHOLE
WHOM
WHOSE
WHY
WISH *
WITH
WITHIN
WITHOUT
WONDER
WORD *
WOULD
WRITE *
YES
YET
YOU
YOUR

---

[1]Where an asterisk appear in character position eight of a word, that word plus any of the suffixes shown in section 3.1 will be a stop word.

APPENDIX B

Keypunch Symbols Arranged in the Standard Collating Sequence

blank	.
0	)
1	- (minus)
2	J
3	K
4	L
5	M
6	N
7	O
8	P
9	Q
=	R
≠ or '	$
%	*
+	/
A	S
B	T
C	U
D	V
E	W
F	X
G	Y
H	Z
I	'
	(

Note: The colon is converted to the digit 9 by BIBCON.

APPENDIX C

Encoding Scheme for OE 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).

$=n_1,\ldots,n_m$	where n_i is the letter position of an expanded letter in the resulting form.

For example, am̃ 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)

$n_1,\ldots,n_m$	where $n_1,\ldots,n_m$ is as described in Expansions.

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

$=n_1,\ldots,n_m$	where $n_1,\ldots,n_m$ is as described in Expansions.

Examples

OE	:	BVTAN	tweon	lar
Punched:		BVTAN)A	TWEON	LAR

OE	:	lufan	ge	wanig̃ 7	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

APPENDIX D

Computing Center Control Cards Needed for Running on the UWCC 3600[1]

All Control Cards in this section require a 7 and a 9 punch in Col. 1. The other information must begin in Col. 2.

Card 1 Job Card

JOB, $n_1n_2n_3n_4$, $p_1p_2p_3p_4$/name, time

$n_1n_2n_3n_4$ is the project number assigned by the UWCC; in other words, the source of funds for running the concordance.

$p_1p_2p_3p_4$ is the user's UWCC number, also assigned by the UWCC.

name is the last name of the user, or any other name that does not contain a comma.

time time is the time limit in minutes for the job. If the job is not finished by the time it has run the specified limit, it will be terminated at whatever stage it is at. Since the program must be rerun from the beginning, this time will be lost.

It is difficult to estimate accurately the amount of time required for a concordance. This is a function of the size of the logical records, the number of logical

---

[1] As of July 10, 1968. UWCC stands for University of Wisconsin Computing Center.

records, the relative sizes of the fields within a logical record, the number of keywords per logical record, and other options selected by the user. The following table is intended as a guide and not as an exact estimate. It is better to specify too much time than too little.

For logical records of 80 characters; 125 stop words and no Notes field.

No. of records	Run time
250	2 min. 30 sec.
5000	33 minutes

Card 2 Equip card for BIBCON program

EQUIP, 9 = (BIBCON 1.1), HI

Card 3 Tape card for BIBCON program

TAPE, LUN 9 IS UWCC S8137

If input is from tape, rather than cards, another EQUIP and another TAPE card must be provided. See the UWCC consultant for these cards. Likewise, if output is to be on magnetic tape, or if the keyword-frequency list is to be saved, additional EQUIP and TAPE cards must be provided.

Card 4 Load card

LOAD, 9

Card 5 Run card

RUN, time, print

time should be the same as time on Card 1. This is the maximum time in minutes that you will allow the program to run and is identical to time on the Job Card (Card 1).

print this is the maximum number of print lines that you will allow. To be safe, punch 999999 here.

The BIBCON control cards follow immediately after the UWCC control cards. If data input is from cards, then the data cards follow the BIBCON control cards. If data input is from tape, the last BIBCON control card is the last card of the deck.

figure 11 here

```
'JOB,9999,1234/TCHAIKOVSKY,5
'EQUIP,9=(BIBCON 1.1),HI
'TAPE, LUN 9 IS UWCC TAPE S8134
'EQUIP,10=(WHITMAN),HI
'TAPE, LUN 10 IS PRIVATE TAPE TCH-6
'EQUIP,11=(WHIT-CONC,,,999),HI
'TAPE,LUN 11 IS PRIVATE TAPE TCH-8 RING IN BLANK LABEL
'EQUIP,7=(WW-WORDS,,,999),HI
'TAPE,LUN 7 IS PRIVATE TAPE ABBA-H RING IN BLANK LABEL
'LOAD,9
'RUN,5,999999
      I      S
    160     -2       1       9     11    160
     60    120      16                0      0     11                    10
      1      1       0
A CONCORDANCE TO WALT WHITMAN S, TO A COMMON PROSTITUTE
WALT WHITMAN--TO A COMMON PROSTITUTE
      0      0       1
********
```

Figure 11

UWCC and BIBCON control cards for the Whitman concordance shown in section 4.9. Input is on unit 10, (TCH-6 is the external label for this tape), output is to be on unit 11 (external label "WHIT-CONC" will be written on the tape during output), and the key words will be saved on unit 7 ("WW-WORDS" will be written as a label). The single quote mark in column 1 is the 405 listing for a 7, 9 punch.

APPENDIX E

Tape Formats

Keywords

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

Records are written at 556 bpi in even parity, 999 words (8 character words) per physical record, terminated by an eof mark. Unused words in the last physical record are filled with zeroes. Logical records are three words in length. The first two contain a keyword, left justified with blank fill; the third contains a frequency number in binary.

Concordance listing

If the concordance output is equated to any unit except 61, it will be blocked in 450 word physical records. Each logical record is 120 characters in length, the last character being an end of record character (32_8). Tapes are written in even parity, 556 bpi. File one will contain the concordance listing, followed by a single eof mark. File 2 will contain the keywords, unblocked, followed by an eof mark.