DOCUMENT RESUME

ED 044 121 LI 002 115

AUTHOR	Standera, Oldrich
TITLE	COMPENDEX Profiling Guide.
INSTITUTION	Information Retrieval Research Council of Alberta, Edmonton.
PUB DATE	Jun 70
NOTE	41p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$2.15
DESCRIPTORS	*Information Science, *Information Services, *Information Systems, Manuals
IDENTIFIERS	Canada, CIS, COMPENDEX, *Computerized Engineering Index, Current Information Selection

ABSTRACT

This manual provides instructions for completing the COMPENDEX (Computerized Engineering Index) Profile Submission Form used to prepare Current Information Selection (CIS) profiles. An annotated bibliography lists nine items useful in searching for proper profile words. (AB)

ALBERTA INFORMATION RETRIEVAL ASSOCIATION

COMPENDEX PROFILING GUIDE

by

OLDRICH STANDERA

INFORMATION SYSTEMS
THE UNIVERSITY OF CALGARY

Issued by

Information Retrieval
Research Council of Alberta
Edmonton 7, Alberta
Canada

June, 1970

CONTENTS

ILLUSTRATIONS

PURPOSE OF THIS MANUAL

The purpose of this manual is to give instructions for successfully completing the COMPENDEX Profile Submission Form.

This manual is primarily intended for search editors, although some users may do their own search editing.

In our opinion no manual can be an exhaustive answer to all questions. But those readers who will study it from the beginning to the end will be able to prepare Current Information Selection (CIS) profiles and so gather experience in solving practical problems encountered.

To facilitate the study of this manual we have included a series of examples and a completed COMPENDEX Profile Submission Form.

I GENERAL INFORMATION ABOUT COMPENDEX

COMPENDEX: COMPENDEX is the computerized version of the Engineering Index Monthly (starting with January 1969 issue), the leading source of abstracts on the world's engineering developments.

Among the 6000 abstracts of The Engineering Index Monthly there are answers to many engineering problems. More than 3,500 sources of engineering literature are read and their important contents are abstracted and indexed. Coverage includes journals, publications of engineering organizations, papers from conferences and symposiums, books, government reports and patents. This information is put on tapes to be disseminated to users according to their interest profiles. Engineering Index Inc., 345 East 47th. Street, New York 10017 supplies the tapes.

TEXT-PAC: COMPENDEX tapes are processed using the TEXT-PAC system. Dr. S. Kaufman wrote the TEXT-PAC programs and the documentation of the system (See Annotated Bibliography).

THE UNIVERSITY OF CALGARY, INFORMATION SYSTEMS, will run your profiles against the Engineering Index tapes. We do search editing work for The University of Calgary campus only.

For other users, the COMPENDEX Current Information Service is available through the Alberta Information Retrieval Association (AIRA) at the following address:

AIRA,
INFORMATION RETRIEVAL,
RESEARCH COUNCIL OF ALBERTA,
EDMONTON 7, ALBERTA.

AIRA will accept questions from search editors who have familiarized themselves with this manual. In many cases, a search editor would be a technical librarian. If an organization or firm does not have a library and/or a search editor, an individual may act as his own search editor and contact AIRA directly until a more efficient arrangement can be devised.

The flow of information in the system is shown in Fig. 1.

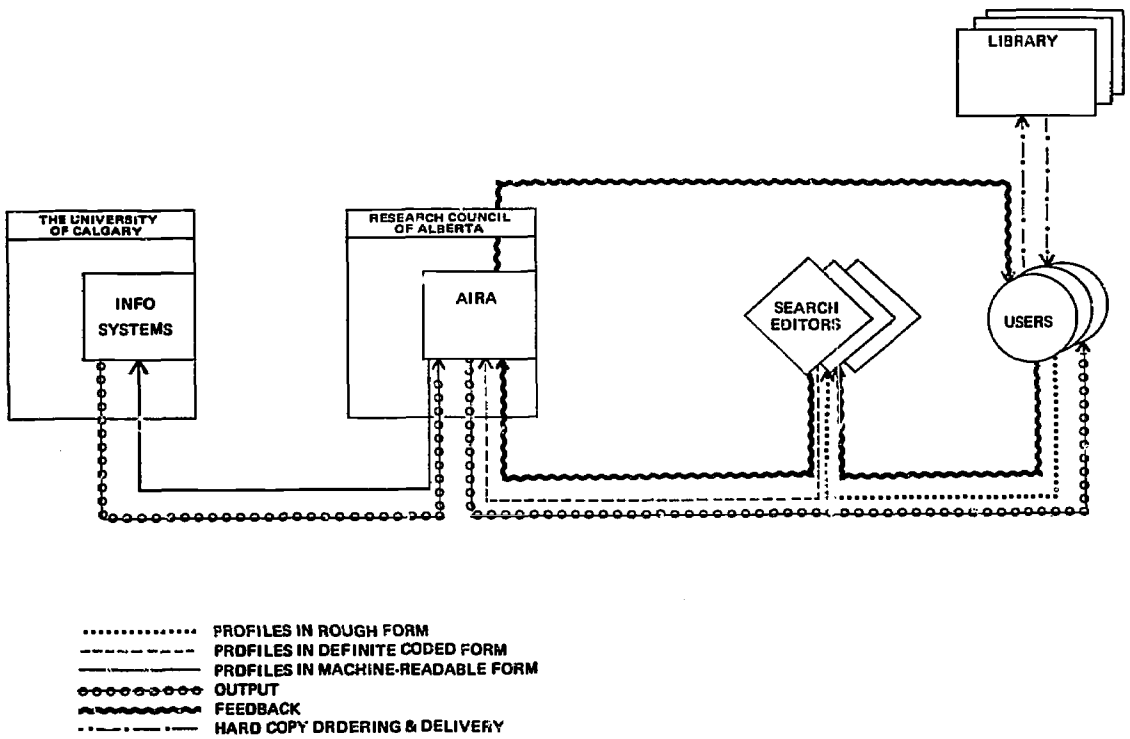


Fig. 1 – Compendex Flow

COST: The cost of COMPENDEX current awareness searches will be $100.00 per year for 40 terms as of July 1st, 1970. Each additional ten terms would be $20.00 per year. Charges would be based on the average number of terms used throughout the year. Most subscribers are able to prepare highly effective profiles within four months.

We have observed that users who put the most effort into constructing and maintaining their profiles derive superior results and are most satisfied with the system.

FEEDBACK: Feedback from the user is of the utmost importance for successfully running any SDI (Selective Dissemination of Information) service.

By means of this communication link the service centre is kept posted of the overall relevance of the information which is disseminated to users and may take corrective steps if it becomes apparent that the user is not getting what he wants.

Fig. 2 shows a sample double-card.

70X 51497 30/05/70

INTERACTIVE SEARCH STRATEGIES AND DYNAMIC FILE ORGANIZATION IN INFORMATION RETRIEVAL
EIX70X051497
IDE E
SALTON G
CORNELL UNIV-DEPT OF COMPUTER SCIENCE-SCI REPORT ISR-16 SEPT 1969 SEC XI, 34 P
SEVERAL SEARCH AND RETRIEVAL STRATEGIES ARE DESCRIBED THAT USE FEEDBACK INFORMATION SUPPLIED BY THE USER DURING THE RETRIEVAL PROCESS TO MODIFY THE QUERY OR DOCUMENT SPACES. IN EACH CASE, THE SPACE MODIFICATION IS INTENDED TO INCREASE THE CORRELATION BETWEEN QUERIES AND RELEVANT DOCUMENTS, WHILE DECREASING THE QUERY CORRELATION WITH NONRELEVANT ITEMS. EXPERIMENTAL EVIDENCE INDICATES THAT THE IMPROVEMENTS IN RETRIEVAL EFFECTIVENESS OBTAINABLE WITH THESE HEURISTIC SEARCH STRATEGIES ARE MUCH LARGER THAN THE IMPROVEMENTS IMMEDIATELY DERIVABLE FROM THE MORE FORMAL DETERMINISTIC METHODS BASED ON BETTER DOCUMENT AND QUERY ANALYSES AND MORE SOPHISTICATED LINGUISTIC NORMALIZATION TOOLS. 19 REFS. 21620

THE UNIVERSITY OF CALGARY—INFORMATION SYSTEMS

IBM E12324 PRINTED IN U.S.A.

Left portion

OR STANDERA CALG UNI INFOSYS 1[illegible] 0022141
21620 70X 51497

0022141
CARD NO.

INSTRUCTIONS:
1. READ ABSTRACT.
2. PUSH OUT THE APPROPRIATE BOX WITH A SHARP PENCIL.
3. RETURN THIS HALF TO SPONSORING AGENCY.

YOUR COOPERATION WILL IMPROVE OUR SERVICE

ABSTRACT RELEVANT
ABSTRACT IRRELEVANT
DOCUMENT WANTED
DOCUMENT NOT WANTED
COMMENTS, QUESTIONS, ADDRESS CHANGE
(USE REVERSE SIDE)

THE UNIVERSITY OF CALGARY — INFORMATION SYSTEMS

IBM E12324

Right portion

Fig. 2 — Sample Double-Card Output

The feedback in the COMPENDEX service does not require much effort on the part of the user. All he has to do is read the abstract, judge its relevance and push out the appropriate box with a sharp pencil.

The abstract may be either relevant or irrelevant. If relevant it remains to be decided whether the document is wanted or not wanted. Comments, questions or address change may be given on the reverse side.

The left-hand portion is designed for your file, but the right-hand portion should be forwarded without delay to the search editor and through him to AIRA.

HARD COPY: Note that neither The University of Calgary nor the Alberta Information Retrieval Association are able to provide you with original documents or copies. Your nearest library should provide you with the most effective service in this regard.

RETROSPECTIVE SEARCHING: The University of Calgary, Information Systems, is also experimenting in retrospective searches using TEXT-PAC. The COMPENDEX tapes began in January 1969. Therefore, any retrospective searches would be run from that date.

II HOW TO PREPARE A PROFILE SUBMISSION FORM

The Compendex Profile Submission Form is simple and easy to use. In the following text we will deal with each individual section of this form.

Please refer to Fig. 3, which shows a completed example.

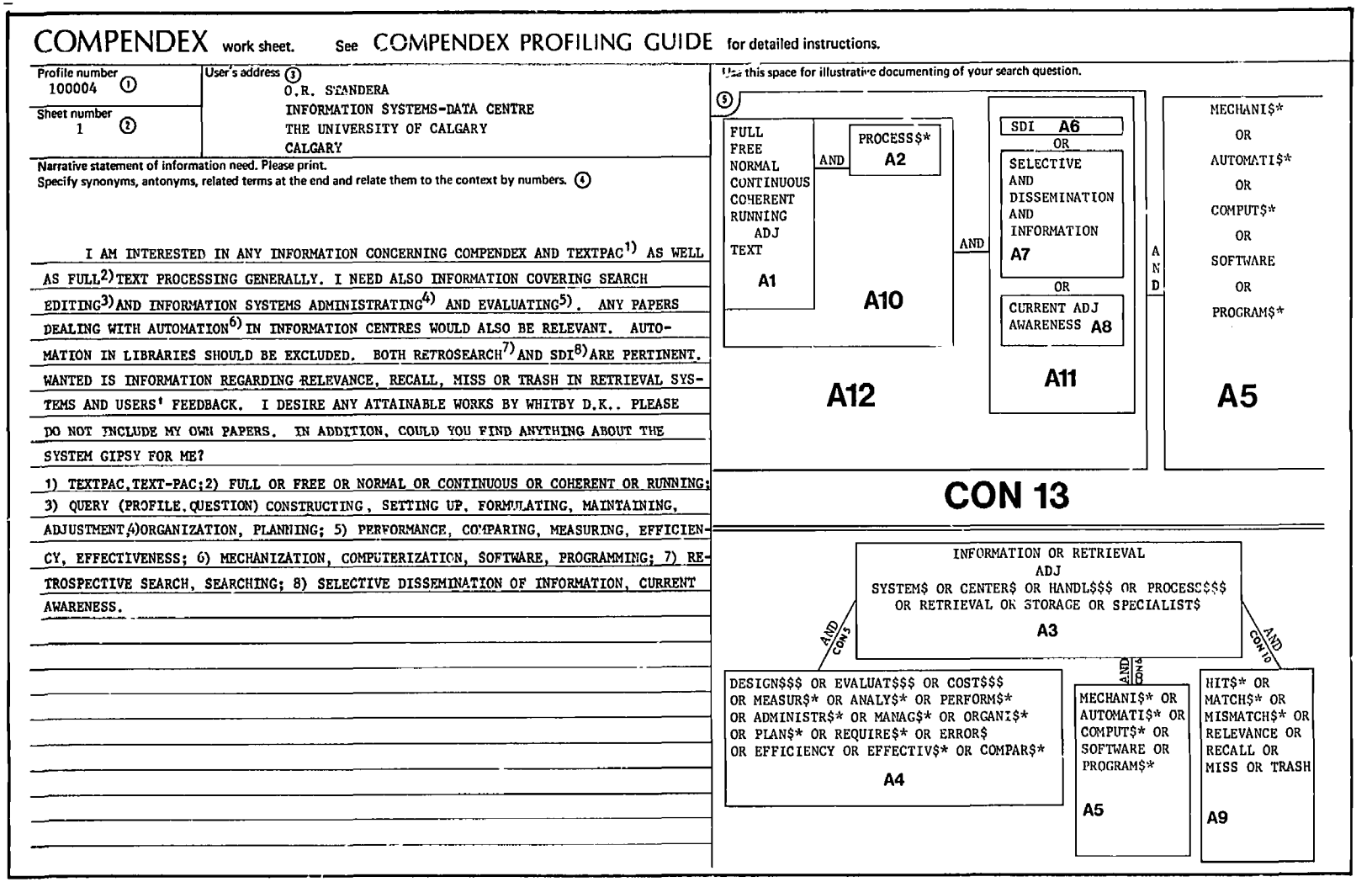


Fig. 3 - COMPENDEX Profile Submission Form

(7a)

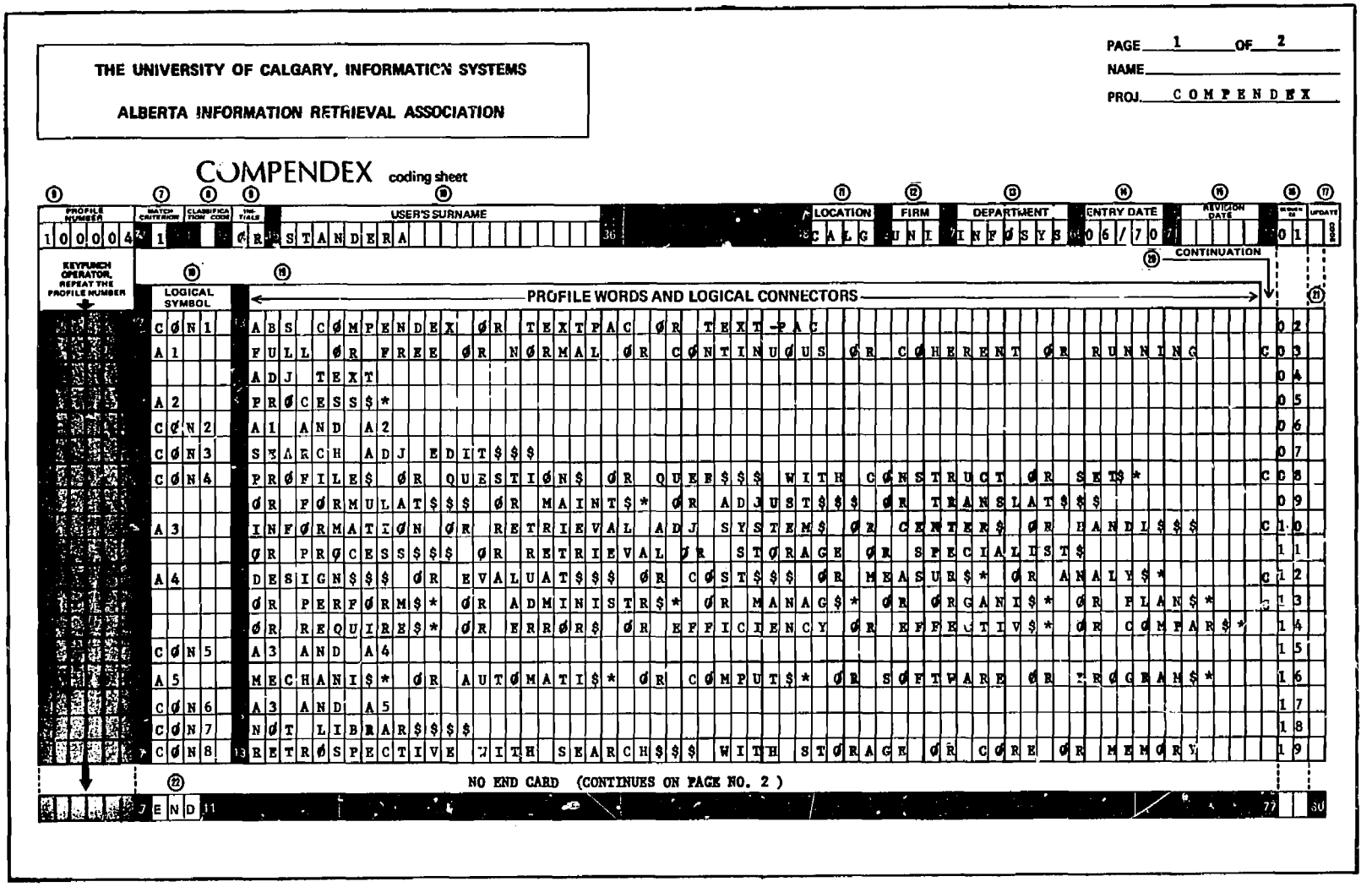


Fig. 3 - COMPENDEX Profile Submission Form

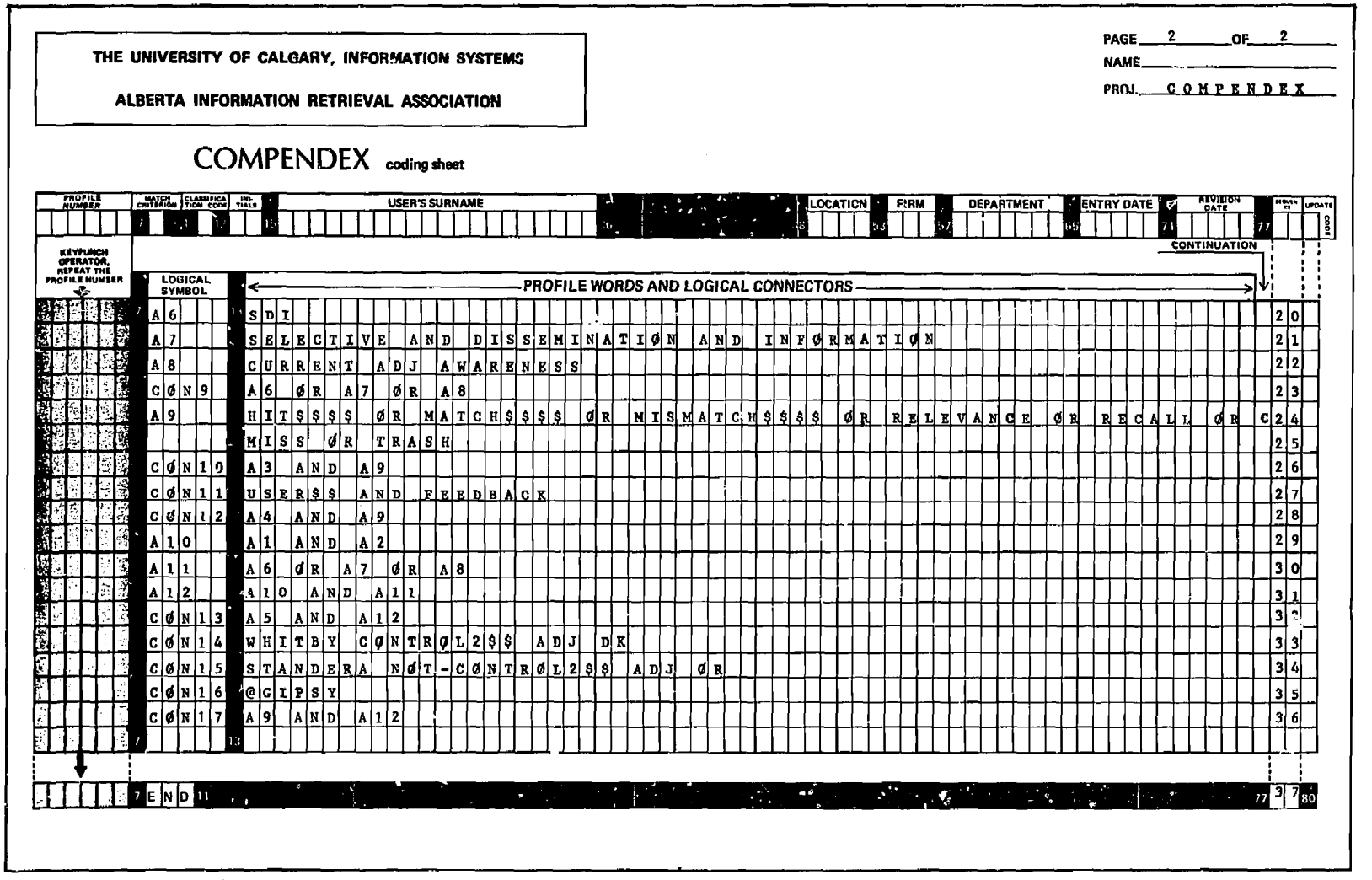


Fig. 3 - continuation

The front side of a Profile Submission Form is a work sheet; the reverse side is a coding sheet. Please note that the encircled numbers after each field name provide a direct link to the Submission Form.

(1) WORK SHEET (FRONT SIDE OF PROFILE SUBMISSION FORM)

Profile Number ①

You will be assigned a profile number. If you already know your profile number (e.g. if you are making adjustments to your existing profile), fill it in here. Also refer to your profile number in your correspondence.

Sheet Number ②

If you need more space than is available on one sheet, use as many sheets as needed and number them consecutively.

User's Address ③

Give in full the address at which you wish to receive your search results.

Narrative Statement ④

In this space you should state your search request in narrative form and specify all the profile words you would like to have included in the search expressions. The proper and exhaustive statement of your search request will help in finding the proper words.

Pay due attention to filling out this space. Even the best formulating of your search expressions will not make up for any omissions at this stage.

Be sure to include all words characterizing your information need and interest. Add all synonyms (e.g. full or free or normal etc, text), related terms (relevance, recall), antonyms if applicable (matching, mismatching), acronyms (SDI - Selective Dissemination of Information). Relate synonyms, acronyms, antonyms and related terms to the context by numbers. An example is given to illustrate the proper way of doing this. See Fig. 3, Section ④

In preparing a profile the following aids may prove useful:

1. Your own articles on the topic,

2. Other authors' works which you consider relevant to your information needs,

3. Subject indexes of books on this theme,

4. Handbooks and thesauri of synonyms and antonyms,

5. The Dictionary of COMPENDEX terms serves primarily to check all new words for correct spelling, but is a valuable help for IR specialists in formulating a question. The same holds true of the Statistical lists of all used words. Print-outs would involve an additional charge of $25 for the latest issue printed.

6. For additional literature see the Annotated Bibliography.

Do not be scared by the rather involved profile we are following throughout this manual. We have established it to show you all the facilities in one example. Most of your profiles will have a simple structure. We recommend a straightforward simple structure as it is easy to establish and maintain. Rather than one complicated concept use two simpler ones. The same applies to search expressions. After some time you will find it easy to set up profiles of any degree of sophistication shown below. A few examples of simplified profiles are:

(1) Narrative statement: I need information pertaining to synthetic (plastic) foam, as far as it is related to the manufacture. Also properties of synthetic foam are of interest.

Profile

A1 SYNTHETIC OR PLASTIC
A2 FOAM$
A3 A1 WITH A2

A4 PROPERT$$$ OR CHARACTERISTIC$ OR MANUFACTUR$$$ OR PRODUC$$$$
CON1 A3 AND A4

Explanation

Dollar signs e.g. in PRODUC$$$$ mean that this formulation covers "PRODUCTION", "PRODUCE", "PRODUCER", etc. "FOAM$" covers both "FOAM" and "FOAMS". A3 connects "FOAM$" with either "SYNTHETIC" or "PLASTIC". "WITH" implies occurrence of both A1 and A2 in the same sentence. CON1 links A3 and any of the terms under A4. Terms or symbols connected by "AND" must occur in the same record to produce a hit.

(2) Narrative statement: The same as under 1.

Profile

A1 SYNTHETIC ADJ FOAM$
A2 PLASTIC ADJ FOAM$
A3 A1 OR A2
A4 PROPERT$$$ OR CHARACTERISTIC$ OR MANUFACTUR$$$ OR PRODUC$$$$
CON1 A3 AND A4

Explanation

"ADJ" in A1 requires that both "SYNTHETIC" and "FOAM$" be close together in the order shown, to produce a hit. A3 indicates that either A1 or A2 are acceptable.

CON1 states that A3 and A4 may occur at any place in the same record to meet the information need. Only one type of logical connector is used in any one concept.

(3) Narrative statement: The same as under 1 and 2.

Profile

A1 SYNTHETIC OR PLASTIC ADJ FOAM$
A2 PROPERT$$$ OR CHARACTERISTIC$ OR MANUFACTUR$$$ OR PRODUC$$$$
CON1 A1 AND A2

Explanation

This is the concise way of setting up a profile from the statement given.

"OR" logical connector may be used with "ADJ" or "WITH" in the way shown in A1. (In the search expression CON1 you may use only logical connector "AND" between A1 and A2. "WITH" and "ADJ" could be used if A1 and A2 contained words connected by "OR").

(4) Narrative statement: the same as under 1-3 but we do not wish to receive the information as far as marketing is concerned (and some other related terms).

Profile

A1 SYNTHETIC OR PLASTIC ADJ FOAM$

A2 PROPERT$$$ OR CHARACTERISTIC$ OR MANUFACTUR$$$ OR PRODUC$$$$

CON1 A1 AND A2

CON2 NOT MARKET$$$ OR SALE$ OR BUY$$$ OR CONSUM$*

Explanation

CON2 contains "NOT" which excludes all documents dealing with MARKET$$$ as well as other terms specified. These documents will not be matched by the profile.

(5) Narrative statement: the same as above. We request, however, any information regarding polyurethane(s).

Profile

A1 SYNTHETIC OR PLASTIC ADJ FOAM$

A2 PROPERT$$$ OR CHARACTERISTIC$ OR MANUFACTUR$$$ OR PRODUC$$$$

CON1 A1 AND A2

CON2 NOT MARKET$$$

CON3 ABS POLYURETHANE$

Explanation

CON3 contains "ABS" logic. This means that any document dealing with "POLYURETHANE($)" will be picked out for the user. It overrides any other logic used.

After this basic introduction to profile constructing you will find the following more detailed explanation easier.

Length of words, spacing: the length of any profile word may be up to 38 characters. However only the first 20 characters are searched.

Leave a space after every word and after every logical connector. Two words must always be separated by any of the logical connectors.

Truncation: it is sometimes desirable to search on word stems rather than on the full words.

TEXT-PAC allows right end truncation only. Truncation can be done in two ways:

Selective truncation may extend as far as six characters past the root. ORGANI$$$$$$ will cover ORGANIZE, ORGANIZER, ORGANIZERS, as well as ORGANIZING, ORGANIZATION. As we may use only six dollar signs, we have to use unconditional truncation if we want also ORGANIZATIONAL to be included in our profile formulation (See A4, Fig. 3).

Unconditional truncation ORGANI$* will cover all possible endings of the given root as far as twenty characters. The root may consist of a minimum of one character.

When using this profiling facility always carefully consider all possible words that you might match. You might save several seconds by indiscriminate truncation but loose a considerable period of time getting through irrelevant information produced.

For example if you are interested in programming and programs of retrieval systems. Specifying PROGR$* would find not only desired programming and programs, but also unwanted progress, progression, etc.

Illustrative Documenting of Search Question ⑤

Place synonyms etc. in boxes and then link these boxes together with logical connectors as shown in Section ⑤ of Fig. 3. If you formulate your search request

in this way, you will find it easy to complete the item 19 on the reverse side which is the main part of this form.

Terminology: Profile words (terms) are connected to each other by means of logical connectors forming the concepts which constitute search expressions. One or more search expressions form a question (profile).

Profile words (terms), concepts or search expressions may be represented by logical symbols. Notice that the search expressions are denoted by CON in the COMPENDEX Profile Submission Forms. (The original TEXT-PAC documentation uses "concepts" where we introduced "search expression").

The following example is designed to clear up terminology:

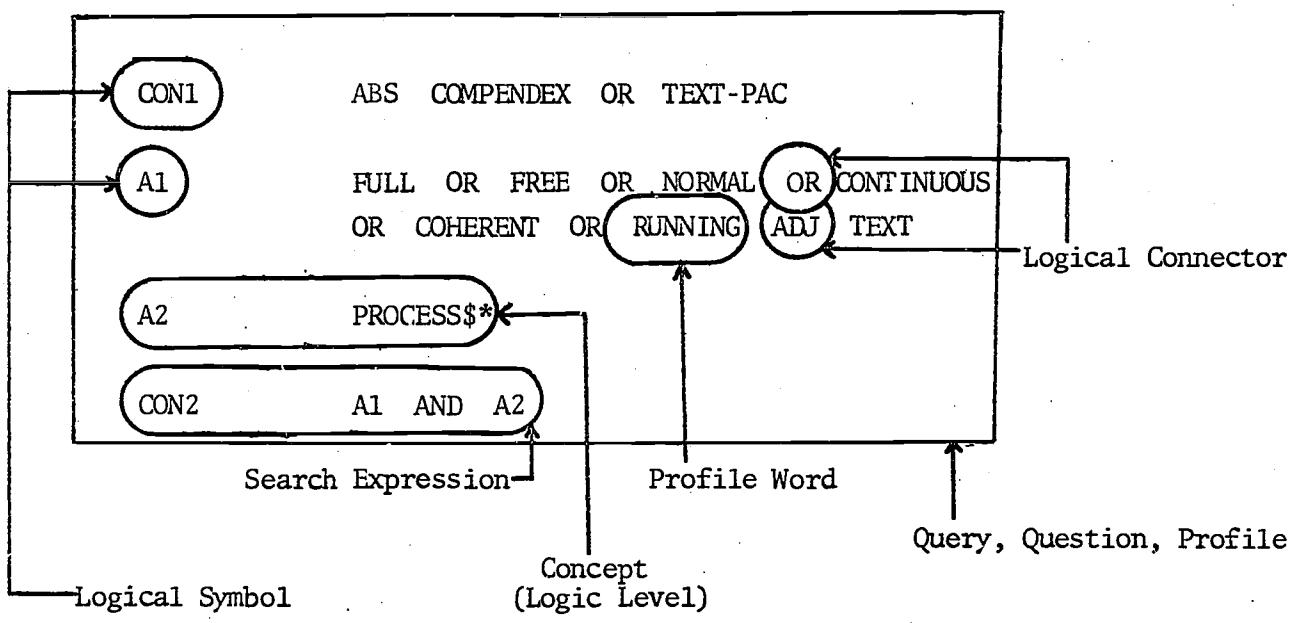


Fig. 4 - Query terminology

Please note that we are using the following logical connectors:

OR
AND
WITH
ADJ
NOT
ABSOLUTE
CONTROL
NOT-CONTROL

OR: Logical connector OR combines profile words or logical symbols indicating that any of them will satisfy the user's requirement. In our example we are interested in "text" which may be specified as:

FULL OR FREE OR NORMAL OR CONTINUOUS OR COHERENT OR RUNNING.....

See A1. Fig. 3.

AND: Logical connector AND identifies the profile words or concepts which must jointly be present in a data base record for the hit to occur. A maximum of 15 profile words may be connected by AND.

For example, "USERS' AND FEEDBACK" means that the hit will only result if both of these words occur in the same document. It is evident that we might get some irrelevant hits if one sentence dealt generally with "USERS' REACTION" and another described "FEEDBACK" in electronics.

ADJ, WITH: Two profile words or concepts linked by ADJ must occur in the order specified to bring about a match.

SEARCH ADJ EDIT$$$

The logical connector WITH will cause a hit if the connected profile words or concepts are found in the same sentence of the document.

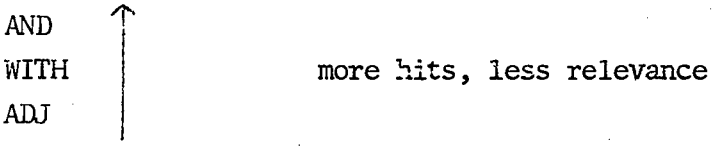
RETROSPECTIVE WITH SEARCH$$$ WITH STORAGE OR CORE.....

will produce a hit in any of the following contexts:

STORAGE REQUIRED BY RETROSPECTIVE SEARCHING.....
RETROSPECTIVE SEARCH NEEDS MORE STORAGE THAN.....
STORAGE CONSIDERATIONS FOR CURRENT AWARENESS, RETROSPECTIVE SEARCHING.....

Concerning the use of ADJ or WITH jointly with CONTROL and NOT-CONTROL logic, see the paragraph on CONTROL.

After you have formulated a few profiles in TEXT-PAC system, you will appreciate the way you can make your concepts and search expressions broader or narrower, thus obtaining more or less hits.



This arrow shows the direction of obtaining more hits, although you may get more irrelevant information too at the same time.

Remember two rules for proper use of ADJ or WITH:

(1) Only one type of logical connector may occur in a concept or search expression. There is one exception: you can use OR logic inside ADJ or WITH logic provided you connect profile words and not logical symbols denoting concepts (See A3; CON4).

RIGHT:

INFORMATION OR RETRIEVAL ADJ SYSTEM$ OR CENTER$.....
PROFILE$ OR QUESTION$ OR QUER$$$ WITH CONSTRUCT$$$ OR SET$$$$.....

WRONG:

A1 OR A2 WITH A13

(2) Using ADJ or WITH logical connectors to connect two or more logical symbols which denote concepts, always make sure that the logical symbols cited represent words joined by OR logic (Another formulation of the above example):

RIGHT:

A1 INFORMATION OR RETRIEVAL
A2 SYSTEM$ OR CENTER$ OR.....
A3 A1 WITH A2
ALSO: A3 A1 ADJ A2

WRONG:

A1 INFORMATION ADJ RETRIEVAL
A2 SYSTEM$ OR CENTER$
A3 PROFILE$ OR QUESTION$ OR QUER$$$ WITH CONSTRUCT$$$ OR SET$*
CON1 A1 WITH A2 WITH A3

The proper way to formulate a search expression such as this would be
CONL A1 AND A2 AND A3

ABS: The logical connector for the ABSOLUTE logic is identified as ABS. When using ABS the hit will result with occurrence of any word accompanied by ABS in any context whatsoever. Remember that ABS may be used only in search expressions (not in concepts) and must be the first word of logic data. In this case any document containing the profile words COMPENDEX or TEXT-PAC will be quoted as a hit regardless of all the other logic.

CON1 ABS COMPENDEX OR TEXT-PAC

NOT: The NOT logical connector denotes the profile words which we do not wish to cause a hit. It overrides any other logical connector except ABS. This implies that if a given document contains a profile word which was denoted by NOT and another profile word specified by ABS, this document will become a hit. Keep in mind that you can only use NOT in search expressions and it must be the first word of logic data, e.g. the user wants all the information specified but he has enough information dealing with "libraries" already available and desires it to be excluded:

CON 7 NOT LIBRAR$$$$

GENERAL REMARKS ON QUESTION FORMULATION: It should be noted that matching profile against data base is done against the search expressions.

When constructing your profile, remember to include terms which are synonymous or closely related to your basic terms. Then formulate as many search expressions as needed to cover your information request.

Label the concepts with logical symbols A1, A2, A3 and so on. Label the

search expressions CON 1, CON 2, CON 3.....

Any concept may contain either logical symbols or words but not both together.

e.g. CON 1 A1 AND A2
CON 9 A6 OR A7 OR A8
A8 CURRENT ADJ AWARENESS

WRONG:

A21 ECONOMICS AND A20

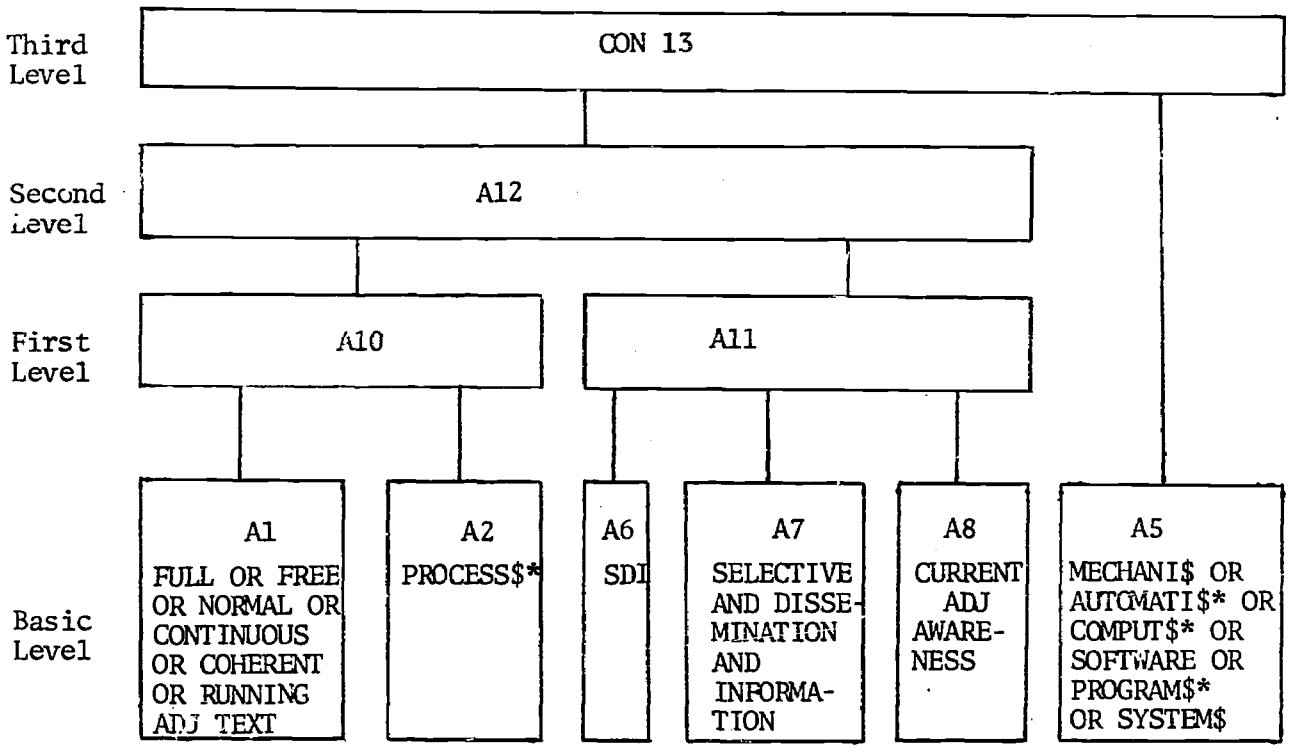
The same rule applies to search expressions. Remember you may use only one type of logical connector in any one logic level, (concept or search expression). The only exception is mentioned in the section dealing with ADJ and WITH.

Levels of back-referencing: when constructing your search expressions you may use the concepts and search expressions in three levels. This is an excellent feature of TEXT-PAC and the Fig. 5 will clear up the principles involved.

You will notice on the following figure that you may reference e.g.:

(1) the search expression back to A12
(2) A12 back to A10 and A11
(3) A10 back to A1 and A2

More levels of referencing will cause an error reported by an error message of the computer.



A1 FULL OR FREE OR NORMAL OR CONTINUOUS OR COHERENT OR RUNNING ADJ TEXT

A2 PROCESS$*

A5 MECHANI$* OR AUTOMATI$* OR COMPUT$* OR SOFTWARE OR PROGRAM$* OR SYSTEM$

A6 SDI

A7 SELECTIVE AND DISSEMINATION AND INFORMATION

A8 CURRENT ADJ AWARENESS

A10 A1 AND A2

A11 A6 OR A7 OR A8

A12 A10 AND A11

CON 13 A5 AND A12

Fig. 5 - Levels of back-referencing

There are two more rules concerning back-referencing in the three level structure:

(1) You may reference back to profile words or to concepts (A1, A2, A3.....) but not to search expressions!

(2) Any logical symbol (expressing a word or a concept) may be referenced a maximum of fifteen times.

Remember also that you must not specify more than fifteen logical symbols in any one concept. If more than fifteen should be connected, establish a new concept! You cannot use more than ten cards (= lines in the profiling form) to specify any one concept.

It is not permitted to back-reference a logical symbol to another one standing alone, but it is allowed to identify a search expression by one logical symbol:

RIGHT: CON 14 A6

WRONG: A13 A6

(2) CODING SHEET (REVERSE SIDE OF PROFILE SUBMISSION FORM)

FIRST CARD

Profile Number ⑥

The user is not concerned with this field. The search editor will place the profile number here and the keypunch operator will continue all the profile down.

Match Criterion ⑦

Fill the match criterion in the card column 8. Although it may be anything between 1 and 9 you will almost exclusively use 1. It implies that any one of the search expressions specified may cause a hit.

Classification Code ⑧

Leave this field (card column 11) blank. As Compendex does not contain

confidential files, all files may be searched for you and that is specified by a blank.

User's Initials (9)

These should be placed in columns 13 and 14.

User's Surname (10)

This should be placed in column 16 through 35.

Location (11)

(Card Columns 49-52) Contains the first four letters of your location (e.g. CALG for Calgary, OTTA for Ottawa).

Firm (12)

Firm comes next in the card columns 54-56. It denotes the firm, institute, or whatever the organization of the user may be. Use either first characters (XER for XEROX) or initial characters if composed of more than one word (UOC for The University of Calgary).

Department (13)

Department is self-explanatory; it allows you to fill in the department to which you belong (e.g. MECHENG if you are a member of the department of Mechanical Engineering). Use columns 58-64 for this purpose.

Entry Date (14) Enter date in this format MM/YY (e.g. 01/70) in columns 66 - 70.

Revision Date (15) Complete only if appropriate. The format is the same as for "Entry Date". (Columns 72 - 76).

Card Sequence Number (16) (Columns 78 - 79). You fill 01 in these columns of the First Card.

Update Code ⑰ is placed in Column 80. Leave this column blank in a new profile.

You may change your profile by:

(1) simply punching a new card(s) and substituting it (them) for the existing one(s) in the profile card deck,

or

(2) you may indicate the change in the Update Code column of the First Card (and Logic Data Card(s)) and the program will take care of your change. (See also Update Code of Logic Data Cards).

"R" in the column 80 of the First Card means that:

(a) the profile is to be revised (if all fields are filled in),

(b) the whole profile is to be resequenced (if only the profile number and "R" are specified on the First Card).

If you wish to change the entire profile by a new version, simply substitute new profile card deck and the old one will be automatically erased.

LOGIC DATA CARDS

Logical Symbol ⑱

The field encompassing the columns 8 - 12 is reserved for logical symbols. Logical symbols identify concepts or search expressions.

A1, A2, A3, A4 etc. denote concepts:
CON1, CON2, CON3 etc. identify search expressions.

Logical Data ⑲

Logical Data is printed in columns 14 through 76. It is the main field in the COMPENDEX profiling form and it is composed of terms and logical connectors. The rules for completing it were given in previous sections.

Continuation Column ⑳

There are 63 columns reserved for logic data. Should it happen that it is not enough for your concept (A1, A2....) or search expression (CON1, CON2...), specify "C" in column 77 ("C" means continuation) and carry on your logical data on the next line. Be sure to leave the field ⑱ (logical symbol columns 8 - 12) blank on the next line.

Remember that each line must end and any new line must begin with a whole word. You must not divide words between lines (i.e. cards). There is a limit imposed on the number of cards pertaining to one logical symbol: never use more than 10 lines to define one logical symbol.

Card Sequence Field (columns 78 - 79) was already mentioned with the First Card (field ⑯). Here the purpose remains the same: the number you fill in here identifies a certain line (card) inside any profile (02, 03, 04 and so forth all the way down, including the End Card).

Update Code ㉑

Update Code is contained in the last column of any line. If you submit a new profile or if you make a change in your profile by changing a card in your profile cards deck, then this column remains blank.

If you wish to correct your profile on the tape by specifying an update code in this column, then you should proceed in the following manner:

(a) Deletion. To delete a line, you indicate the profile number (1-6), continuation (77) if applicable, card sequence number (78-79) and you print "D" in the update code column (80).

Note that neither logical symbol nor logical data are required for a line to be deleted.

(b) Revision. If you want to make a revision, you provide the new card with "R" in the update code column 80.

(c) Addition. If you want to add a line, print an "A" in the column 80. This line will be added before the sequence number you have indicated.

Let us take our profile as an example. For the sake of simplicity we will use only the two starting search expressions:

```
100004  CON1  ABS COMPENDEX OR TEXT-PAC ..................................  02
100004  A1    FULL OR FREE OR NORMAL OR CONTINUOUS OR COHERENT OR RUNNING....  C03
100004        ADJ TEXT.....................................................  04
100004  A2    PROCESS$*....................................................  05
100004  CON2  A1 AND A2 ...................................................  06
```

You may wish to:

(1) change the 02 line
(2) delete the 04 line
(3) add a new line before the 07 line.

Your profile adjustment will be coded this way:

```
100004  CON1  ABS COMPENDEX OR TEXT-PAC OR TEXTPAC ..........................  02R
100004  A1    FULL OR FREE OR NORMAL OR CONTINUOUS OR COHERENT OR RUNNING....  03R
100004        .............................................................  04D
100004  CON3  NOT COST$$$ .................................................  07A
```

Your new profile will look like this:

```
100004  CON1  ABS COMPENDEX OR TEXT-PAC OR TEXTPAC ..........................  02
100004  A1    FULL OR FREE OR NORMAL OR CONTINUOUS OR COHERENT OR RUNNING....  03
100004  A2    PROCESS$*....................................................  04
100004  CON2  A1 AND A2....................................................  05
100004  CON3  NOT COST$$$...................................................  06
```

END CARD (LAST CARD)

You will find completing this card very easy. The only objective it serves is to formally separate the profiles from each other.

END (22) "END" has been already preprinted in the columns 8 - 10 of the end card.

Card Sequence The same applies to these columns (78 - 79) as was stated for the field (16). This card is the last for this profile in question and has the highest sequence number consequently.

* * *

For illustration we have included a sample output from the profile print programs (Fig. 6) which shows you what the end product of your profiling form looks like.

```
6/10/70                          CIS QUESTION LISTING                          PAGE    1

100004 1     OR STANDERA                                  CALC UNI INFOSYS 06/70      C1
100004 CON1  ABS COMPENDEX OR TEXTPAC OR TEXT-PAC                                     C2
100004 A1    FULL OR FREE OR NORMAL OR CONTINUOUS OR COHERENT OR RUNNING          C  3
100004       ADJ TEXT                                                                 4
100004 A2    PROCESS$*                                                                5
100004 CON2  A1 AND A2                                                                6
100004 CON3  SEARCH ADJ EDIT$$$                                                       7
100004 CON4  PROFILE$ OR QUESTION$ OR QUER$$$ WITH CONSTRUCT OR SET$*             C  8
100004       OR FORMULAT$$$ OR MAINT$* OR ADJUST$$$ OR TRANSLAT$$$                    9
100004 A3    INFORMATION OR RETRIEVAL ADJ SYSTEM$ OR CENTER$ OR HANDL$$$          C 10
100004       OR PROCESS$$$ OR RETRIEVAL OR STORAGE OR SPECIALIST$                    11
100004 A4    DESIGN$$$ OR EVALUAT$$$ OR COST$$$ OR MEASURS* OR ANALY$*            C 12
100004       OR PERFORM$* OR ADMINISTR$* OR MANAG$* OR ORGANI$* OR PLAN$*         C 13
100004       OR REQUIRE$* OR ERROR$ OR EFFICIENCY OR EFFECTIV$* OR COMPAR$*           14
100004 CON5  A3 AND A4                                                                15
100004 A5    MECHANI$* OR AUTOMATI$* OR COMPUT$* OR SOFTWARE OR PROGRAM$*             16
100004 CON6  A3 AND A5                                                                17
100004 CON7  NOT LIBRAR$$$$                                                           18
100004 CON8  RETROSPECTIVE WITH SEARCH$$$ WITH STORAGE OR CORE OR MEMORY              19
100004 A6    SDI                                                                      20
100004 A7    SELECTIVE AND DISSEMINATION AND INFORMATION                              21
100004 A8    CURRENT ADJ AWARENESS                                                    22
100004 CON9  A6 OR A7 OR A8                                                           23
100004 A9    HIT$$$$ OR MATCH$$$$ OR MISMATCH$$$ OR RELEVANCE OR RECALL OR        C 24
100004       MISS OR TRASH                                                            25
100004 CON10 A3 AND A9                                                                26
100004 CON11 USER$$ AND FEEDBACK                                                      27
100004 CON12 A4 AND A9                                                                28
100004 A10   A1 AND A2                                                               29
100004 A11   A6 OR A7 OR A8                                                          30
100004 A12   A10 AND A11                                                            31
100004 CON13 A5 AND A12                                                           32
100004 CON14 WHITBY CONTROL2$$ ADJ DK                                             33
100004 CON15 STANDERA NOT-CONTROL2$$ ADJ CR                                       34
100004 CON16 @GIPSY                                                               35
100004 CON17 A9 AND A12                                                           36
100004 END                                                                        37

TOTAL NUMBER OF PROFILES -     1
```

Fig. 6 - Profile printout

III. EXTENDED COMPENDEX FEATURES

(1) PRINT CONTROLS

Records on the COMPENDEX data base have the following structure:

```
 00   Design of mechanical linkage using analogue computer
 09   Mechanisms 40373
 10   EIX69X120761
201   REA D.P.
 3    40373
4 Z   Simulation V13 N1 JULY 1969 P13-23
401   Electronic Associates Ltd., Burgess Hill, England.
 50   Application of steepest descent technique to solution of linkage
      problems is considered, using, as example, typical five-bar geared
      linkage; in linkage design study it is necessary to be able to
      access quickly effects of changing linkage parameters on motion of
      linkage; a way in which analog computer can do this, using logical
      control and high speed repetitive operation is fully discussed. 40373.
 60   Mechanisms
610   00-A112
611   00-A300
650   Computers, Simulation
```

The numbers on the left side denote various parts of this data base item and are called "Print Controls". TEXT-PAC enables the user to search any or all of these elements, which are explained below:

```
 00   title
 09   subject heading (subheading may also be present), EI number
 10   ID (identification number) which is the internally assigned
      sequential number
201   Author (as many as 99 authors may be specified under 201-299)
 3    EI number
4 Z   Citation (Source)
401   Author affiliated (of first author if more than one specified)
 50   Abstract
```

60 Subject heading (and subheading)

610 (to 649) Sales codes relating to the former Card Service of EI. These will soon be replaced by CAL identifying areas in the CARD-A-LERT service of EI.

650 Access words or keywords

(2) CONTROL AND NOT-CONTROL

The outstanding feature of the TEXT-PAC system is its ability to search the entire record as we have shown. Searching limited to one or more of these print controls is possible, although not typical. It may only be justified for example if we need all papers published by an author. Then we search only in the print control 2$$ e.g. CON14 WHITBY CONTROL2$$ ADJ DK

The means for conducting a search in this way is called "CONTROL" logic. If we use "CONTROL" then the hit will only be achieved if the logic specified in the profile match the logic in the specified print control of any data base record.

The rules governing use of CONTROL logic are:

1. We can use it only with profile words (not logical symbols)
2. The CONTROL is followed by print control without blank
3. As many as seven print controls may follow a word. They are separated by commas without blanks
4. Print controls are listed in ascending order
5. Print controls may be masked by dollar signs on the second and third character
6. When using ADJ or WITH logical connectors, you may only use CONTROL logic with the first profile word to the left of the first ADJ or WITH.

The "NOT-CONTROL" logic is subject generally to the same regulations. It is used if we do not want the search to be conducted in a certain print control e.g. we do not need our own papers because we have them thoroughly documented.

CON15 STANDERA NOT-CONTROL2$$ ADJ OR

Keep in mind that limiting the search uses only partially the capabilities of the system.

(3) CAPITALIZATION

Because TEXT-PAC handles both lower case and upper case printing you may refine your profile even more by taking advantage of the following rules. It should be pointed out that capitalization is not widely used though there are specific cases where it is warranted.

Assuming you have not specified capitalization, the profile word will be a match if there is such a word in the data base, no matter if the letters are in upper case or lower case or in any combination.

"GIPSY" in the profile formulation will match any information about gipsy as well as about the acronym GIPSY denoting an information system.

If you specify one "at sign", it will match only all upper case characters (GIPSY) or initial capitalization (Gipsy). See our example CON16 in Fig. 6.

Correct specification: @ GIPSY

Two "at signs" (@@GIPSY) will find only all upper case characters.

If you wish to have hits only with a word containing mixed upper and lower case letters, then you may use the number sign. #PH will match only pH which means the concentration of hydrogen ions.

As you may have recognized you may make your job a lot easier without specifying the capitalization unless it is necessary to do so.

(4) SOME LIMITATIONS IN THE TEXT-PAC SYSTEM

The match criterion	1-9 in the Current Information Service (CIS): Retrospective Search 1-19
The query word length	Maximum printable 38 characters. Internal truncation to 20 searchable characters

Selective truncation	Maximum 6$ (6 characters)
Unconditional truncation	$* matches all words with the root specified to a total of 20 characters
CONTROL, NOT-CONTROL	A maximum of 7 print controls can be used per question word
AND	Connects maximum 15 query words
Back-referencing to logical symbols	Any logical symbol may be cited maximum 15 times in a search expression
Length of a logic level	Maximum 10 cards (9 continuation cards) Maximum 15 logical symbols
Levels of back-referencing	Maximum of 3.

(5) EXERCISE: QUIZ AND ANSWERS

Now after explaining the principles of the TEXT-PAC profiling logic and acquiring some skill in completing the COMPENDEX PROFILING FORM, let us take a quiz which will show you the extent you have understood the text. In addition, it will point out some limitations you have to bear in mind when constructing COMPENDEX profiles.

QUIZ

1. How many characters may any profile word consist of and how many of them are searchable?

2. How many levels are permitted for back referencing?

3. What numbers may be specified as match criterion?

4. What is the difference between selective and unconditional truncation?

5. How many times may you reference any one logical symbol (A1, A2....)?

6. How many logical symbols may any one concept contain? (A7 A1 OR A2 OR.

7. How many lines (=cards) may specify any one concept?

8. How many profile words may be connected by the logical connector AND?

9. What are the capabilities of the Capitalization feature?

10. What are the logical connectors you can use in COMPENDEX profiling?

11. What are, briefly stated, the functions of the logical connectors?

12. Is this correct?
A1 LIBRAR$$$ AND INFORMATION ADJ CENT$$$

13. Is this right?
A1 LIBRAR$$$ WITH AUTOMATION OR MECHANIZATION

14. Is this statement right or wrong? A7 ABS A3

15. May you specify the search expression this way?
CON5 CON3 AND A16

16. Find out what is the meaning of the Print Control 50.

17. If you decide to confine your search to certain print controls, how many print controls are permitted per profile word?

18. What should you remember when using the CONTROL facility?

19. What are the Update Codes used for?

20. In the paragraph "Update Code" we have shown an example of how to code a change in the existing profile. Assume that this profile adjustment has not proven to be a happy solution and you would like to return to the original profile formulation, how would you code it?

ANSWERS

1. 38 (20 of them searchable).

2. Maximum of three levels.

3. 1 through 9.

4. Selective masking (up to six$) covers six characters maximum, unconditional masking ($*) up to twenty characters.

5. Maximum 15 times.

6. Maximum 15.

7. Maximum 10 (9 continuation cards).

8. 15.

9. It enables us to find in the data base the profile words
   (1) regardless of whether in upper or lower case,
   (2) only in upper case,
   (3) in upper case or initial upper case,
   (4) mixed lower and upper case.

10. OR, AND, WITH, ADJ, NOT, ABS, CONTROL, NOT-CONTROL.

11. Answer to this question see in the previous text under the logical connector respective.

12. NOT

13. YES

14. WRONG

15. NOT

16. An abstract

17. Seven

18. It makes the COMPENDEX not to utilize all of its inherent capabilities.

19. It allows us to delete, revise or add lines to our existing profile.

20.

```
100004  CON1  ABS COMPENDEX OR TEXT-PAC..................................... 02R
100004  A1    FULL OR FREE OR NORMAL OR CONTINUOUS OR COHERENT OR RUNNING.C03R
100004        ADJ TEXT..................................................... 04A
100004        ..............................................................06D
```

ANNOTATED BIBLIOGRAPHY

You may find some of the following literature useful in your search for proper profile words, synonyms, antonyms, acronyms and related terms.

CRISPIN, F.S. (1970) - Dictionary of Technical Terms. New York, The Bruce Publishing Company.
("....guide to definitions and illustrations not found in standard dictionaries or ordinary technical books. Terms used in modern trades, technical procedures, industry, shopwork and occupations of mass production are explained").

ENGINEERS JOINT COUNCIL (1969) - Thesaurus of Engineering and Scientific Terms, New York, Engineers Joint Council.
(A List of Engineering and Related Scientific Terms and their Relationships for Use as a Vocabulary Reference in Indexing and Retrieving Technical Information.

GRAHAM, E.C. (1967) - The Basic Dictionary of Science. New York, The MacMillan Company.
(The dictionary gives the senses of more than 25,000 words used in science and technology).

KAUFMAN, S., & Others (1968) - TEXT-PAC, S/360 Normal Text Information Processing, Retrieval and Current Information Selection System.
ITIRC, IBM Corporate Headquarters, Armonk, N.Y. 10504.
(Gives all the instructions needed for the implementation of the TEXT-PAC system. This system provides a natural text information processing capability).

MANDEL, S. (1969) - Dictionary of Science. New York, DELL Publishing Co. Inc.
(Terms and definitions from the following fields: aeronautics, astronomy, biology, chemistry, engineering sciences, geology, mathematics, physics).

McGRAW-HILL Encyclopedia of Science and Technology (1966). New York, McGraw-Hill, Inc.
("....The modern multivolume encyclopedia aimed at authoritative, comprehensive coverage of the physical, natural and applied sciences...."). It consists of 15 volumes.

NEWMAN, J.R. (1967) - The Harper Encyclopedia of Science. New York, Harper & Row Publishers.
(Covers various branches of science in one volume).

UNIVERSAL DECIMAL CLASSIFICATION - Various editions and Divisions.
(Provides a systematic hierarchical view of scientific and technological fields).

VAN NOSTRAND'S Scientific Encyclopedia (1968). New York, Van Nostrand Company, Inc.
(".... defines and explains 16,500 alphabetically arranged terms of fundamental interest. ....a basic reference work on science, engineering, mathematics and medicine..."). One volume.

INDEX