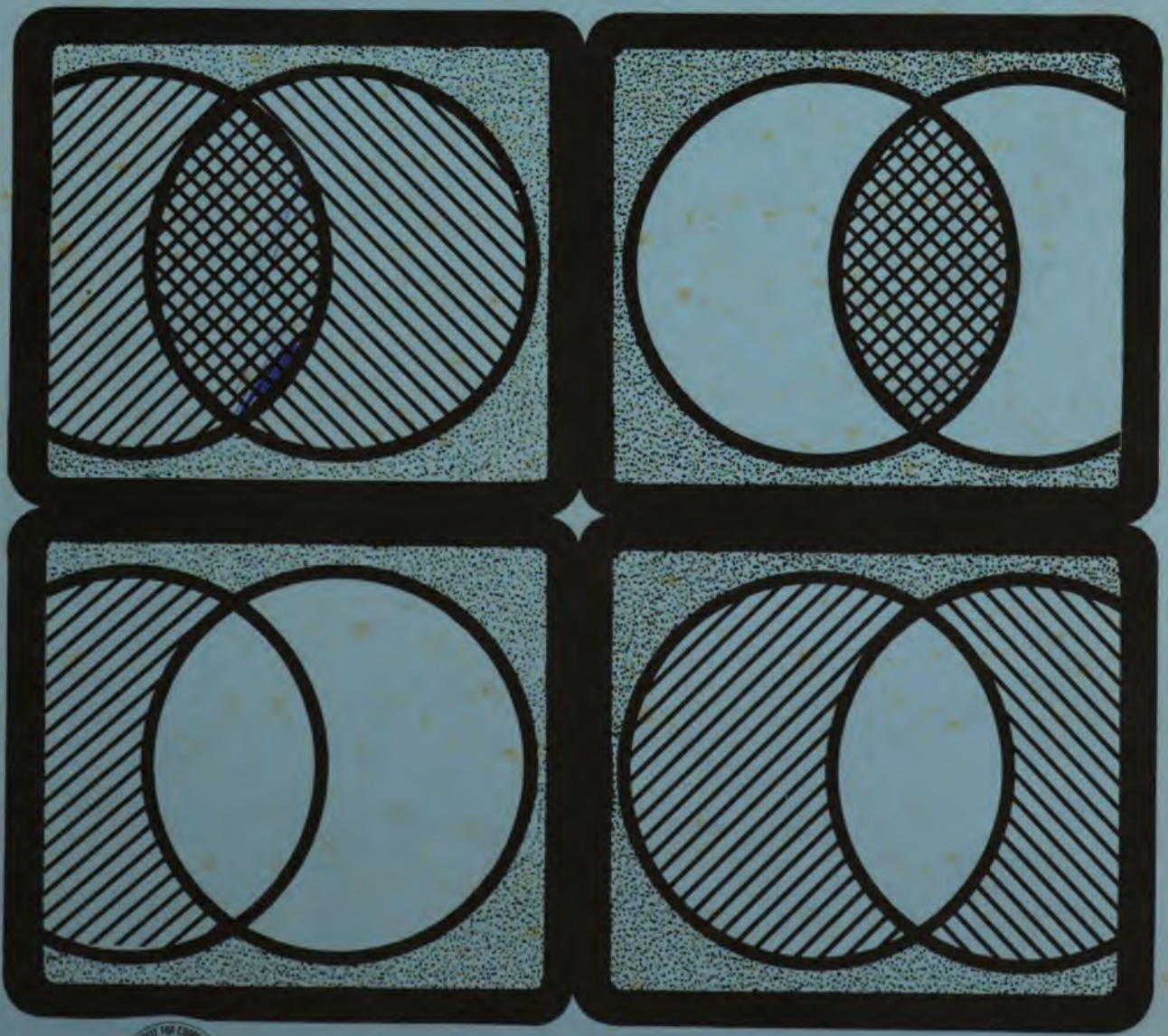
AGRINTER-AGRIS
Centro Interamericano de Información Agrícola
IICA—CIDIA

PIADIC
Information Management Series

3

Managing Information for Selective Dissemination



IICA
20
29

IICA
INSTITUTO INTERAMERICANO DE COOPERACION PARA LA AGRICULTURA

IICA—CIDIA

Managing Information For Selective Dissemination

Construction of Interest Profiles

Text Prepared by Finn Damtoft

00007708

PIADIC
INFORMATION MANAGEMENT SERIES

This series contains the experiences gained by
Programa Información Agropecuaria del Istmo Centroamericano

A joint project between:
Instituto Interamericano de Cooperación
para la Agricultura (IICA)
and
Regional Office for Central America and Panama (ROCAP)
(Contract AID 596 - 0048)

Title	Author/Editor	Series No.
Rural Area Profiling	Franklin E. Rosales	1
Area Frame Probability Sampling	PIADIC	2
Managing Information for Selective Dissemination	Finn Damtoft	3
PIADIC Bibliography	PIADIC	4

The series can be obtained from
IICA
Apartado Postal 55, 2200 Coronado, San José, Costa Rica
or
ROCAP
8a. Calle 7-86, Zona 9, Guatemala, Guatemala

About This Manual

We have designed this manual as a guide to the construction of "interest profiles" for use in computer-based searching of bibliographic data bases. A "profile" is a statement of the user's search request or specific subject interest that he constructs according to certain rules that are explained in the manual. These "profiles" are the key to recurrent, computer-based literature search.

Our aim is to describe the procedures to follow in developing and actually writing such profiles within a "workshop" setting. We hope the manual will serve as an introduction to novice information specialists, as well as for review by more sophisticated users of information systems that are now coming "on-line" in many developing countries. The methodology and procedures have been tested over several years of intensive application in Central America.

The manual first provides an overview of the fundamentals of profile development and defines the mechanical rules followed in creating a computer-compatible search request. Next, the reader is shown how to develop a statement of search topic, using keyword lists and examples from the AGRINTER and AGRICOLA data bases. The Appendix provides actual examples of profile forms used for retrieval of citations.

Contents

Basic Concepts

The computer retrieval of bibliographic citations from bibliographic data bases such as AGRINTER, AGRIS or AGRICOLA, is based on the same principles used when we search a coordinated index system or Termatrex.

To provide a brief illustration of the indexing process, we can assume that the document to be indexed is the publication *"Notes on the Corn and Bean Production Courses"* (Figure 1).

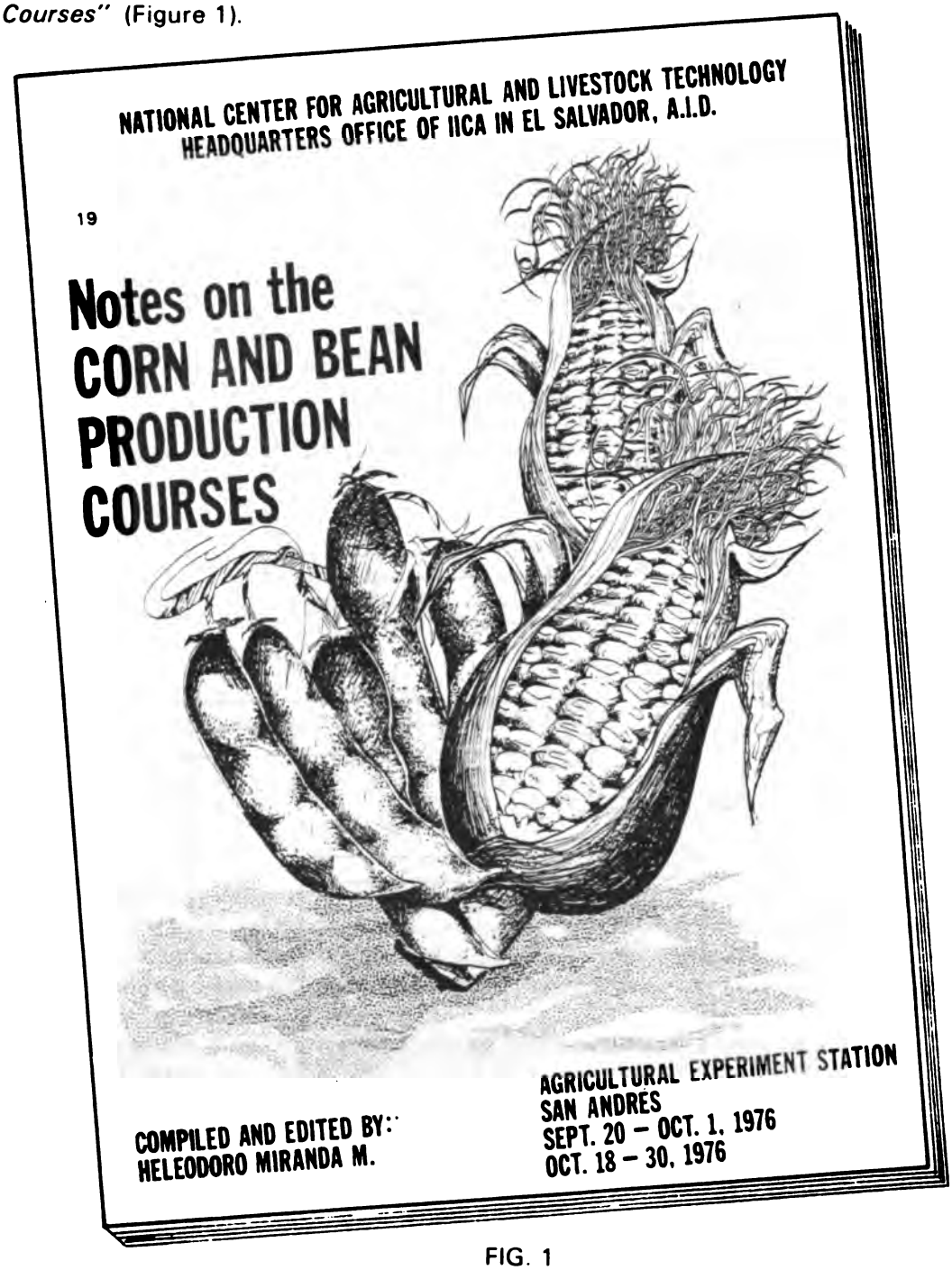


FIG. 1

The Termatrex (or Optical Coincidence System) is somewhat more sophisticated, because it uses a ray of light to identify the appropriate citations. Instead of writing a sequential number on the index cards, a "punch hole" is perforated in the appropriate location with a precision perforator. Such systems can manage a larger number of documents (5,000 - 10,000) and more keywords (500-800).

Use of Computer

When we require an index that covers 10,000 to 100,000 articles, or more, with searches using 40 to 50 index terms to cover all aspects of a topic, we must seek the help of the computer. Work on this scale is impossible to control and search manually.

The computer, however, is ideally suited to the operations necessary for control, storage, and retrieval of a large number of documents. Although the computer performs no operations that cannot be done by humans, the computer can operate with a speed and precision that we humans could never match.

When we provide the computer with a program of instructions, it will function with these instructions at the rate of perhaps millions of additions per second. Thus, the computer is equipped for any tasks that involve large quantities of computation, manipulation, comparison, and retrieval. (But we should remember that all of the computer procedures can be written beforehand, as a set of instructions. Actually, the computer's system is only a more sophisticated version of the manual system that we discussed previously, but performed with incredible speed.)

Index for Information Storage

Let's now take a look at the various files that the computer generates and maintains to function as an index for stored information (Fig. 4).

(1) **The File Index**—is a sequential file of the terms used and arranged in alphabetical order. For each term, it indicates:
   —the number of files for the term
   —the last or first number filed sequentially.

(2) **The Search File**—is the equivalent of the index cards in the manual system. This file also is arranged sequentially in alphabetical order of the terms used. However, after every indexed term, there is a listing in ascending or descending order of the sequential numbers that represent the article or document in which the indexed term appears.

(3) **The Entry Unit File**—contains the bibliographic information referring to each of the bibliographic citations that have been entered in the file. This file is arranged sequentially in ascending order, with the unique sequential number assigned to each entry. This sequential order is the most common because it makes the most effective use of an inexpensive method of filing (magnetic tape) and because it is very efficient for use in "batch" operations at periodic intervals.

The sequential processing method is best suited for "batch" operations because the entire file is searched by the computer; even if only a few operations are required, the whole file is searched. Thus, search operations are more efficient when a reasonable quantity can be processed.

The document has now been indexed; however, to register and control this information, we must:

— assign a unique sequential number to the indexed document (Figure 1).
— select the appropriate keyword cards and note the sequential number on each one (Figure 2). The unique number is noted within the appropriate column on the index card, to correspond with the last digit of the sequential number. Noting in this format makes it easier to identify the numbers.
— create an Author's index card with all necessary bibliographic information, and perhaps a summary of the article's contents, to correspond with the unique sequential number (Figure 3).

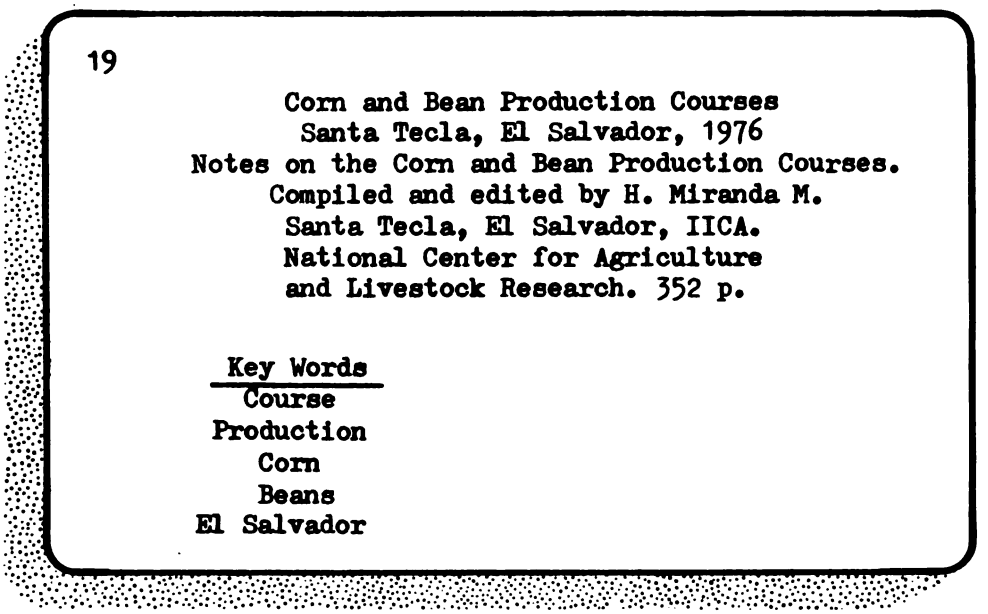


FIG. 3

We have now integrated the document in the files and it is retrievable. To retrieve the information from the index, a user selects the keywords that best describe the subject matter needed. The appropriate index cards are then selected from the files and the user makes a visual check to see if some sequential numbers appear in all the cards selected.

For example, if our user is interested in the production of corn, he should select the keyword cards for "Production" and "Corn". These two keyword index cards are removed from the file and the user notes whether a unique document number appears on both cards. If the unique number of a document is repeated, it probably will be of interest to the user.

The manual system that functions as we have described is useful if the document collection is not very large (less than 2000), or if the number of keywords for the file search is no more than 200-250.

How We Use the System

Let's suppose that a user is interested in the production of beans and corn in El Salvador. The computer is programmed to identify the citations with the key words:

Production - Beans - Corn - El Salvador

It will also identify all the coordinates of these four key words. After we give the computer these keywords, it goes to its index file to obtain some preliminary information. For example, it asks:

(1) Are these keywords in the system?
(2) How many entries are there?
(3) What is the first or last sequential number?

This initial phase could be represented as shown below:

Keyword File	Keyword	First Sequential Number
15	Production	11
402	Beans	19
300	Corn	18
7	El Salvador	10

The computer then goes to the search file and obtains each of the four keywords with all the unique sequential numbers listed for each keyword:

Production	Beans	Corn	El Salvador
11	19	18	10
15	20	19	19
17	25	20	20
18	27	35	35
19	29	37	37
20	30	40	39
25	35	-	42
35	37	-	-
42	39	-	-
110	-	-	-
230	-	-	-
232	-	-	-
315	-	-	-
326	-	-	-
402	-	-	-

Each of these keyword lists is compared with the others—very much like slot machines that operate in casinos. Each time identical numbers are indicated in the four columns, the computer notes the coordinates.

When all numbers have been compared, the computer prints the coordinates.

For our example above, it would list coordinates: 19 - 20 - 35

With these coordinates, the computer searches the entry units to obtain bibliographic information about entries 19, 20 and 35. When located, this information is printed and made available to the user.

Boolean Logic

Having demonstrated how the computer makes its search, we can examine the use of Boolean logic to instruct the computer in combining terms.

In general, we use three logical operators related to Boolean logic in the search for Selective Dissemination of Information (SDI): "OR", "AND" and "NOT". When we use these logical operators in our search, we can increase the precision of the final output by eliminating or reducing unwanted elements, or through limiting the number of parameters to be searched. Of course, we also can increase and widen these parameters to increase the recall of final output.

Logical "OR"—is generally used to widen or expand searching concepts. It allows the searcher to combine terms or concepts into a single searching concept. The symbol used for "OR" is " ❙ ". The "OR" concept can be illustrated as shown in Figure 5.
In day-to-day use, "OR" implies exclusive selection of one item as opposed to something else. (We can stay here *or* go to the movies, but cannot do both things at the same time.) In Boolean logic, all terms and concepts combined through "OR" have the same value. In the Venn diagram, we see a small darker area where the two circles intersect. This area indicates those points that contain both oranges and lemons.

Generally, a searching program is designed so as not to have such intersecting areas. If the program has not been adjusted, we would retrieve all the citations containing "oranges" and all with "lemons"; however we will also receive all those containing both oranges and lemons.

Logical "AND"—serves to narrow the searching stage. It implies that the two terms must be presented together in a single citation to be ready for retrieval. The symbol for "AND" is "&". This indicator is shown in Figure 6.

The only retrievable items in this case, using the "AND" relationship operator, would be the citations represented in the intersecting area of both circles.

Those items contain both "A" and "B". If we keep using oranges and lemons as keywords, we would be retrieving only citations with oranges and lemons together.

Logical "NOT" (AND NOT). This relationship operator is used to eliminate the elements not required in the search. The symbol for NOT is "& ¬". The "NOT" logic is very difficult to use. It is a very strong operator, but if not properly used could eliminate desired citations. One of the better uses for & ¬ is when the user wants to exclude the search of specific types of publications such as monographs, specific institutionals publications, or articles of identified journals (Figure 7).
The versatility of the relationship operator "NOT" can be graphically presented as in Figure 8.

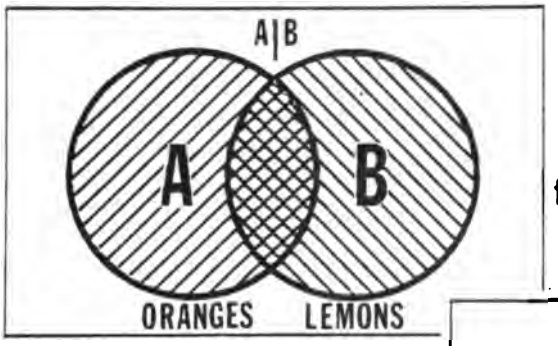


fig. 5

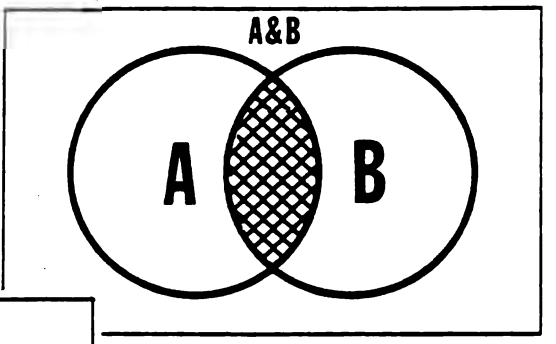


fig. 6

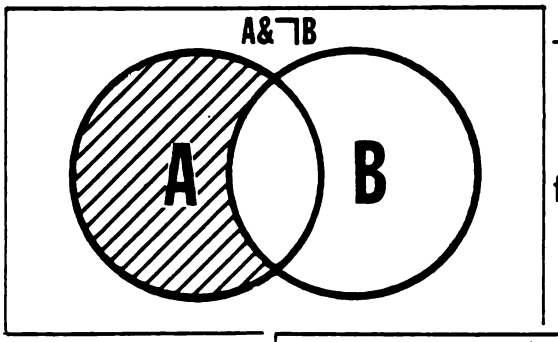


fig. 7

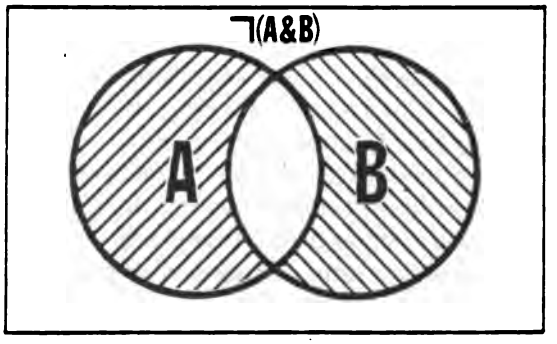


fig. 8

SHADED AREAS EXPRESS THE UNIVERSE OF THE LOGICAL ARGUMENT
&¬ = NOT, & = AND, | = OR

Profile Construction Sequence

With the above general background, we can begin the process of constructing an interest profile for SDI. In sequential steps, we:

1. **describe a clear and concise topic to be searched.**
2. **develop a list of key words that adequately express all aspects established in the original topic.**
3. **divide these key words logically in groups of concepts.**
4. **design a searching strategy that uses Boolean logic formulas.**
5. **evaluate the results.**

Statement of the Search Topic

Generally, the user asks for the information he needs in the most precise and detailed manner possible. We expect that the user has the capacity to describe the topic he needs without major problems. However, this is not always the case. If the user cannot write a concise description of his needs, he might simply be inexperienced with the formulation of an SDI profile. We should be prepared, then, to assist the user from the beginning of the "profile" construction. A typical error is to describe the topic too generally. In such cases, the request usually begins with: "All that the data base has. . .". But, if the search is conducted on this basis the end result might be the retrieval of 1500 articles or more. The user would still have a lengthy task of selecting the articles of interest. Another common mistake is to express interest in more than one topic. Our solution here is to write several precise search topics.

A trained search analyst can help the user to refine his search request, revising it according to the user's needs and the capabilities of the system. The profile construction for SDI can be done either by *specialists* on the subject or by *generalists*. Our experience suggests, however, that the specialists can produce better profiles. This is probably because the user is usually a specialist on the subject and has more confidence in dealing with another specialist.

A well developed statement of the search topic should include the following information for the searcher.

1. **The specific subject to be searched.**
2. **Specific aspects of the search topic; for example, production, fertilizing requirements, weed control, and other related to the topic.**
3. **The purpose of the search: for example, use in a research paper, a thesis, or as a support element on a research project.**
4. **Any special requirements: language limitations, age or dates, or information coming from only one specific author.**

Keyword List

After the searcher has a clear idea of the search topic, the next step is to devise a list of appropriate keywords for the computer search.

Before we can select keywords, however, we must research the capacity and the structure of the data bases that we plan to use. We also need to know the capacity and the structure of the search program(s) to be used.

A data base used for SDI purposes is formed by a number of bibliographical citations. When a citation is recorded on tape, it is divided into units called fields. Citation fields would include: author, title, date of publication, title of the journal, index term, and analytical summary. Each field contains a specific type of information, and the summation of all fields represents the complete citation.

For our search, we identify the individual fields, with each field coded with a specific label. This procedure is standard for all data bases; however, for our search we need to know if all the fields can be searched individually, or only partially.

Let's suppose that the user is interested only in publications written by E. Blanco. The computer should receive instructions according to the field to be searched for this term. If not limited to this specific field, the computer will search the author field and find E. Blanco; but it will also search in the title field and retrieve any title with the word "blanco" in it. Unless the profile designer indicates the exact field of interest, the computer will retrieve all related fields without making a distinction.

In the AGRICOLA Data Base the following logical indicators are used:

CCC = Classification code
ATI = Author(s) name(s)
TXT = Title and Indexing Term
PDL = Primary Document Language (original language)
PRC = Primary citation (Journal Title)
PDD = Primary Document Date

We must know the method by which authors have been entered in the system. The AGRICOLA data base lists author's entries by the anglo-american rules, using the full name to avoid confusion among authors with very common names. Both AGRIS and AGRINTER enter author's names using their last names and initials. We find that this practice has disadvantages when we search, as it is very difficult to note a difference between authors when using only initials.

Both AGRICOLA and AGRIS have citations in several different languages: English, Spanish, German, Russian and others. A search program functions, in general, by comparing the profile's keywords with the data base terms, on a character-to-character basis. The titles, indexing terms and summaries have been translated into a common language (English) so that the data base can be searched by all users. However, if a user feels that an efficient search can be made in only one language, Spanish for example, the profile could be done in Spanish—then, however, only the citations in Spanish would be retrieved.

The AGRINTER Data Base serves a predominantly Spanish-speaking-region— 90% of AGRINTER citations are in Spanish. The other 10% of the citations, those that appear in Portuguese, French and English, have been registered in the original languages. Improvements are made to these terms in Spanish. AGRINTER profiles are done in Spanish, also.

Free Vocabulary or Controlled Language?

With the free-vocabulary approach, all the words that appear in the citation are useful as keywords. In fact, a small program eliminates small words such as "and", "of", "to", etc. from the index.

In a free-vocabulary data base, the keyword selection is open. There is no organization such as in a thesaurus or hierarchical list. Both AGRICOLA and AGRIS/AGRINTER Data Base use the free-vocabulary approach. When an indexing controlled language is used, such as that developed for MEDLINE, only certain words may be used as keywords. Acceptable keywords are listed in a thesaurus or in a hierarchical system.

The user or the search analyst should also know if the search program will let us truncate keywords to avoid the listing of all possible variations of a keyword that has prefixes, suffixes or infixes.

The truncation may be made from the right, as in methyl*, which would retrieve terms such as:

methyl*
methyl* ene (methylene)
methyl* ation (methylation)

Truncation from the left, as in *methyl, would retrieve terms such as:

methyl
4, 5, 8 -tri *methyl (4,5,8 - trimethyl)

Combined left-right truncation, as in *methyl*, would retrieve terms such as:

methyl
de *methyl* ase (demethylase)

Asterisks appended to the profile term indicate where truncation should occur.

***Keyword** Only the characters following the asterisk (*) must be matched. Blanks or other characters may precede the asterisked position.

Keyword* Only the character preceding the asterisk must be matched. Any combination of blanks or characters, may occur after the asterisked position. This truncation made is useful for retrieving words in both singular and plural.

Keyword The characters enclosed by asterisks (*—*) must be matched but it can be preceded or followed by any combination of blanks or characters.

Truncation of keywords is economically useful because it allows the search program to do fewer character-to-character comparisons.

AGRICOLA: Generally, the user who has asked for the search is the best informed source for appropriate keywords for that search. This user knows the topic and the related terminology.

When the user has provided all appropriate keywords, the search analyst's work begins. Using a free-vocabulary approach, the analyst should try to extend and define the question so as to include all specific terms that are identified with the topic. This would include the identification of specific and generic terms, plus synonyms and associated terms. If a means is not available for truncating the singular and plural forms, they should also be included in adjective forms. The scientific name of any insect, crop pest, virus, etc., should be included along with their common names. In some cases (a plant for example), various names may be commonly used within a country that will indicate the information required. If a search is related to one specific geographic area, the search should also include the district and, perhaps, the country or state. All of the citations of AGRICOLA and AGRINTER are classified and the number or numbers of these classifications can be used as a keyword.

If for example, the search topic related to the word "cattle", the user or search analyst should, of course, use "cattle" as a keyword, but also "cow, cows, calf, calves, bull, bulls, heifer, heifers, bovine, and beef cattle". Depending on the topic, it might also be useful to include the names of the different breeds of cattle.

Most search analysts enjoy this process because of the investigative work included. Sources to be consulted during this work include, a thesaurus, general and specialized dictionaries, bibliographies and vocabularies (for example, the AGRINTER vocabulary, the Canadian Agricultural Vocabulary, etc.). Complete coverage of a topic may require a very large list of key words; it is not uncommon to have 40 or 50 keywords in a search.

AGRINTER: The AGRINTER system also uses a free-vocabulary approach with some modification. During the indexing process, some words that are most important to express the topic are chosen from the title. If the indexer feels that more key words should be added to the title to make the content of the article more comprehensive, these additions should be made.

The key words chosen are entered into "A List of Keywords for AGRINTER". This computer listing represents the only keywords that can be used to search the AGRINTER data base; however, there is no hierarchy in the ordering of the terms in this list, thus AGRINTER does not have a thesaurus at present.

When using the AGRINTER system, the user must select his keywords from this list. Because AGRINTER receives input from Brazil and the Caribbean area, the user and the search analyst should include key words in Portuguese, French, and English to assure total access to the data base.

Grouping by Definition of Concepts

If we feel that a sufficient number of keywords have been selected, we can begin grouping keywords by concept.

"Grouping by concept" is the process of assembling terms with similar meanings that we want the computer to treat as synonymous.

The computer can only look for a series of letters, it does not have the capability of judging the content of the terms. The search analyst should try to resolve this deficiency of the computer by increasing the number of parameters that must be filled to retrieve a citation. He can do this by dividing the keyword list into groups of synonyms or terms that are conceptually associated.

Let's consider a user who would like information about how size and weight of the seed affect the growth and yield of the corn plant. His list of keywords might include the following:

Corn	Dimension	Production	Germination
yield	size	maize	germinate
yields	mass	corn	germ
growth	length	mays	growing
seed	width	zea	growth
seeds		weight	growth
			producing
			produce

After grouping within these concepts, the list should read:

Corn	Seeds	Size	Yield
zea	seeds	weight	yields
maize		weights	growth
mays		dimension	production
		dimensions	germination
		length	growing
		width	growth
		mass	increase
			increases
			producing
			produce

We have now divided the different keywords into four groups of larger concepts in which the keywords are associated in meaning or are synonymous. This process requires a great deal of common sense, as well as a knowledge of the material. Often we find that the grouping will provide additional appropriate keywords; these can be placed within appropriate groups.

If truncating facilities are available, we would set up a program to reduce the number of keywords—being extremely cautious, however, because we might accidentally include irrelevant words.

We can now number each of the groups **G001**, **G002**, **G003** and **G004**, and we are ready to apply Boolean logic.

In the previous example we have four groups of principal concepts, each one with a number of keywords. We now want to instruct the computer to search for all the citations containing all four concepts. For example, the citation that describes the relationship between the production and the size of *barley* seed will be rejected, because we are interested only in the relationship between the production and the size of *corn* seed. Within each group, any of the keywords is acceptable.

We tell the computer to accept any of the keywords within each group, by combining them with the logical operator "OR".

The list now looks like this:

G001	G002	G003	G004
corn	seed	size	productivity
OR	OR	OR	OR
zea	seeds	weight	productivities
OR		OR	OR
mays		weights	production
OR		OR	OR
maize		dimensions	germination
		OR	OR
		length	growth
		OR	OR
		width	grew
		OR	OR
		mass	growing
		OR	OR
		mass	growth
			OR
			growing
			OR
			produce
			OR
			producing

Because we are interested in only these citations that include all four concepts, we should instruct the computer that all retrieved citations should be a combination from all four groups using:

- any keyword of the 001 Group,
- any keyword of the 002 Group,
- any keyword of the 003 Group,
- any keyword of the 004 Group.

We can accomplish this through the linkage of all four groups of concepts, using the logical operator "AND". The formulas for the final search would be:

G001 & G002 & G003 & G004

By substituting concept numbers for the keywords of the group, we eliminate the tedious listing of all terms and provide the computer with a shorthand expression of our question. The logical relationship operator "OR" is pre-programmed within each one of the keywords in a group and the computer accepts it automatically.

We have used, so far, the logical relationship operators "OR" and "AND". Now, we will review the "NOT" operator.

The "NOT" operator will cause a citation to be rejected if a term is used that is designated by "NOT" logic. "NOT" is always used in conjunction with the "AND" operator. For example, if we want all the citations on rust or mildew affecting corn or barley, but we are not specifically interested in their effects on wheat, the profile would be as follows:

G001	G002	G003
Mildew	Corn	Wheat
Rust	Barley	

(Mildew or rust) and (corn or barley) and not (wheat)

G001 & G002 & G003

Logical operators are interpreted by the computer according to certain rules. The computer reads the Boolean expressions from left to right, with "NOT" and "AND" operators considered before "OR".

Parentheses can be used to alter the order in which logical operations are performed. Parenthetical expressions are operated on first, then the standard order of priority applies. The last step in the profile formulation is the evaluation of the printed result: Are the retrieved citations useful?

The final user generally is the best judge of the results. With his experience as specialist on the subject, he can determine the quality and use of the retrieved citations. Often the user will have also a better idea of the completeness of the search. If he notes some missing citations, we have an indication that we should have included more keywords in the search.

In spite of the best efforts of the user and the search analyst to develop a complete strategy for the initial search, the search may not be totally adequate. Each search provides additional experience to increase our efficiency. SDI is a valid tool for research, but it requires a complete understanding of its principles and an expert management of its methodology to be really effective. The examples that follow can serve as an introduction.

EXAMPLE 1.

Retrieve all citations dealing with animo acids and/or amides and sugar beets. Each concept should be defined as fully as possible to avoid missing relevant citations.

Concept 1. Sugarbeets

Expansion of this concept is relatively simple. By including both possible spellings (sugar beet and sugarbeet) and the Latin name (Beta Vulgaris) we have fully expanded the concept.

Concept 2. Amino Acids or Amides

This concept is easy to expand also, but it is a more tedious job. We must list the individual amino acids in which we are interested as well as any general terms that might be used. Thus we obtain the following list for Concept 2.

Amino Acid*	Methionine
Aspartic Acid*	Leucine
Threonine	Lysine
Serine	Ornithine
Glutamic Acid*	Histidine
Citrulline	Trytophan
Glycine	Arginine
Alanine	Asparagine
Cystine	Glutamine
Valine	Amide*

The asterisks indicate where truncation occurs and in the example are used to retrieve both singular and plural occurrences.

Since the classification codes (CCC) are not as specific as the two concepts in which we are interested we will not use them in this profile. The logical relationship we want between our two concepts is the Boolean operator AND: we want to include one term from Concept 1 *AND* one term from Concept 2 before a citation is retrieved.

If we assign concept numbers, we obtain the following:

Concept 1:

G001 Sugarbeet*
G001 Sugar Beet*
G001 Beta Vulgaris

Concept 2:

G002 Amino Acid*	G002 Methionine
G002 Aspartic Acid*	G002 Leucine
G002 Threonine	G002 Lysine
G002 Serine	G002 Ornithine
G002 Glutamic Acid*	G002 Histidine
G002 Citrulline	G002 Tryptophan
G002 Glycine	G002 Arginine
G002 Alanine	G002 Asparagine
G002 Cystine	G002 Glutamine
G002 Valine	G002 Amide*

Substituting concept numbers for the individual terms and using the Boolean notation, our Boolean expression as entered on the Profile Information Form would be: G001 & G002.

Additionally, on the same form, we must enter the code for the Library data base (CAIN).

On the Profile Coding Form, we enter the actual individual terms, with their concept numbers and term tags. In this example, the term tag for all terms would be TXT since the search terms may occur either in the title or as additional indexing terms.

On the following pages are completed Profile Information and Profile Coding Forms for this example.

USER: Fill in only sections outlined by *heavy lines.*

☐ REVISED PROFILE
PROFILE NO.

—OR—

☐ NEW PROFILE
STATE LAST PROFILE NO. ASSIGNED

(If not previously assigned a number, complete User Information Form)

UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL RESEARCH SERVICE
DATA SYSTEMS APPLICATION DIVISION

CURRENT AWARENESS SEARCHING SYSTEM
PROFILE INFORMATION FORM

NAME
TEL. NO. *(Include Area Code)*
ORG. UNIT

Punch this Identification No. on each line in Card Columns 75 – 80

DATA DESCRIPTIONS		Card Columns 1 thru 35 or 36 thru 70	73-74	75-80
Control Card *(Optional)*		E R A S E P R O F I L E (User No.) (Prof. No.)	0 1	
Control Card *(Optional)*		N E W P R O F I L E (User No.) (Prof. No.)	0 2	
Threshold Weight			0 3	
Boolean Expression	1-35	G 0 0 1 & G 0 0 2		
B. E. Continuation	36-70		0 4	
B. E. Continuation	1-35			
B. E. Continuation	36-70		0 5	
Data Base Codes	1-35	C A I N		
DBC Continuation	36-70		0 6	
DBC Continuation	1-35			
DBC Continuation	36-70		0 7	
Output Form			0 8	
Output Class			0 9	
Profile Review Date			1 0	

BOOLEAN OPERATORS
OR = |
AND = &
NOT = &¬

DATA BASE CODES
BIOLOGICAL ABSTRACTS = BA BIORESEARCH INDEX = BRI
CHEMICAL ABSTRACTS = CAO, CAE ENGINEERING INDEX = EI
NATIONAL AGRICULTURAL LIBRARY = CAIN
FOOD SCIENCE AND TECHNOLOGY ABSTRACTS = FSTA
GOVERNMENT REPORTS ANNOUNCEMENT = GRA

USER: Fill in only sections outlined by heavy lines.

UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL RESEARCH SERVICE
DATA SYSTEMS APPLICATION DIVISION

CURRENT AWARENESS SEARCHING SYSTEM
PROFILE CODING FORM

NAME	
TEL. NO. (Include Area Code)	
ORG. UNIT	
Page 1 of 1	

Punch this Identification No. on each line in Card Columns 75 – 80

CONCEPT NO. 1-4	WEIGHT (Optional) 5-7	TERM TYPE 8-10	TERM – Card Columns 11 thru 70	71-74	75-80
C 001		TXT	SUGAR BEET*	01	
C 001		TXT	SUGAR BEET*	02	
C 001		TXT	BETA VULGARIS	03	
C 002		TXT	AMINO ACID*	04	
C 002		TXT	ASPARTIC	05	
C 002		TXT	THREONINE	06	
C 002		TXT	SERINE	07	
C 002		TXT	GLUTAMIC	08	
C 002		TXT	CITRULLINE	09	
C 002		TXT	GLYCINE	10	
C 002		TXT	ALANINE	11	
C 002		TXT	CYSTINE	12	
C 002		TXT	VALINE	13	
C 002		TXT	METHIONINE	14	
C 002		TXT	LEUCINE	15	
C 002		TXT	LYSINE	16	
C 002		TXT	ORNITHINE	17	
C 002		TXT	HISTIDINE	18	
C 002		TXT	TRYPTOPHAN	19	
C 002		TXT	ARGININE	20	
C 002		TXT	ASPARAGINE	21	
C 002		TXT	GLUTAMINE	22	
C 002		TXT	AMIDE*	23	
C				24	
C				25	

ARS FORM 424
FEB 1972

EXAMPLE 2.

Retrieve all citations dealing with human nutritional requirements of juveniles and infants.

Again, there are two concepts in this example. The first is the concept of human nutritional requirements. The second concept is a qualifier or delimiter, namely, the age group which includes infants and juveniles.

Concept 1. Human Nutritional Requirements

This concept can be difficult to define under some circumstances since much research on nutritional requirements relates to non-humans. However, from scanning the classification codes, we find that 151000 specifically refers to human nutrition. Also 153500 refers to human nutrition education where we might also find relevant materials. By using these codes we can fully express Concept 1.

Concept 2. Infants and Juveniles

Expressing this concept is not so neatly accomplished. There is no auxiliary code to limit human nutrition to this age group, so we must attempt to list those specific terms which might occur in titles or as added indexing terms and which refer to the designated group. We obtain the following partial list:

infant
juvenile
child
children
boy
girl
teenager
baby
young adult
youth

Another likely inclusion is the term "student" and perhaps, "school", since many studies deal with grade school students or school lunch programs, etc. "Pupil" as a synonym for student should also be added. And so we continue the process until we obtain as definitive a list as we can compile:

infant*
juvenile*
child
children
teenager*
boy*
girl*
baby
babies
young
youth*
student*
pupil*
school*

After receiving our initial output, if we find that terms like "student" or "school" are causing retrieval of irrelevant citations, we can eliminate them.

Thus, our concepts linked by concept numbers would be:

G001 151000
G001 153500
G002 infant*
G002 juvenile*
G002 child
G002 children
G002 teenage*
G002 boy*
G002 girl*
G002 baby
G002 babies
G002 young
G002 youth*
G002 student*
G002 pupil*
G002 school*

Again, our logic would be Concept 1 *AND* Concept 2, so that our Boolean expression would be: G001 & G002.

On the Profile Coding Form, we would enter Concept 1 as G001, term type = CCC (Classification code), and the codes:

G001 CCC 151000
G001 CCC 153500

The remaining terms which comprise Concept 2 would all be TXT terms since we are willing to have them occur either in the title or as added indexing terms. Thus,

G002 TXT Infant*

On the following pages are completed Profile Information and Profile Coding Forms for this example.

USER: Fill in only sections outlined by *heavy lines.*

☐ REVISED PROFILE
PROFILE NO.

OR

☐ NEW PROFILE
STATE LAST PROFILE NO. ASSIGNED

(If not previously assigned a number, complete User Information Form)

UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL RESEARCH SERVICE
DATA SYSTEMS APPLICATION DIVISION

CURRENT AWARENESS SEARCHING SYSTEM
PROFILE INFORMATION FORM

NAME	
TEL. NO. *(Include Area Code)*	
ORG. UNIT	

Punch this Identification No. on each line in Card Columns 75 – 80

DATA DESCRIPTIONS		Card Columns 1 thru 35 or 36 thru 70	73-74	75-80
Control Card *(Optional)*		E R A S E P R O F I L E (User No.) (Prof. No.)	0 1	▲
Control Card *(Optional)*		N E W P R O F I L E (User No.) (Prof. No.)	0 2	▲
Threshold Weight			0 3	▲
Boolean Expression	1-35	G0 0 1 i G0 0 2		
B. E. Continuation	36-70		0 4	▲
B. E. Continuation	1-35			
B. E. Continuation	36-70		0 5	▲
Data Base Codes	1-35	CA I N		
DBC Continuation	36-70		0 6	▲
DBC Continuation	1-35			
DBC Continuation	36-70		0 7	▲
Output Form			0 8	▲
Output Class			0 9	▲
Profile Review Date			1 0	▲

BOOLEAN OPERATORS
OR = |
AND = &
NOT = &¬

DATA BASE CODES
BIOLOGICAL ABSTRACTS = BA
BIORESEARCH INDEX = BRI
CHEMICAL ABSTRACTS = CAO, CAE
ENGINEERING INDEX = EI
NATIONAL AGRICULTURAL LIBRARY = CAIN
FOOD SCIENCE AND TECHNOLOGY ABSTRACTS = FSTA
GOVERNMENT REPORTS ANNOUNCEMENT = GRA

USER: Fill in only sections outlined by heavy lines.

UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL RESEARCH SERVICE
DATA SYSTEMS APPLICATION DIVISION

CURRENT AWARENESS SEARCHING SYSTEM
PROFILE CODING FORM

NAME	
TEL. NO. (Include Area Code)	
ORG. UNIT	
Page 1 of 1	

Punch this Identification No. on each line in Card Columns 75 – 80

CONCEPT NO. 1-4	WEIGHT (Optional) 5-7	TERM TYPE 8-10	TERM – Card Columns 11 thru 70	71-74	75-80
C 0.01			CCC 151000	01	
C 0.01			CCC 153500	02	
C 002			TXT INFANT*	03	
C 002			TXT JUVENILE*	04	
C 002			TXT CHILD	05	
C 002			TXT CHILDREN	06	
C 002			TXT TEENAGE*	07	
C 002			TXT BOY*	08	
C 002			TXT GIRL*	09	
C 002			TXT BABY	10	
C 002			TXT BABIES	11	
C 002			TXT YOUNG	12	
C 002			TXT YOUTH*	13	
C 002			TXT STUDENT*	14	
C 002			TXT PUPIL*	15	
C 002			TXT SCHOOL*	16	
C				17	
C				18	
C				19	
C				20	
C				21	
C				22	
C				23	
C				24	
C				25	

ARS FORM 424
FEB 1972

	FECHA DE DEVOLUCION		

IICA
U20
329
Autor
Managing Information for
Selective Dissemination
Título
Fecha Devolución | Nombre del solicitante

Editorial services by
(n) Put International, Inc.
ROCAP PIO/T #596-0000.3-3-10004

DOCUMENTO
MICROFILMADO
Fecha: 12 SET 1982