

The B F Goodrich Information Retrieval System and Automatic Information Distribution Using Computer-Compiled Thesaurus and Dual Dictionary*

PAUL J. HORVATH

B F Goodrich Chemical Company, Development Center, Avon Lake, Ohio 44012

ALICE YANOSKO CHAMIS

B F Goodrich Company, Research Center, Brecksville, Ohio 44141

ROBERT F. CARROLL and JOYCE DLUGOS

B F Goodrich Chemical Company, Computer Department, Cleveland, Ohio 44115

Received July 21, 1966

The design and computer aspects of the B F Goodrich information system are discussed. EJC principles are used in indexing and thesaurus construction. The computer program, written for the GE-235, validates input index terms with the magnetic tape thesaurus. After generic expansion or synonym exchange, where necessary, the validated input is used to update the magnetic tape dual dictionary. Searches now consist of manual coordination of terms in the printout of the dual dictionary. The validated input is also used to match interest profiles in an automatic information distribution system.

A thesaurus, a dual dictionary, and a dissemination system are three elements common to many information systems. Seldom does a small- or medium-sized company have sufficient knowledgeable manpower available to develop and automate these interrelated elements simultaneously. As a result, some costly system redesign is often necessary to make the various elements of the total system fit together properly.

For the past three years, the B F Goodrich Chemical Co. and the B F Goodrich Co. Research Center have been working on a coordinated, automated information system containing the three elements noted above. In this paper, we discuss the design and operation of our information storage and retrieval system. There are many papers published on thesaurus construction, dual dictionary use, and dissemination systems, but few discuss the actual automation process. We discuss a coordinated system with all three elements directly and purposely related.

SYSTEM DESIGN

Early in our study, we decided to adopt a common scheme for indexing our internal documents. Two separate retrieval systems were actually planned—one for the B F Goodrich Chemical Co., Development Center, Avon Lake, Ohio, and one for the B F Goodrich Research Center, Brecksville, Ohio. The compatibility factor became

important because it was possible that the two retrieval systems might someday be merged.

We decided to use the Engineers Joint Council (EJC) system (1, 2) of indexing for the following reasons:

1. The EJC system was gaining wide acceptance among chemical industry-oriented publishers and associations, especially the A.I.Ch.E. (3, 4, 5). If desired, we could use indexes that they publish without reindexing articles. Initially, the keywords would have to be key punched using our format. However, someday it will be possible to purchase these indexes on magnetic tape as well as the pertinent computer programs.
2. The use of links and roles in the EJC system offered us a high degree of discriminating power, both during searches and during information dissemination.
3. The English language-based system appeared easy to use. Indexers did not have to learn elaborate coding schemes. New personnel we hire could be sent to the EJC training sessions conducted by Battelle Memorial Institute. The use of the EJC system could also be easily explained to those not directly involved in information work.

In addition to using the EJC system we designed the B F Goodrich information system to meet the following criteria:

1. The indexing language must be open-ended but controlled.
2. The index, or dual dictionary as we call it, must be searchable manually as well as by computer. We did not want to depend entirely on computer department schedules for all of our searches. However, since the day of the real time, shared time computer is rapidly approaching, this may not be a problem for long.
3. Our information system must incorporate a current awareness service.

* Presented before the Division of Chemical Literature, 153rd National Meeting of the American Chemical Society, Miami, Fla., April 12, 1967.

THE THESAURUS

Our thesaurus is patterned after the A.I.Ch.E. thesaurus (6). As shown in Figure 1, the final printed form nearly duplicates that of the A.I.Ch.E. thesaurus. The abbreviations used to denote the relationships between the terms are similar to those in the A.I.Ch.E. thesaurus, and therefore will not be discussed in detail. Our thesaurus has Sections 1 and 2 as in the A.I.Ch.E. thesaurus. However, because we are dealing with our own internal literature, the B F Goodrich products would make Section 2 voluminous. Thus we have placed those products which have many designations, such as Geon vinyl resins, in a separate section. We have three additional sections for the personnel, companies, and projects we deal with. Summarizing, we have,

- Section 1: General terms. Equipment, processes, properties, etc.
- Section 2: Chemicals, polymers, materials, etc.
- Section 3: Personnel list.
- Section 4: Company list.
- Section 5: Project numbers.
- Section 6: B F Goodrich products with numerous designations.

We did not attempt to draw up a list of all the terms we felt belonged in our thesaurus. Rather, the terms encountered in indexing are evaluated for inclusion in the thesaurus. We consider the value of the term as a searching aid, the predicted frequency of occurrence of the term, as well as the existing synonyms and related terms prior to inserting the term in the thesaurus. A committee of three to five persons periodically reviews the consistency and accuracy of the thesaurus.

Computer Compilation of the Thesaurus. We decided to store our thesaurus on magnetic tape for two reasons. First, the thesaurus, stored on magnetic tape and computer processed, can be used in the automatic preparation of a dual dictionary index. We refer to the thesaurus stored on magnetic tape as an "electronic" thesaurus. Second, the thesaurus requires frequent updating and printing. With the thesaurus on magnetic tape, only the new terms need be entered; the old terms previously stored are ready for subsequent printings.

All of our programs except the sort routines are written in the General Electric GECOM language. The sort routines are part of the software furnished by General Electric. The computer system we use is a GE-235, with 16K storage and six tape drives. We have a 900-line per minute printer as well as card input-output devices. Like all other computers, the GE-235 has its own peculiarities, one of which is its alphabetizing process. It handles spaces, slashes, dashes, parentheses, etc., as if they were alphabetical characters. For example, the terms "heating," "heat-agglomeration," and "heat" would be arranged in that order by the computer, rather than the order we prefer: "heat," "heat-agglomeration," and "heating." Thus, we speak of words being either in "alphabetical" order or "machine-alphabetical" order.

For printing our thesaurus, we want our terms arranged in alphabetical order. When we enter new terms into the thesaurus, we assign to each a seven-digit term code that will place it in the desired alphabetical order. We allow sufficient gap between numbers for new terms. If

there is an influx of terms having close alphabetical proximity, a program could be written to reassign new numbers to the terms either in that section or in the entire thesaurus. This has not been necessary yet.

With the seven-digit number as a base we designed the format of the card used to enter and position each term in the thesaurus.

The input card is as follows:

Columns*	Contents
1- 7	Cluster code
9	Use code
11-17	Term code
19	Deletion code
24	Thesaurus section number
25-80	Term

* All columns not accounted for are left blank.

THE CLUSTER CODE. The cluster code (columns 1 through 7) is the seven-digit number that we assign to the main term which heads a given group of terms. The cluster code is the same for all terms under the main term. For example, in Figure 1 we have clustered together "seat belts, belts, safety devices, fasteners, and straps."

SEALS

SF SEALANTS
GT GASKETS
GT O-RINGS
RT BOTTLE CAPS
RT CAPS (NOT EXPLOSIVES)
RT CLAMPS
RT FASTENERS
RT FLASHING (MATERIALS)
RT JOINTS
RT PACKINGS
RT STOPPERS
RT STRAPS

SEAT BELTS

PO BELTS
PO SAFETY DEVICES
RT FASTENERS
RT STRAPS

SEAT COVERS

PO COVERS
RT UPHOLSTERY

Figure 1. Portion of the thesaurus.

The computer lists these as a group because they have a common—*i.e.*, the same—seven-digit number in columns 1 through 7. In this case, the seven-digit number is that assigned to the term "seat belts" (Figure 2).

THE USE CODE. The use code can range in value from 0 through 8. It signals the computer to treat the term as a main term, or an SF, GT, PO term etc., the same abbreviations used in the A.I.Ch.E. thesaurus. When the thesaurus is printed, these numeric use codes are translated into the appropriate alphabetical symbols and spacing as listed below. Refer to Figures 1 and 2 for examples.

Code 0. A main term, but the computer must add "(TRADE NAME)" after the term. Otherwise, it is to be handled as a code 1 term.

THE B F GOODRICH INFORMATION RETRIEVAL SYSTEM

- Code 1. A main term. Shift term 3 spaces to the left of print position for codes 2 to 8. One line is skipped before and after this term. For example, main term SEALS in Figure 1 was code 1 in Figure 2.
- Code 2. Place "SF" (seen from) ahead of term. Example: SF SEALANTS has a code 2.
- Code 3. Place "PO" (post on) ahead of term. Example: PO BELTS has a code 3.
- Code 4. Place "GT" (generic to) ahead of term. Example: GT GASKETS has a code 4.
- Code 5. Place "RT" (related term) ahead of term. Example: RT BOTTLE CAPS has code 5.
- Code 6. Place "SEE" ahead of term.
- Code 7. Place "SEE ALSO" ahead of term.
- Code 8. This is similar to code 3. However, a term with a use code 8 does not appear in the printed version of the thesaurus, although it is recorded on the "electronic" thesaurus. Because this "Post On" term is not printed but is on the tape, we call it a "Ghost Postie." More will be said about this later.

These numeric use codes also determine the print position of terms within a cluster. This means that code 0 or 1 terms are followed in a cluster by code 2, 3, 4, 5, 6, 7 terms. Not all of these codes occur within a cluster. For example the term "Sealants" (code 2) in Figure 1 is followed by two "GT" terms (code 4) and nine "RT" terms (code 5) in the cluster for "Seals" (code 1).

THE TERM CODE. The term code is the seven-digit number that we assign to a term to place it in the desired alphabetical order. By sorting on this seven-digit number, the terms within a subcluster are thus ordered

"alphabetically" within each use code group. Note that all the "RT" terms under the word "Seals" in Figure 1 are indeed in alphabetical order.

THE DELETION CODE. Occasionally, a term is entered erroneously into the thesaurus. These terms can be deleted by punching a "1" in column 19 in addition to the cluster number, the use code, the term code, and the thesaurus section number. If this input card corresponds to an existing record on magnetic tape, the record is deleted from the tape file. A term which is to be accepted as a new valid term is punched with a "2" in column 19.

THE THESAURUS SECTION. This directs the term to be entered or deleted in the proper section of our thesaurus.

THE TERM. Here is written out the term as we want it to appear in the thesaurus. This is how the term must be written when indexing. However if the term is a trade name, the designation (TRADE NAME) is not necessary because use code 0 makes this entry.

Thesaurus Printout. In Figure 2 is shown the coding which produces the printout in Figure 1. By sorting in ascending numerical order within five specified card fields, one ends up with a file arranged by thesaurus section number (col. 24), cluster number (col. 1 to 7), use code (col. 9), term code (col. 11 to 17), and deletion code (col. 19). We merge these sorted new input term cards with the existing term records stored on magnetic tape. We are then ready to print a new version of the thesaurus. All of the sorting operations are carried out on the computer, rather than with a card sorter.

The Short Thesaurus. When the thesaurus is updated, an abridged thesaurus tape, called a Short Thesaurus,

Card Cols.	1 to 7 Cluster No.	9 Use	11 to 17 Term Code	19 Del.	24 Sect.	25 to 30 Term
	7 9 6 8 7 0 7	1	7 9 6 8 7 0 7	2	1	Seals
	7 9 6 8 7 0 7	2	7 9 5 7 4 9 8	2	1	Sealants
	7 9 6 8 7 0 7	4	3 8 1 7 4 9 9	2	1	Gaskets
	7 9 6 8 7 0 7	4	6 3 6 0 1 1 1	2	1	O-Rings
	7 9 6 8 7 0 7	5	0 9 7 7 9 3 5	2	1	Bottle Caps
	7 9 6 8 7 0 7	5	1 3 6 7 9 9 9	2	1	Caps (Not explosives)
	7 9 6 8 7 0 7	5	1 7 1 2 0 0 0	2	1	Clamps
	7 9 6 8 7 0 7	5	3 4 2 9 7 4 8	2	1	Fasteners
	7 9 6 8 7 0 7	5	3 5 5 3 5 3 1	2	1	Flashing (Materials)
	7 9 6 8 7 0 7	5	5 2 3 1 4 9 9	2	1	Joints
	7 9 6 8 7 0 7	5	6 4 0 9 0 0 0	2	1	Packings
	7 9 6 8 7 0 7	5	8 5 7 8 1 2 4	2	1	Stoppers
	7 9 6 8 7 0 7	5	8 5 9 1 2 5 0	2	1	Straps
	7 9 6 8 9 4 3	1	7 9 6 8 9 4 3	2	1	Seat Belts
	7 9 6 8 9 4 3	3	0 8 6 8 4 5 1	2	1	Belts
	7 9 6 8 9 4 3	3	7 8 0 1 7 5 0	2	1	Safety Devices
	7 9 6 8 9 4 3	5	3 4 2 9 7 4 8	2	1	Fasteners
	7 9 6 8 9 4 3	5	8 5 9 1 2 5 0	2	1	Straps
	7 9 6 9 1 8 0	1	7 9 6 9 1 8 0	2	1	Seat Covers
	7 9 6 9 1 8 0	3	2 1 3 7 7 4 7	2	1	Covers
	7 9 6 9 1 8 0	5	9 3 6 4 3 7 3	2	1	Upholstery

Figure 2. Card format for thesaurus.

is automatically prepared in machine alphabetical order for use in the dual dictionary and dissemination programs. This tape contains only those terms bearing use codes needed for subsequent preparation of the dual dictionary. More will be said about this later.

THE DUAL DICTIONARY

Dual dictionaries, used in many information systems, are lists of index terms showing the accession number of each document that contains each index term. For systems in which role indicators are used, such as the EJC system, each term is listed for each role in which it has been used. These combinations of terms and roles are conveniently called term/roles. The use of these is described elsewhere (2) and we will not go into detail here.

Computer Compilation of the Dual Dictionary. Figure 3 is a sample of the B F Goodrich dual dictionary. Note that the roles precede the terms to facilitate selection of the desired role. Note also that the document numbers are modified with alphabetical link indicators. Thus, for a proper match in multiterm searches, both the document numbers and these links must match before one is certain of a "hit."

The document numbers are arranged in a "terminal digit posting" order. That is, all documents ending in zero are in the first column, those ending in one are in the second column, those ending in two are in the third column, etc. This posting format greatly speeds up the manual coordination of document numbers.

The assembly and printing of our dual dictionary is completely automatic. The procedure is as follows:

After a document is indexed, each term/role listed on our index form is key punched. Each card contains one term, the section of the thesaurus in which the term appears, the link, the role, and the accession number of the document in which the term appears. The input card format is as follows:

Columns ^a	Contents
2	Thesaurus section number
3-56	Term
60-61	Role
62	Link
64-71	Document accession number

^a All columns not accounted for are left blank.

The terms (col. 3-56) on these dual dictionary input cards are then sorted by the computer into machine alphabetical order. The other information on each card is carried along with the term. Similarly, the short thesau-

DATE 11-30-65 BF GOODRICH CHEMICAL CO. - DUAL DICTIONARY - CONFIDENTIAL - PART 1 PAGE 100										
00	ELECTRICAL PROPERTIES									
1910A	2161A	0272A	0203A	0014A	2515A	0276D	2147H	1668A	2269A	
2271A	3931A	1192A	0763A	0034B	4135A	0006A	2417A	2416A	2569A	
2451A	3781A	1592A	0853A	0304A	4315B	1606A	2437A	2610A	3279A	
2560A	3031A	2432A	1137A	0404A		0006A	2507A	2610A	4029A	
4190A	4171B	2512A	2523A	0904A		3676B	4417A	4136A	4609A	
4140A	4671A	3872C	4143A	2604A		4650A	4707C	4208A	4909A	
5390A	4741B	4322A	5733A	3034A		5046A	5307A	4948A	5279A	
6160A	5391A	4412A	6064C	3054A		5056C			5409A	
	5441A	4922A	6373C	3074A		5075C			5409A	
		5102A		5064A		6326A			6469A	
		5402A		6024A						
		5562A		6154B						
		5742A								
10	ELECTRICAL PROPERTIES									
		0742A	8001G	0404A				3766A		
		0932C		3674A						
				5714A						
04	ELECTRICAL TAPE									
								0187A		
09	ELECTROPHOTOGRAPHIC COATING									
		5912A	4863A							
09	ELECTRIC MOTORS									
		6371A								
09	ELECTROFAX PRINTING									
						1226A	4897A			
10	ELECTROFAX PRINTING									
						1226A				
08	ELECTROLYSIS									
			4283B	0284C						
09	ELECTROLYSIS									
			4283B							
09	ELECTRON MICROSCOPE									
1100D	1751A	5912A						4007A		
10	ELECTRON MICROSCOPE									
4840A	1751A	0932D	0283A		2145A	0756A	1337A	4838A	4839A	
	4841A	5012A			3975A		4007A			
	6301A									
09	ELECTROPHOTOMETERS									
		3311F								
10	ELECTROPHOTOMETERS									
		3311F								
04	ELECTROPLATING									
				6354A						
				6354B						
				6354C						
08	ELECTROPLATING									
				6354E						
				6354F						
09	ELECTROPLATING									
				6354G						
06	ELECTROSMOSIS									
		0311C								
08	ELECTROSTATIC									
				0574C						
				3344A						
06	ELECTROSTATIC									
								3A46B		

Figure 3. Portion of dual dictionary.

rus, prepared during the production of the thesaurus, is also in machine alphabetical order.

As mentioned previously, the short thesaurus contains only terms having use codes needed for the preparation of the dual dictionary. These are use codes 0 for main terms (TRADE NAME), 1 for main terms (normal), 3 for "Post On" terms, 6 for "See" terms, and 8 for "Ghost Postie" terms. Thus the term "Seals" in Figure 1 would be entered on the short thesaurus tape by itself because SF code 2, GT code 4, and RT code 5 terms are not accepted for this tape. On the other hand, the term "Seat Belts" would be followed by "Belts" and "Safety Devices" because these are "Post On" code 3 terms.

After the dual dictionary input data and the short thesaurus are both in machine alphabetical order, the term validation, generic expansion, and synonym exchange operations begin. All of these operations are accomplished with one pass of the short thesaurus tape. Each incoming term is compared with the main terms (code 0 or 1) in the short thesaurus.

MATCHED TERMS. When an incoming term matches a main term in the short thesaurus, the seven-digit number assigned to the term is added to the incoming term. The computer deals with this number from this point on. It is easier for the computer to manipulate numbers than alphabetic characters. When terms match, the computer also scans the use codes of the terms in the same cluster as the main term. If a use code of 3 (Post On) or 8 (Ghost Postie) is found, the computer generates a new record with the accession number of the incoming term. This new record contains, in addition, the thesaurus section number, the role, and the link. A use code 6 (See) in that cluster signals the computer to exchange synonyms. The incoming term is discarded and the preferred "See" reference term is substituted.

UNMATCHED TERMS. Incoming terms without a counterpart in the short thesaurus are printed as a list of invalid terms. This list includes new terms and misspelled terms. The new terms are evaluated for possible inclusion in the next updating of the thesaurus. This term validation also serves as an error screen because we do not verify the key punching of incoming terms. All like errors are together, making correction easy.

IDIO T TERMS. Some terms are frequently entered erroneously—for example, singular forms of words instead of plural forms in the thesaurus, or chemical terms which have widely used abbreviations. Instead of having the indexer spend considerable time checking the thesaurus for correct terms, we have built an "idiot section" into our thesaurus. Through the use of our "See" reference (use code 6), we can easily handle abbreviations, frequently misspelled terms, etc. We enter the erroneous form as a main term in the "idiot section" of the thesaurus, as well as a "See" reference to the correct term. The exchange is then made automatically by the computer. These terms are accepted as input and are not printed on the list of invalid terms. We have placed the "idiot section" of the thesaurus as a post-Z alphabetical character—that is, this section of the thesaurus is printed after terms beginning with the letter Z. Thus, we can remove this section from the thesaurus before it is duplicated. Since this section is not compiled alphabetically, it is of no use for searching.

VALIDATED TERMS. As an option, a list of the validated input terms for each document can be printed. This list includes all terms for a document which have a counterpart in the short thesaurus, including the "idiot section." The original indexing form need not be retained since we have this information on the updated magnetic tape.

Dual Dictionary Printout. After the matching, generic expansion, and synonym exchange are completed we have the correct, expanded indexing terms for the dual dictionary. However, the incoming terms must be rearranged into alphabetical order by sorting on the seven-digit term numbers. The next step in our program is a shrinking operation to remove duplicate records from the tape. If one document contained information on "acetone" and "methyl ethyl ketone," both of these terms would have generated the generically higher term "ketones." Thus, we would have two records for the term "ketones" from the same document. One of these records would be removed at this stage. We are now ready to merge the new tape with the old dual dictionary tape.

This merging consists of numeric comparisons and insertions. The updated tape is now ready for printing. At this final stage, the document numbers are arranged in terminal digit posting order, mentioned previously. The dual dictionary is printed in the format shown in Figure 3.

The Bibliography. As an option to this program, a bibliography of our documents can be printed. This is a list of the document numbers, giving for each the authors, issue date, project number, title, and abstract. In a manual or a computer search of the dual dictionary, one obtains a list of matching document numbers. Abstracts of these documents can be found in the bibliography. These abstracts are used by the searcher to help him decide whether or not he should refer to the original document. The bibliography can also be listed by author, project number, or issue date, if desired. We find our author listing very useful in searches where the author is known but the subject is indefinite. Two card formats are used for this bibliographic section, as follows:

BIBLIOGRAPHIC DATA CARD:

Columns ^a	Contents
1	1 = Bibliographic data
2	Deletion code
3-56	Authors
57-60	Month, year
62-67	Project numbers
70	Type of document
72-79	Document accession number

^a All columns not accounted for are left blank.

TITLE AND ABSTRACT CARD

Columns ^a	Contents
1	2 = Title or abstract card
2	Deletion code
3-69	Title or abstract
70-71	Sequence number
72-79	Document accession number

^a All columns not accounted for are left blank.

AUTOMATIC INFORMATION DISTRIBUTION

The thesaurus program and the dual dictionary program were purposely written in modular form to allow using

sections of these for our program of automatic information distribution (AID). Our AID program has characteristics similar to those used in the SDI-4 (7) and SDI-5 (8) programs, developed by the International Business Machines Corp. We use a modification of the weighted profile system described in SDI-4, and a controlled vocabulary as is built into SDI-5. AID consists of matching index terms with personnel profiles. The index term cards described previously for dual dictionary input are also used for the AID program. We can also key punch any published EJC index and use it as an input to AID. Thus, we can selectively disseminate information on published as well as internal documents without additional indexing effort.

Profiles. Each individual who is to be served by AID prepares his own profiles. Standard profile forms are used to simplify the job. Figure 4 is a portion of our form.

0171 _____ Clamps
 0174 _____ Cleaners
 0175 _____ Clips
 0178 9 _____ Coagulators
 0179 _____ Coaters
 0186 _____ Columns (Process)
 0187 7 _____ Compactors
 0192 _____ Compressors
 0193 _____ Computers
 0196 _____ Condensers (Heat Transfer)
 0199 _____ Contactors
 0205 -9 _____ Conveyors
 0206 _____ Coolers

Figure 4. Portion of AID input sheet.

A participant is instructed to read the form and place a weight factor beside each term that interests him. The weight factors can range from +9 to -9. A weight of +9 indicates the highest interest, +5 a moderate interest, +1 a slight interest. The negatively weighted terms are used to suppress documents containing undesired terms. A -9 indicates the term does not want to be seen, -5 a less negative interest, etc.

After the form is filled out, a card is punched containing each participant's seven-digit identification number. Interest in seven terms can be coded on the same card by punching the four-digit number and the weight factors shown preceding each term on the form. A four-digit number is used to code these terms rather than the seven-digit number assigned to the term in the thesaurus, to save key punching time by putting more information on each card. These cards enter an auxiliary computer program that contains a look-up table where the thesaurus terms and their seven-digit number are matched with four-digit numbers on the profile input cards. This program is a high-speed key punch routine, which then punches

profile term cards for each term in the profile, with the following format:

Columns ^a	Contents
1	A "2" is inserted to indicate a new term card. A "1" is inserted to indicate this term card is to be deleted.
2	The number "2" indicates a term card.
5-11	Seven-digit number assigned to the participant.
12	Profile number. (Each person may have more than one profile.)
15	Thesaurus section number.
18-24	Seven-digit thesaurus term code assigned to the term.
26-27	Role indicator (optional).
29-30	Weight factor.
32-80	The term.

^a All columns not accounted for are left blank.

An additional "profile header" card must be prepared for each profile. It contains data that are common to all the term cards in the profile. Data from the header card and the term cards are later combined by the computer to speed up subsequent manipulations. The header card has the following format:

Columns ^a	Contents
1	May contain a 1, 2, or 3 with the following meaning: 1 = delete entire profile from the file. 2 = an entirely new profile. 3 = see column 14 for further instructions.
2	"1" indicates a header card.
5-11	Seven-digit number assigned to the participant.
12	Profile number.
14	May contain a 1, 2, or 3 with the following meaning: 1 = change the hit level on the old record to that on the incoming card. 2 = change spelling of participant's name. 3 = change participant's address.
15-40	Participant's name.
43-44	Participant's address.
50-51	Hit level.

^a All columns not accounted for are left blank.

The hit level (col. 50-51) is the threshold value that the sum of the weights of matched terms must equal or exceed before the participant is sent a document notification.

Personnel profiles can be adjusted every time a run is made. Thus we can quickly modify profiles. An interesting and useful "who-knows-what" file can be obtained as a by-product from the personnel profiles. By combining all the profiles and "inverting" the collection, we can print out a "who-knows-what" file—that is, we can obtain a list of profile terms containing the names of persons expressing an interest in the terms. We can reasonably expect that most persons having an interest in information containing these terms will also have some knowledge in the field. Of course, some of these persons will be newcomers to the field. However, with this knowledge file, we can direct them to persons who might know more than they do about a field of interest. Quite often information is not permanently recorded in written form. In addition, person-to-person contact is the fastest means of communication.

Computer Processing of Profiles. In our AID program, the index term cards of newly indexed documents are

first processed as in the dual dictionary program, but the new input is not merged with the old dual dictionary tape. All of the personnel profiles are sorted by term. Index terms are then matched with profile terms. When they match, a record of the hit level for the profile, the weight factor assigned the term, the seven-digit number assigned to the participant, and the name of the participant are recorded on magnetic tape. These data are sorted by participant and document numbers. For each document link, the terms which matched the profile are compiled. In the AID program, separate links within a document are treated as individual documents. The weight factors for matching terms are then summed. This sum is checked against the profile hit level. If the sum equals or exceeds the hit level, a record is written on an output tape. This record is then checked against circulation lists for that document. The circulation list on each document is punched and entered on magnetic tape. The tape is compared to check if a participant is on the circulation list. If he is, the computer will not print out this document number. We assume the participant has access to the document if he is on the circulation list; if he is not, the matching information is printed. We send him a printout containing the number and the link of the document, the terms which caused the match, and the sum of the weight factors.

COMPUTER SEARCHES OF THE DUAL DICTIONARY

At present, our literature searches are performed by manual coordination of terms in the printed dual dictionary. However, the AID program can be used to run retrospective searches as well as the automatic information distribution just described. In this case, a search question is structured as a personnel profile. If several aspects of a search question exist, each can be stated separately. The search can be made on the index terms from the newly indexed documents, as in AID for current awareness information. The search can also be made on the complete updated dual dictionary of all our documents.

However, the AID program for searches is not ideal because it was not originally designed for retrospective

searches. We are planning to write a computer program specifically designed for searches, when we decide that our volume of documents or searches requires the use of computer for efficiency. This program would be so designed that we could get a list of pertinent document numbers and/or a bibliography with or without an abstract.

SUMMARY AND PLANS FOR THE FUTURE

We have described all of the major elements of the B F Goodrich Information System—our thesaurus, dual dictionary, and automatic information distribution. We feel that this is a coordinated but flexible information system with many options. Our future plans call for computer searching. We are also planning to develop coordinated specialized thesauri and dual dictionaries for particular projects and/or locations. We continue to look for improvements to our existing computer programs.

LITERATURE CITED

- (1) Wall, E., "An Action Plan for the Engineering Profession," Panel program sponsored by the Engineers Joint Council and Section M (Engineering), AAAS, Engineers Joint Council, New York, N. Y., Dec. 1962.
- (2) Costello, J. C., Jr., "Training Manual and Workbook for use in Abstracting and Coordinate Indexing," Battelle Memorial Institute, Columbus, Ohio, 1963.
- (3) Abramson, H. I., *Chem Eng. Progr.* **60** (8), 88 (1964).
- (4) *Chem. Eng.* **72**, (21), 187 (1965).
- (5) Liston, D. M., and Halliday, T. C., *Hydrocarbon Process. Petrol. Refiner* **42** (12), 142 (1963).
- (6) "Chemical Engineering Thesaurus," American Institute of Chemical Engineers, New York, N. Y., 1961.
- (7) Kraft, D. H., "A Selective Dissemination of Information (SDI) System using an IBM 1401 Computer to Match Users Interests with Current Abstracts," IBM Corp., 618 S. Michigan Ave., Chicago, Ill. 1963.
- (8) Barnes, A. B., Briggs, A. A., Gauss, J., and Resnick, A., "SDI-5, An Advanced System for Selective Dissemination of Information," IBM Corp., Advanced Systems Development Division, Yorktown Heights, N. Y., 1964.