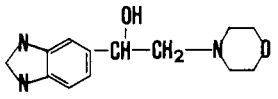
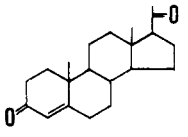
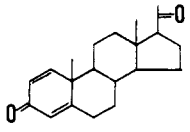
Coded compound	Punch positions erroneously absent	Output
	6/12 6/6	10082S 01 029 ISOLATED RING 014 UNSATURATED HETEROCYCLE
	2/1 25/11 25/Ø 25/7	10437S 01 035 SIDE CHAIN 044 COL. 25
	23/7	10437S 02 003 Y IN COL. 23

Figure 3.

gram. A new listprogram was written. This program allows the printing of accession number and card number of erroneous cards together with all mistakes (see Figure 3).

Contrary to a fixed check program, the procedure described here gives the possibility of changing the logical conditions very easily. Thus it is possible to consider the changing of coding rules a.s.o. without changing the program.

DRS—A User Oriented Information Retrieval System

GEZA SZONYI

Polaroid Corporation, Cambridge, Massachusetts 02139

Received December 6, 1973

DRS (Data Retrieval System), a user-oriented computerized information retrieval system, is described. The system allows data base generation and information retrieval by users without programming and systems knowledge. As an illustration, a chemical information system using the Wiswesser Line Notation (WLN) is generated, and retrieval examples are given using this data base. Because of its easily understandable English command language, DRS is suitable for a wide variety of information retrieval systems. Polaroid was instrumental in many of the advanced features of DRS developed during the past three years.

FAMULUS, a user-oriented computerized information retrieval system was presented in the fall of 1969.[1] This system, consisting of eight main and three peripheral computer programs, allows the user to set up data bases with relative ease. Having done-this, data manipulation, such as addition, correction, and deletion of information units or records, can be carried out relatively simply. Subsequently, searching of the data base by Boolean logic or by keywords is accomplished by English commands, and the retrieved subset can then be sorted alphabetically. In addition, data subunits or files can be merged, and printouts of the retrieved subset can be displayed in several output formats.[2] The system was mainly designed to handle personal reference collections but appeared to have applicability for all types of computerized information retrieval systems.[3] Unfortunately, the system was designed for IBM System/360 Control Data Corp. 6400 and 6600, UNIVAC 1108, or Xerox Sigma 7 type computers.[1,2] These are large machines and not particularly suited for operation by individual users without system and programming knowledge. Also, searches with this system are carried out in the batch mode, rather than interactively, making the information retrieval rather awkward. A more serious drawback of FAMULUS is its magnetic tape orientation; *i.e.*, searching is done sequentially rather than at random. In addition, typical data bases contain 3000–5000 records, each record having about 4000 characters.[2]

We were interested in FAMULUS-type retrieval systems, *i.e.*, having all the attributes of such a system, but one which could be implemented on minicomputers, since these could potentially be operated by single users without system or programming staff support. Further, data bases

should be randomly accessible using disks or drums, rather than sequentially accessible, using tape. A survey of the literature during 1970 revealed that only a few computer program packages existed having these attributes.[4] From these, we selected Data Retrieval System (DRS), distributed by the Aeronautical Research Associates of Princeton, Inc. (ARAP).[5] Our selection of this computer software was based on its versatility and simplicity. English is used for all commands so that no programming knowledge is required, and the users can operate this system by themselves. Searching is carried out using randomly accessible disk cartridges or disk packs, rather than the sequentially accessible magnetic tapes. Further, it could be implemented on our IBM 1130 16K computer, and it had the one-time price of $2800.00 for the basic system. DRS was later implemented on the General Automation 18/30 and on the Digital Scientific Meta 4 minicomputers. Here, the price for the basic system is the same as mentioned above. Recently, this system was implemented in a non-minicomputer environment using the IBM System/370 OS/VS and VM/CMS computers. This version is purchasable for $9500.00 and $15,500.00, respectively. Negotiations are currently in progress to make DRS available through commercial time-share houses. Also, ARAP will provide, on request, time-sharing or remote job entry (RJE) access to DRS.

Within Polaroid, DRS has proven suitable for the organization, retrieval, and reporting of information of various kinds, such as library journal circulation, laboratory data, reports, collections of individual reprints, etc. Figure 1 gives representative list of information retrieval systems implemented. Data input into DRS can be carried out in fixed field, free field, or mixed fixed-free field formats, using textual, numerical, or calendar data.[6] As in FAMULUS, the data unit is the record, and DRS allows up to 250 subunits of fields per record. The maximum length of a record in DRS is 5000 characters. Depending on the number of characters per record, 4000–8000 records can be accomodated on the average per a 0.5-megabyte disk platter.[8] Retrieval operations can be carried out interactively from a computer console or in the batch mode, using punched cards. The former is usually employed for exploratory runs, the latter for standard repetitive runs. Sorting of the retrieved data subset can be carried out on numerical or textual fields.[6] Data manipulation, such as adding, modifying, or deleting whole records or individual fields is simple and quick. Protection of the whole data base or part thereof is possible, using passwords. DRS also supplies the user with straightforward, easily understandable error messages if incorrect command syntax is used.[9] Provisions are also made to prevent the input of incorrect data (in the DRS sense) to any data base.[6]

We have been working in close cooperation with ARAP since 1970, and since that time the DRS software has been expanded considerably. Many of the improvements made were added at our suggestions. One of these is the LINK modules allowing the user to add individualized features to the DRS system.[7] Thus, specific computational commands may be added, useful for financial reports, statistical analysis, etc. LINK modules go beyond the basic DRS system, requiring Fortran IV programming knowledge from its users. The price for the LINK module capability is $250.00. Another innovation is the extension of the sort (ARRANGE) command. By inserting into any data base specified lists of NOISEWORDS (the, and, not, by, etc.) or THESAURUS words (selected keywords), text analysis can be carried out on individual fields or merged groups of fields.[6] With NOISEWORDS, the analysis selects all noniseword words from a text, whereas, with THESAURUS words, only those words are selected which match the words of the thesaurus list. Thus, text analysis can serve to generate an automatic keyword list from the input data. Obviously, one can have keywords or key phrases as record fields and use these alone or together with the computer generated keywords for information retrieval. A major improvement of the system is the DRS System Generator,[8] obtainable for the one-time charge of $350.00. This allows the generation of DRS-based information retrieval systems from a set of simple instructions, again requiring no programming knowledge. Prior to that, the knowledge required for the implementation of a specific retrieval system was considerably more than the user wished to obtain and was done by persons having DRS system knowledge or by ARAP on a contractual basis with a charge of $10.00 per field. For this a deck of punched cards was obtained, which could be loaded by the user on the computer, following simple instructions. Subsequently, appropriate data could be loaded, and the data base in question was operational.

At present, DRS can be used at two major levels. The user who does not want to be concerned with systems or programming work can generate a retrieval system using the DRS System Generator[8] and operate the thus generated system using the simple English commands described in the DRS Users' Manual.[6] The more sophisticated user can carry out several manipulations on the data base generated, design individualized output formats[10] or build LINK modules[7] fitting specific requirements. Thus, DRS provides a wide latitude for a variety of users.

As an example of the application of the DRS system, at its present level, the generation and usage of a chemical information retrieval system based on the Wiswesser Line Notation (WLN) will be outlined. This will also serve to illustrate one of the main advantages of this system, namely, that a single individual, without computer programming or systems knowledge, can design and implement a DRS-based information retrieval system.

We selected as input data for this projected data base the "Educator Deck" distributed by W. J. Wiswesser,[11] based on the easy availability of an already punched card deck. This deck contains 1,855 compounds selected at random from the Merck Index.[12] Each punched card has the following information, punched in fixed field format:

Card columns	Information
1–10	WLN notation
13–32	Shortened computer name; registered names or synonyms were enclosed by asterisks
35–38	Alphanumerical serial number used to sort the deck by parent names, inverted names, or shifted names
40–43	Sequence designation, which contains codes for aliphatic chains, branched chains, benzene derivatives, cyclic compounds, etc.
44–45	Special marks, not further specified
46–51	Batch numbers, *i.e.*, structural classification designation[13]
53–69	Molecular formula.
71–72	Highest ranking WLN rule, applicable to the compounds in question.
74–79	Merck Index page and location on the page

Out of this information, we used only the following entries for our data base: WLN Notation, Shortened Name, Molecular Formula, and Merck Index page and location. Each of these entries presents a "field" in DRS terminology, the totality of the four fields pertaining to a specific compound, a "record."

The data base was implemented on a 16K General Automation 18/30 Computer, interfaced with a high-speed line printer (600 lines per minute), a disk drive, a magnetic tape drive, a card reader, a card punch, a teletype, and a Selectric typewriter. The disk drive employs a Wright Line 1361-00 or similar five-platter disk, having 2.5-megabyte capacity, *i.e.*, 0.5 megabyte per platter. This computer has replaced our IBM 1130 computer.

Name	Purpose	Remarks
LIBRARY	To obtain a computerized routing list for the circulated library journals.	In addition to preparing routing lists, this data base also gives a listing of journal holdings, the number of journals circulated to any individual, the listing of all individuals on the circulation list, etc.
STATREP	To organize an individual reprint collection in the area of applied statistics and mathematics.	Reprints are organized by authors, titles, date of publication, keywords, location (file cabinets and files proper), author affiliation or address, reference designation (open or government literature), abstracts and remarks.
REPORT	To organize research reports.	Research reports are organized by their authors, titles, keyphrases, date of issuance, accession number and number of pages.
LABORATORY	To organize laboratory data.	Testing data were organized by their various attributes using up to 64 fields per record.
PHOTOKWOC	To provide an index to selected photographic articles.	Two output formats are produced: a table of contents type bibliography and a KWOC (Keywords-out-of-Context) index.
WLNDATA	To organize a chemical information retrieval system, using the Wiswesser Line Notation (WLN) for structure description.	Information pertaining to chemicals listed in the Merck Index are organized by WLN notation, name, molecular formula, Merck Index page, remarks, boiling point, melting point, alternate name, synthetic information, color, uses, molecular weight, solubility and literature reference.

Figure 1. Implemented information retrieval systems using the DRS Computer Program package.

To set up the desired data base, the following procedure was used:

1. A new disk pack was initialized, using a card deck supplied by the computer manufacturer and following standard simple instruction.[14] This will give each platter on the disk its unique identification number (ID number).

2. Deposit onto any two platters of a five-platter disk pack the software corresponding to the DRS Basic System and to the DRS System Generator, respectively. This can be done using magnetic tape or appropriate card decks.

3. Using the DRS System Generator,[8] design the cards needed for data base generation. In particular, design a master card (Card Type 1) containing the data base name, WLNDATA; its abbreviated name, WLN; the satellite data base name and abbreviated name; TWLDATA and TWL; ID of the disk platter on which the data base will be located; and indications that a NOISEWORD list will be used and that all information connected with the data base generation should be printed out. (The satellite data base is located on a different disk platter than the data base proper and serves to extend the information handling capacity of the data base.) Secondly, a submaster card (Card Type 2) is designed, containing the designated main heading for the data base: WLN EXPERIMENTAL DATA BASE. This heading will appear on each printout of this data base. Thirdly, design a card (Card Type 3), which indicates that data input will be in fixed field format. This card is omitted for data bases where data input occurs in free field format. Fourthly, design a card (Card Type 4) for each field considered within a record. It is customary to design more fields than those present in the original data. Such fields serve to expand the information content of the data, if needed. Field names are expressed in the DRS system by two-character symbols: the first character is always alphabetic; the second alphabetic or numeric. Card Type 4 cards contain designation whether data input is by fixed or free field; field names; description of the nature of the field (textual, calendar date, integer or real numbers, or definable, *i.e.*, used to store simple computations performed on numerical fields); scanning depth, *i.e.*, how many sequential alphanumeric characters will be considered in scanning this field in searching and sorting operations; field width expressed in a number of characters and indication of multiple lines for field width, if desired; and, field headings. (If these are left empty, headings can be put in interactively.) As an illustration, the card designed as

Card column:	1	2	3	4	5	6	7	21	23	24	34	35	41	42	43
Punch:	4	1	W	L	T	1	2	1	1	2	1	2	W	L	N

indicates (reading from left to right) that it is a Card Type 4 card, that data input is fixed field, that the field name is WL, that it is a textual field, that scanning depth is 12 characters, that information for this field is to be read from the data card from column 1 to column 12, that the field width is 12 characters and the field heading is WLN. The complete design for this system is presented in Figure 2, showing how simple and easy it is to design an information retrieval system.

4. Assemble the following card deck:

```
Cold Start Card
// JOB 0100  0101  0102  0103  0104
// XEQ IRDBG
Cards punched out as per Figure 2
```

This deck is put in the card reader and read into the computer, with the appropriate disk in the disk drive (having the ID numbers given above) and the card punch in readiness. (In

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80				
1 WLNWLNDATA TWLTWLNDATA				Y Y 0102
2				WLN EXPERIMENTAL DATA BASE
31	0/1			
41WLT12	1 12	12		WLN
41NAT19	1433	19		NAME
41MLT13	5365	13		MOLECULAR # FORMULA
41MET7	7480	6		MERCK # INDEX # PAGE
4 RET20		520		REMARKS
4 BPR		6		B.P.
4 MPR		6		M.P.
4 ANT20		520	B/	ALTERNATE NAME
4 SIT30		730	B/	SYNTHETIC INFORMATION
4 COT20		20		COLOR
4 UST30		530	B/	USES
4 MWR		6		MOL. # WT.
4 SOT20		520	B/	SOLUBILITY
4 RFT20		520	B/	REFERENCE
4 III Y				
4 IJI Y				
4 R1R Y				
4 R2R Y				
4 XXT20		520		
4 YYT20		520		
7/9				

Figure 2. Design and card layout for the data base WLNDATA.

Commands	Explanations
S,A,L,E,OR O? OP:WLNDATA .;	S,A,L,E,OR O? Stands for SELECT(S), ARRANGE(A), LIST(L), EXECUTE(E), or OTHER(O) commands. The underlined statements represent user's responses to the system's questions. OP(OPEN) - This opens the data base and selects automatically all the information contained therein.
S,A,L,E,OR O? S:WL@'SW'|WL@'WS'&ML<='C20';	S(SELECT) compounds containing SW or WS in WLN notation, i.e. $-SO_2-$ groups and having in their molecular formula, less than or equal 20 C atoms.
NUMBER THAT QUALIFY = 66	System response. Out of 1855 compounds, 66 comply with the above search requirements.
S,A,L,E,OR O? D:¬WL@'R';	Command: D(DEFINE) used here for subselection. Command: not WL contains R, i.e. remove from subset selection those compounds containing within the WLN notation the character R, i.e. benzene rings.
NUMBER OF RECORDS EXAMINED = 25	System response. Out of 66 compounds selected, 25 comply with the subselection requirements.
S,A,L,E,OR O? D:¬WL@'T';	2nd subselection: Remove all compounds with heterocyclic rings.
NUMBER OF RECORDS EXAMINED = 20	System response. Out of 25 compounds subselected, 20 comply with the second subselection requirements.
S,A,L,E,OR O? A:ML/F;	A (ARRANGE)(sort) by molecular formulas forward, i.e. from lowest to highest.
S,A,L,E,OR O? H:H2='WLNDATA - COMPOUNDS CONTAINING SULFONYL GROUPS#NOT CONTAINING BENZENE OR HETERO RINGS#ARRANGED BY MOLECULAR FORMULAS';	Input a subtitle.
S,A,L,E,OR O? H:DT='JUNE 27, 1973';	Input a date, for documentation as to when this run was carried out.
S,A,L,E,OR O? L:ML,WL,NA,ME;	L(LIST) on the printout in the LIST output format, in this sequence: ML (molecular formula), WL (WLN), NA (name) and ME (Merck Index page and location).
S,A,L,E,OR O? E:EJ,SK3;	E(EXECUTE) -Print onto a new computer page, and skip 3 lines between each record.
S,A,L,E,OR O? Y	Expedited exit from the system. Used for fast logout.
YOU ARE NOW LOGGED OUT OF DRS. THANK YOU FOR CALLING.	

Figure 3. Commands used to obtain compounds containing sulfonyl groups and not containing benzene or heterocyclic rings.

other computer configurations, the above Cold Start and JOB cards are replaced by appropriate system cards.)

5. This results in the partial generation of the data base, printout of appropriate documentation, and production of punched cards. The latter are interpreted, preceded as in step 4 by a Cold Start and JOB card, put into the card reader, and read into the computer. This completes the data base generation. The DRS System Generator supplies two output specifications, *i.e.*, formats in which the retrieved data are printed out. One lists the field headings side-by-side, with the field contents below in the conventional columnar mode (LIST format). The other lists the field headings with their field contents below one another, each content going horizontally across the page (DOCLIST format).

6. At this point, one adds a NOISEWORD list to the data. Since this procedure is described in detail in the DRS Users' Manual,[6] it will not be discussed here further.

7. Having done this, the actual data, *i.e.*, the Educator Deck, are placed in the card reader and the data read into the computer. The commands for adding the initial data to the data base are described in detail in the DRS Users' Manual,[6] as is the procedure for adding data at a later time. The data base is now operational.

It should be pointed out that the total procedure described required approximately 3 hr, including disk pack initialization, deposit of the two basic softwares, data base design and generation, and finally data input. All aspects of the above procedure are described in suitable DRS Manuals,[6,8,10] and the details given here merely serve to indicate the simplicity and straightforwardness of these operations.

To illustrate the search capabilities of the system, the generated WLNDATA data base will be used. A search was carried out for compounds having within their structure sulfonyl groups, $-SO_2-$, and containing neither benzene nor heterocyclic rings, and also having 20 or less carbon atoms. The commands for this search in their interactive form are presented in Figure 3. The commands for the batch mode are identical with those used interactively, but instead of typing the commands on the Selectric typewriter, they are punched onto cards. The printout corresponding to the search results is presented in Figure 4. Only a partial printout is given. It should be mentioned that the total time for this search was about eight minutes interactively and about four minutes for the batch operation. In another search, compounds were retrieved containing within their structure a pyridine ring and having 15 or less carbon atoms. These were sorted by every symbol within their WLN structure, separated from each other by blank spaces. The

WLN EXPERIMENTAL DATA BASE
WLNDATA - COMPOUNDS CONTAINING SULFONYL GROUPS
NOT CONTAINING BENZENE OR HETERO RINGS
ARRANGED BY MOLECULAR FORMULAS

SELECTION: WL@'SW'!WL@'WS'&ML<='C20';
DEFINITION: ¬WL@'T';

MOLECULAR FORMULA	WLN	NAME	MERCK INDEX PAGE
C H 3 I O3	WSQ1I	METHIODAL	M0666B
C H 4 O3 S	WSQ1	ME -SULFON&ACID	M0663G
C H 4 O4 S	WSQO1	ME -SULFUR&ACID	M0683G
C H 4 O6 S2	WSQ1SWQ	METHIONIC ACID	M0666D
C 2 H 4 O5 S	WSQ1VQ	SULFO/ACET&ACID	M1000B
C 2 H 6 O4 S	1OSWO1	ME2 SULFATE	M0373A
C 2 H 6 O4 S	WSQ2Q	ISETHIONIC ACID	M0569C
C 2 H 6 O4 S	WSQO2	ET -SULFUR&ACID	M0436E
C 2 H 6 O6 S2	WSQ2SWQ	1-2-ET -DISULFON&A	M0419G
C 2 H 7 N O3	Z2SWQ	TAURINE	M1013D
C 2 H 9 N O4	WSQO2 &ZH	AMM/ ET-SULFATE	M0066C
C 3 H 7 N O5	WSQ1YZVQ	CYSTEIC ACID	M0316A
C 3 H 9 N O3	WSQ2M1	N-ME -TAURINE	M0684A
C 4 H10 O3 S	WS2&2Q	2-/ET -S-&/-ET OL	M0436D
C 4 H10 O4 S	2OSWO2	ET2 SULFATE	M0355D
C 5 H 9 N O2	WS1&3NCS	CHEIROLIN	M0230B
C 6 H14 O6 S2	WS1&O 24	BUSULFAN	M0175B
C 7 H16 O4 S2	WS2&X&&SW2	SULFONMETHANE	M1001C
C10 H 8 O3 S	L66J BSWQ	A&-NAP SULFON&ACID	M0703C
C10 H 8 O3 S	L66J CSWQ	B&-NAP SULFON&ACID	M0703D

Figure 4. Computer printer output.

WLN EXPERIMENTAL DATA BASE
WLNDATA - COMPOUNDS CONTAINING PYRIDINES
WITH MOL. FORM. LESS THAN C15
ARRANGED BY WLN FRAGMENTS AND MOL. FORM.

SELECTION: WL@'T6NJ'&ML<='C15';

WLN FRAGMENTS PERMUTED	WLN	MOLECULAR FORMULA	NAME	MERCK INDEX PAGE
B	T6NJ B	C 6 H 7 N	A&-PICOLINE	M0816B
	T6NJ B F	C 7 H 9 N	2-6-LUTIDINE	M0620B
	T6NJ B O2	C 8 H11 N	4-ET PICOLINE-2	M0435A
	T6NJ B E2	C 8 H11 N	5-ET PICOLINE-2	M0435B
	T6NJ B D F	C 8 H11 N	246-ME3-PYRIDINE	M1068F
	T6NJ B- ?	C10 H 8 N2	&A -DIPYRIDYL	M0385F
BVQ	T6NJ BVQ	C 6 H 5 N O2	PICOLINIC ACID	M0816E
BX	T6NJ BX FX	C13 H21 N	26-TBU2-PYRIDINE	M0343C
BZ	T6NJ BZ	C 5 H 6 N2	A&-AMI-PYRIDINE	M0060A
	T6NJ BZ D	C 6 H 8 N2	2-AMI-PICOLINE-4	M0058A
C	T6NJ C	C 6 H 7 N	B&-PICOLINE	M0816C
CS	T6NJ CS 2	C10 H 8 O3 S	3-3 S2 -DIPYRIDINE	M0370C

Figure 6. Computer listing.

commands used for this search are presented in Figure 5, and a partial printout of the search results is given in Figure 6.

As seen from these figures, information retrieval with the DRS systen can be accomplished using a small number of easily understandable English commands. The user only needs, in addition to these, commands for data addition, data correction or extension (modification), and data deletion, all described adequately in the DRS Users' Manual,[6] to operate a data base. The retrieval may be simple, using only a restricted number of commands, or highly sophisti-

Commands	Explanations
S,A,L,E,OR O? S:WL@'T6NJ'&ML<='C15';	Selects compounds containing pyridine rings with molecular formula less than 15 carbons.
NUMBER THAT QUALIFY = 35	System response. Out of 1855 compounds, 35 comply with the above search requirements.
S,A,L,E,OR O? A:WL/(WF,SF,B4),ML/F;	A (ARRANGE) the WLN notation by every character string separated by blanks; do not print the identical fragments retrieved more than once, skip 4 lines wherever the character string composition changes. Make a subarrangement for all records having identical fragments by their molecular formula, i.e. from the lowest to the highest.
S,A,L,E,OR O? H:H2='WLNDATA - COMPOUNDS CONTAINING PYRIDINE#WITH MOL.FORM.LESS THAN C15# ARRANGED BY WLN FRAGMENTS AND MOL. FORM.';	Insert heading.
S,A,L,E,OR O? H:H3='WLN FRAGMENTS# PERMUTED';	Insert heading for the arranged WLN fragments.
S,A,L,E,OR O? L:SWD,WL,ML,NA,ME;	$WD is the designation of the internal field generated by DRS in which the special permuted WLN fragments are stored.
S,A,L,E,OR O? E:EJ,SK2;	Execute (print) on a new page and skip two lines between each record.
S,A,L,E,OR O? Y	Expedited exit from the system.

YOU ARE NOW LOGGED OUT OF DRS. THANK YOU FOR CALLING.

Figure 5. Commands used to obtain compounds containing pyridines with molecular formula less than C15 arranged by WLN fragments and molecular formula.

Commands	Explanations
S,A,L,E,OR O? S:EX(RE)|EX(BP)|EX(MP)|EX(AN)|EX(SI)|EX(CO)|EX(US)|EX(MW)|EX(SO)|EX(RF);	Select compounds containing information regarding remarks, boiling point, melting point, alternate name, synthetic information, color, use, molecular weight, solubility, reference.
NUMBER THAT QUALIFY = 120	System response. Out of 1855 compounds, 120 comply with the above search requirements.
S,A,L,E,OR O?L/A:ML/F;	A (ARRANGE) by molecular formulas forward, i.e. from lowest to highest. Typical error made: LA was typed instead of A.
LA: 00 404 NO SUCH COMMAND NAME 0	
S,A,L,E,OR O? A:ML/F;	Repeat the above command, this time in correct syntax.
S,A,L,E,OR O? L:DOCLIST:WL,NA,ML,ME,RE,BP,MP,AN,SI,CO,US,MW,SO,RF;	List the fields mentioned in DOCLIST format in the sequence: WLN, name, mol.formula, Merck Index Page, Remarks, boiling point, melting point, alternate name, synthetic information, color uses, mol. wt., solubility, reference.
S,A,L,E,OR O? H:H2='WLNDATA - COMPOUNDS CONTAINING ALL THE INFORMATION';	Insert subheading.
S,A,L,E,OR O? E:EJ,SK2;	Execute on a new page and skip two lines between each record.
S,A,L,E,OR O? Y	Expedited exit from the system.

YOU ARE NOW LOGGED OUT OF DRS. THANK YOU FOR CALLING.

Figure 7. Commands used for obtaining compounds containing all information regarding remarks, boiling point, melting point, alternate name, synthetic information, color, use, molecular weight, solubility, reference, etc.

```
                                WLN EXPERIMENTAL DATA BASE
                    WLNDATA - COMPOUNDS CONTAINING ALL THE INFORMATION
SELECTION: EX(RE)|EX(BP)|EX(MP)|EX(AN)|EX(SI)|EX(CO)|EX(US)|EX(MW)|EX(SO)|EX(RF);

                  WLN - GVG
                 NAME - PHOSGENE
      MOLECULAR FORMULA - C    CL2 O
       MERCK INDEX PAGE - M0808D
              REMARKS -
                 B.P. -
                 M.P. - 138
        ALTERNATE NAME - METHACETIN
  SYNTHETIC INFORMATION -
                COLOR -
                 USES -
               MOL. WT. - 165.19
           SOLUBILITY - ALCOHOL,ACETONE,  CHLOROFORM
            REFERENCE -


                  WLN - VHH
                 NAME - FORM/ALDEHYDE
      MOLECULAR FORMULA - C   H 2 O
       MERCK INDEX PAGE - M0460B
              REMARKS -
                 B.P. -
                 M.P. -
        ALTERNATE NAME - METHANAL,OXOMETHANE,FORMIC ALDEHYDE,METHYL ALDEHYDE
  SYNTHETIC INFORMATION - INCOMPLETE COMBUSTION MANY ORGANIC SUBSTANCES
                COLOR -
                 USES -
               MOL. WT. - 30.03
           SOLUBILITY - WATER
            REFERENCE - U.S.2,812,309(1957)/2,849,492(1958)


                  WLN - VHQ
                 NAME - FORMIC ACID
      MOLECULAR FORMULA - C   H 2 O2
       MERCK INDEX PAGE - M0461C
              REMARKS -
                 B.P. -
                 M.P. -
        ALTERNATE NAME - AMEISENSAURE
  SYNTHETIC INFORMATION - HEATING CARBON MONOXIDE&NAOH UNDER PRESSURE&DEOMPOSING RESULTS WITH H2SO4
                COLOR -
                 USES -
               MOL. WT. - 46.02
           SOLUBILITY -
            REFERENCE -
```

Figure 8. Computer listing.

cated using additional commands or enlarging the content of the commands used. An example of the latter is the ARRANGE (A) command used in Figure 5.

We subsequently extended the information contents of this data base by adding to existing records information which we considered pertinent for a chemical information retrieval system. These included boiling point, melting point, alternate name, remarks, synthetic information, uses, molecular weight, solubility, and literature reference of compounds. This information expansion was carried out for only 120 of the 1855 compounds, using the MODIFY (MO) command described in the DRS Users' Manual.[6] Typical data cards used in the MODIFY procedure are given below since these illustrate the concept of free field data input.

```
Card                                                                    Card
Column 1                                                                Column 72
*$k=961*BP=164*AN=LYMPHOCIN*SO=ACTION OF PHOSPHORUS PENTACHLORIDE ON    (Card 1)
N,N-BIS-ANILINE*US=CANCER RESEARCH*MW=218.12*SO=HOT METHANOL,ETHANOL    (Card 2)
*RF=J.CHEM.SOC.1934,1538/J.GEN.CHEM.(U.S.S.R.)14,312(1944)*;            (Card 3)
```

As seen from these, in free field data input, asterisks are used as field separators, and semicolons as record terminators. Fields are listed by their two-character symbols, followed by an equal sign and then by the field contents. The first field above is: $K, *i.e.*, record key, required for the MODIFY operations, since it is a unique numerical identifier given automatically to each record by DRS. It can be listed just like any other field.

After the modifications were completed, a printout of all the modified records was produced. The appropriate commands are shown in Figure 7; a partial printout is given in Figure 8. A typical error message is also shown in Figure 7.

The amount of data which can be stored with the DRS system will be dependent on the computer and on the disk pack used. In our configuration (General Automation 18/30 computer and with a Wright Line 1361-00 or similar five platter disk pack) and for the designed WLN Experimental Data Base, about 7500 records per disk platter can be stored. This means that one can store approximately 30,000 records per disk pack. (The DRS System Generator can be removed from the disk pack after the system generation is completed.) The number of records which one can store is approximate, since the actual number depends on the amount of information contained within each record. The information contents of a data base can be enlarged further by storing data on additional disk packs or on magnetic tapes. Searching in such cases is performed sequentially, *i.e.*, either using one disk pack at a time, or by transferring just enough data from a tape each time to fill a disk pack. The searching medium in both cases is the randomly accessible disk.

ACKNOWLEDGMENT

I would like to express my thanks to Mrs. Diane Farmer of Polaroid for her help and suggestions in connection with setting up this system. Thanks are also due to my manager, Mr. Charles Zerwekh, Jr., for the support and encouragement given and to the Polaroid Corporation for allowing me to publish this work.

LITERATURE CITED

(1) Meeting of the American Society for Information Science, Oct 1–4, 1969, San Francisco, Calif.
(2) Yerke, T. B., *et al.*, FAMULUS: A Personal Documentation System . . . Users' Manual, Pacific Southwest Forest and Range Experiment Station, Berkeley, Calif., 1969, PB-202 534.
(3) Rase, W. A., *Angew. Informatik*, **14**, 459–465 (1972).
(4) Falor, K., "Modern Data," March 1970, 62–72; Aug 1970, 48–59.
(5) DRS is a proprietary information retrieval software, developed by the Aeronautical Research Associates of Princeton, Inc. (ARAP).
(6) DRS Users' Manual, distributed solely by ARAP.
(7) The DRS LINK Command, distributed solely by ARAP.
(8) DRS System Generator, distributed solely by ARAP.
(9) DRS Error Messages, distributed solely by ARAP.
(10) DRS System Manual, distributed solely by ARAP.
(11) CWIK Note, Aug 1967, and the corresponding card deck.
(12) "Merck Index," 7th ed., Merck & Co., Inc., Rahway, N. J., 1960.
(13) Barnard, A. J., Kleppinger, C. T., and Wiswesser, W. J., *J. Chem. Doc.*, **6**, 41–48 (1966).
(14) IBM System Reference Library, IBM 1130 Disk Monitor System, Version 2, Programmer's and Operator's Guide, 10th ed, May 1972.

Performance of an SDI System with Interactive Features

VIERA ŠAŠKOVÁ

Institute of Inorganic Chemistry, Slovak Academy of Sciences, Bratislava, Czechoslovakia

JIŘÍ KOSÍK

Economic Research Institute of Chemical Industry, Bratislava, Czechoslovakia

Received March 7, 1974

A user-oriented interactive system was developed and tested. For two years, 280 profiles were searched in the *CA-Condensates* data base. The performance and effectiveness of the system were evaluated in relation to the data base used, to the hardware configuration and software package, to the user population, and, finally, in relation to the aid offered by the information center. Various ways and means that lead to a better satisfying of the user's needs, such as the iterative way of searching, quantification of user's needs, searching at various specifity levels, etc., are discussed.

It has been generally accepted that the aim of all information systems is to bring the right information to the right man at the right time. In other words, the primary aim of an information system is the satisfaction of the user's needs. Emphasis on the needs of the user in the information process, in contrast to the previous stressing of the document-handling side, is far more than a pure theoretical or terminological question. This shifting of attention brought about great changes in information handling; new methods were developed with the aim of increasing the adaptability and flexibility of information systems. Interactivity of information systems, man-machine dialogues, machine-aided formulation of the user's profiles, all these were introduced in order to respect the user's wishes and to facilitate his search for information. On the other hand, it was found how little we know about the user, his personality, background, and his literature habits.

Summing up the above, a modern information system should fulfill the following requirements:

1. As a response to the user's request, it should give a relevant and complete set of information.

2. In relation to the user, the system should be active—it should, *e.g.*, be able to make suggestions and point out the errors committed in the profile formulation. It is further desirable that the formulation of profiles be machine-aided and the alterations of profiles be easily performed.

We tried to comply with the above requirements and introduced at least some of the above specified features into the system developed in cooperation with the Institute of Inorganic Chemistry of the Slovak Academy of Sciences and of the Economical Research Institute of Chemical Industry in Bratislava. The system ran under the working name CACS. For some 15 months about 280 profiles were matched against the *CA-Condensates* data base, and a current awareness service was supplied for the users.

We evaluated the performance of the system and tried to find the best ways of satisfying the user's needs. Emphasis was put on the interface between the user and the system. From the analysis of the performance, we concluded that the factors influencing the effectiveness may belong to four major categories:

(1) The data base that is searched
(2) The information system, the software package, and hardware configuration
(3) The personality of the user and his needs
(4) The information center, their assistance given to the user

DATA BASE

CA-Condensates in SDF were searched. The advantages of an external ready-made reference service are obvious: results of the work of many highly qualified abstractors, indexers, and editors are at our disposal; we need not analyze the primary literature. The disadvantage of such an external reference service is that, even if it does not suit us, we have to accept the indexing mode of the outside service organization. There is no doubt that a data base prepared by