Illustrative of these terms are "polyesters," "polyethers," "polyesteramides," and "polyetherurethanes." These terms are assigned unique roles which distinguish between nonmodified and modified polymers. Examples are shown in Figures 5, 6, and 7.

Use of the term polyesteramides in Figure 5 prevents retrieval of the patent when we are searching for polyesters from adipic acid and ethylene glycol and propylene glycol. In the latter case, we would search under the terms given in Figure 6.

The system we have described does not eliminate all false drops. It does not give the selectivity that links give. However, it is a relatively simple system which is amenable to computer registration and it affords considerable selectivity in searching.

LITERATURE CITED

(1) Montague, B. A., and R. F. Schirmer, "Du Pont Central Report Index: System Design, Operation, and Performance," J. CHEM. DOC. 8, 33 (1968).

The Atlas Chemical Research Information System*

FREDERIC R. BENSON and ROGER A. WALCK
Atlas Chemical Industries, Inc., Wilmington, Del. 19899

Received January 29, 1968

A punched card system is described for recording, organizing, and retrieving chemical research data. For each research or reference compound, provision is made for recording identity, constitutive properties, analytical data, physical data, results of tests of utility, and compound novelty. The system is open-ended and new properties or tests can be added easily. Input equipment is the IBM 870 Document Writer which also is used for low-volume, variable output. High-volume output is produced by computer. In addition to data on individual compounds, composition, preparation, and performance data of multi-component preparations such as polyurethanes also are recorded on punched cards. System output includes semimonthly reports of new compounds and use test results, cumulative use test results organized both by use test and by compound class, chemical class reports, physical properties, novelty data, and combinations of these data.

An automated system has been used to record, organize, and retrieve chemical research information at Atlas Chemical Industries, Inc., since 1962. Most of the information in the file is internal research and development data. A small portion of the data concerns non-Atlas compounds, usually reference materials for tests of utility. Input is by punched cards; data is manipulated by both punched card equipment and computer.

SYSTEM CHARACTERISTICS AND INPUT

The system is organized around the individual chemical and its properties. In general, research data are recorded rather than references to data location. The system is open-ended with a variable number of punched cards for each compound. Presently, there is an average of about 15 cards per compound. Advantages of a multiple card system have been described (*1*). Each card carries specific data, a card code, and a compound number.

Chemical information is grouped into several categories as shown in Table I. There are presently over 350 card types, each identified by a card code. Figure 1 shows several punched cards pertaining to a particular compound, illustrating specific card types. New types of information are added to the system by creating new card types.

*Presented in the Symposium on "Automation of Information Operations," ACS Middle Atlantic Regional Meeting, Philadelphia, Pa., February 1, 1968.

Included in the general category, identity and general information, are the name of the compound, data on alternate samples, and results of literature searches for the compound. Each chemical in the system is identified by a 9-digit compound number. Numbers for Atlas compounds consist of a digit identifying the department of origin, the laboratory notebook number and page where preparation is recorded, and, if necessary for differentiation, the position on the notebook page.

0	1472	037	0
indicates dept. or source of compound	note-book no.	note-book page	distinguishing mark

Location of the record is thus immediately known, and the identity of the chemist who originally prepared the compound is easily ascertained. Non-Atlas organic

Table I

	Examples	
General Category	Specific information	Card code
Identity and general information	Name	A001
Constitutive properties	Molecular formula	B100
Analytical tests	Hydroxyl no.	D150
Physical properties	Specific gravity	C124
Chemical and physical use test screening results	Pharmaceutical tablet lubricant	GA10

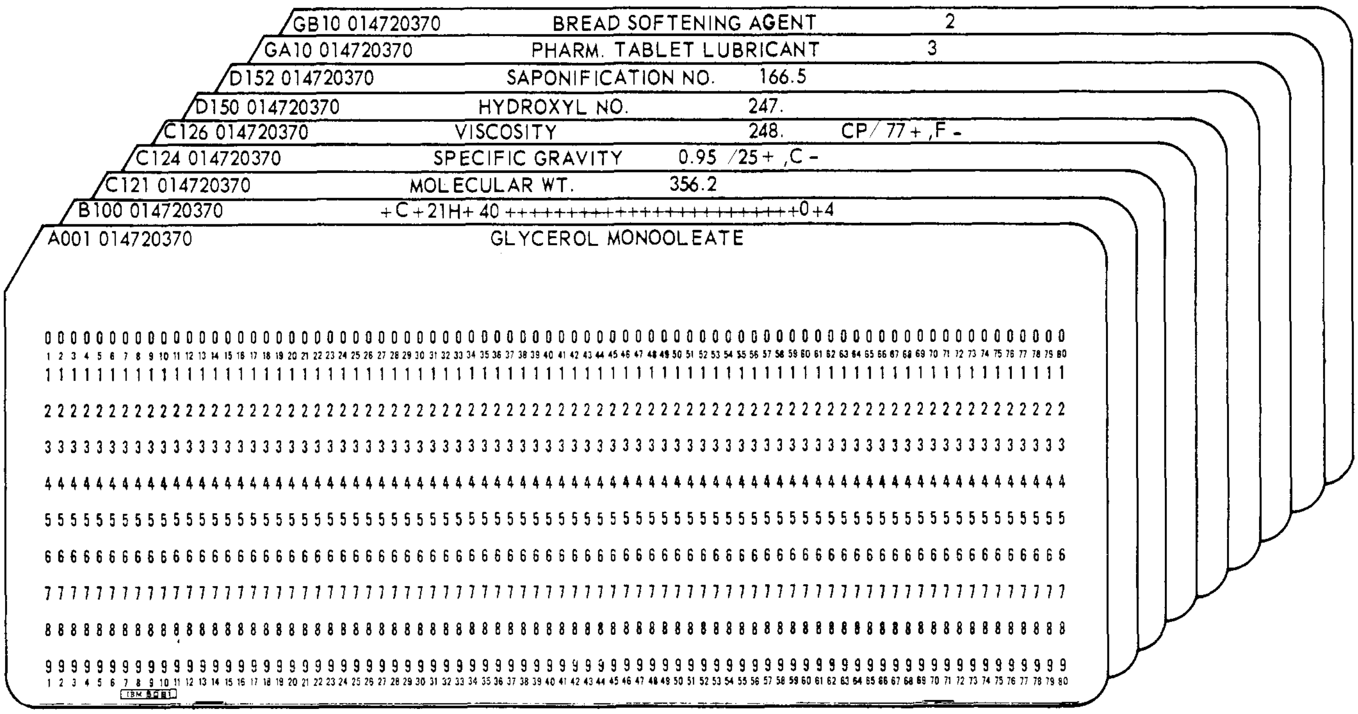


Figure 1. Examples of punched card types

reference compounds are identified by the letter "z," followed by a partial molecular formula, as illustrated by the number for nitrobenzene. The last digit assures a unique number for each compound, distinguishing compounds isomeric in carbon, hydrogen, nitrogen, and oxygen.

z	0 6	0 5	1	0 2	1
indicates organic reference compound	C	H	N	O	distinguishing mark

The compound number identifies the material in the system; additional samples subsequently prepared receive the same number with a suffix (10th digit), indicating first alternate sample, second alternate sample, etc. By this means, analytical data, physical properties, and use test screening results are associated with the particular sample used.

Compound Number	014720370
First Alternate Sample Number	0147203701

numbers permit access to information concerning all preparations of a compound.

showing notebook and page where preparation data pertaining to this second (or later) sample are recorded. These numbers permit access to information concerning all preparations of a compound.

Constitutive properties of each compound are recorded on a molecular formula card and on cards carrying a fragmentation code. The molecular formula card has definite positions for each of the elements, C, H, Br, Cl, F, I, N, Na, O, P, and S, enabling sequencing and selection. Other elements, if present, are recorded in fields beyond the sulfur field. If it is not possible to write a quantitative molecular formula, elements present are indicated by a qualitative formula; thus, a material containing unknown proportions of carbon, hydrogen, nitrogen, oxygen, and phosphorus is recorded as CxHxNxOxPx. The fragmentation code contains approximately 200 fragments. A portion of the form used to code the fragments is shown in Figure 2. Numbers of individual fragments present in a molecule of definite structure are recorded in individual card fields; qualitative presence of a fragment is indicated by x in the same card field. New fragments can be added to the system by adding new cards. At this time, four cards are used.

The constitutive information carried by the molecular formula and the fragmentation code can describe either chemicals of definite structure or those less precisely known, such as mixtures from the partial esterification of polyols. A linear notation or similar structural descrip-

Miscellaneous Structural Information	Branched C Radicals
18,19__3-branch atom (not H)	60_isopropyl
20 _4-branch atom	61_isobutyl
21 _C≡C	62_tert-butyl
22 _nonaromatic C=C	63_tert-octyl
23 _terminal—CH=CH_2	64_"nonyl"
27,28__no. monomers/polymer	65_"dodecyl"
29,30_ _H-bonding centers	66_"tridecyl"
31,32_ _H-donors	
33 _acidic H	

Figure 2. Portion of constitutive properties data sheet
Figures pertain to card columns for recording numbers of individual structural fragments per molecule

tion is not used because these systems are principally valuable only for compounds of known structure.

Analytical results and physical properties are recorded with the name of the determination or property, followed by quantitative values. Only data location is recorded for ultraviolet, infrared, and nuclear magnetic resonance spectra, but not actual spectral data.

Results of use test screening are recorded qualitatively on a scale of 1 to 5. Quantitative data can be recorded when necessary. Over 280 different card types, each representing a test of utility, are presently used. For other use tests performed infrequently, results are expressed in the form of a free write-out of the function tested and performance. A similar card is used for results of infrequent analytical or physical property determinations.

The novelty of a research compound is also a part of its record, just as are its constitutive, physical, or use properties. Results of novelty searches are recorded as either "found" or "not found." Literature references to each compound, if any, are punched. Additional notations to explain the characteristics of the search may be added such as "searched until found," or "references incomplete." A card for each search indicates the main series and supplements of *Beilstein* inspected, and the particular volumes of *Chemical Abstracts*, *Index Chemicus*, or other sources searched. The statement of extent of search qualifies the finding of novelty and gives a starting point for later updating of the novelty search. Novelty searches are recorded both on research compounds already prepared as well as on imaginary compounds. In searches conducted on existing research compounds, the regular compound number is used. Imaginary compounds are identified by a number based on a partial molecular formula, similar to the number used for a non-Atlas compound. Constitutive-properties cards are prepared for these imaginary compounds just as for existing compounds. By this means, novelties of both existing research compounds and speculative materials with similar structural characteristics can be selected and compared, giving preliminary information for patent consideration. If a sample of an imaginary compound becomes available (*e.g.*, from synthesis or purchase), its compound number is changed to one derived from department, notebook, and page.

Information on the preparation, analysis, and properties of each compound submitted for testing is recorded on a data sheet (Figure 3) by the synthetic chemist. After examination by an information chemist, cards are punched directly from the data sheet. Similarly, after the fragmentation code sheet is filled out for each compound by an information chemist, constitutive properties cards are punched. Test results are usually obtained from memoranda or letters and are first recorded on instruction cards, organized for each test, before keypunching. Instruction cards are saved and filed because they may also contain relevant handwritten notes which are not keypunched.

EQUIPMENT

Punched card equipment used in the Atlas Information Section consists of an IBM 082 sorter with counter and two IBM Document Writers. Services of other punched card equipment and the IBM 1401 and 360/30 computers are available in the Atlas Data Processing Department.

The IBM Document Writer serves for input and for low volume, variable output. Essentially this equipment consists of a card punch and attached typewriter, with system function mediated by a control panel (*2, 3*). The output, therefore, has all the flexibility of a conventional typewriter, including upper and lower case. The Document Writers at Atlas have been modified slightly so that all special characters normally in chemical names are in lower case. The upper case of the numbers are subscripts to write molecular formulas in the conventional manner.

Typewriter functions can be wired in the control panel to occur either in specified card columns or at different locations when designated special characters in the card are sensed. In Figure 4, for example, the first letters of the compound, property, or test names are capitalized by wiring from specific card columns while "°C" and "°F" are capitalized by special character punches in the card. The only special characters designated for typewriter functions are for case shifts. Since the typewriter only shifts and does not advance when reading these special characters, their use provides a simple method of closing up unused areas in the card on printout. In the molecular formula card, for example, the quantity of each atom is punched in specified columns for sorting purposes (Figure 1). If a designated element is absent or if not all the alloted columns are used for the coefficient of each element, unused columns are filled with 12 punches. (The 12 punch is the special character used for shift to upper case.) Since the typewriter is already in upper case, these punches only prevent the typewriter from advancing as it would if the columns were blank. Molecular formulas thus are printed in familiar fashion with element symbols and coefficients immediately adjacent.

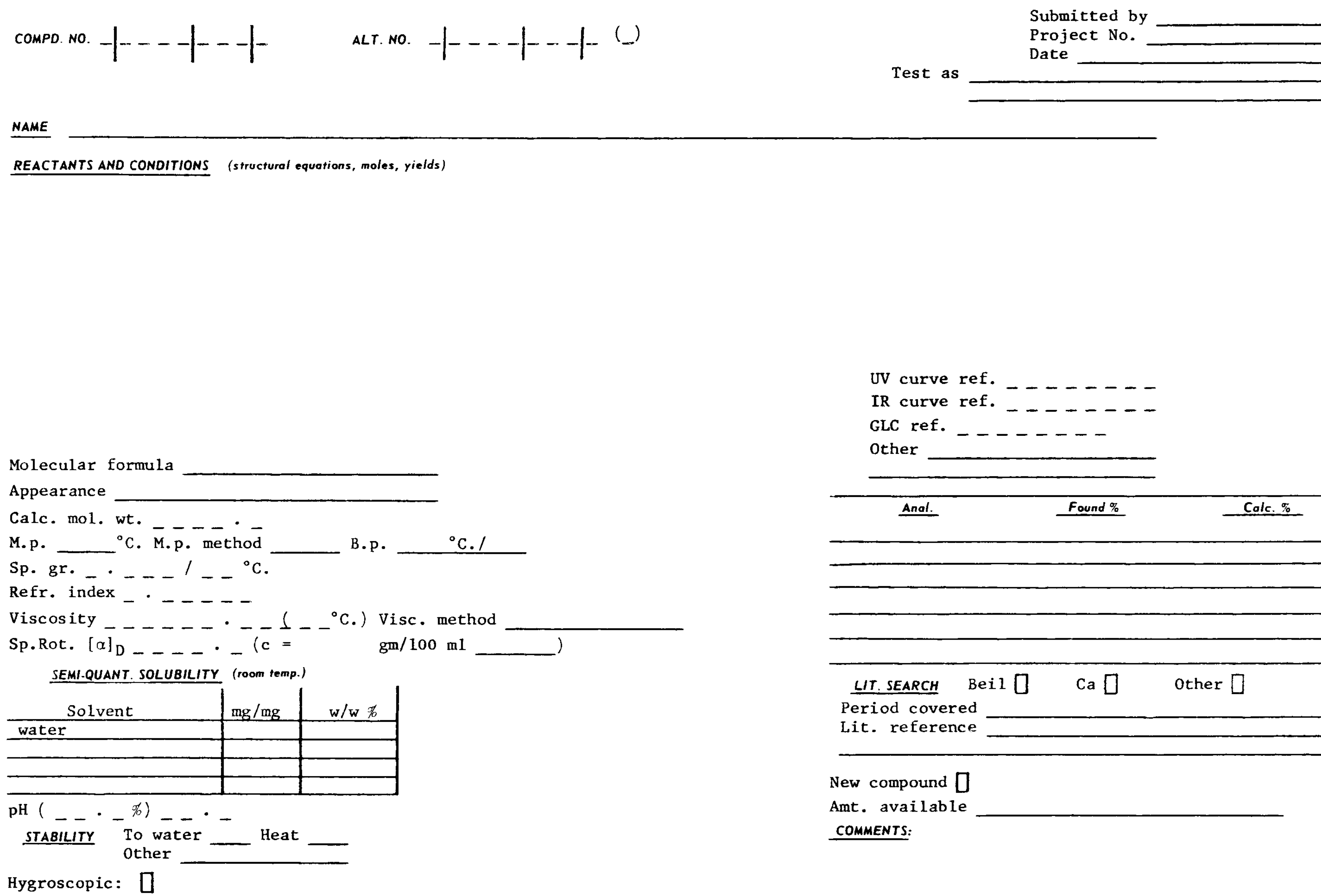
The chemical research information file is duplicated on magnetic tape for use on the Atlas IBM 1401 Computer. A subroutine in the 1401 computer program eliminates the special characters used for typewriter operations. In converting to the more limited print characters of the computer, all lower case characters become upper case and both types of parentheses are converted to the diagonal.

OUTPUT

The output from the system, either by 870 Document Writer or computer, consists of printouts of selected chemical research data, organized variously so that it may be considered from several viewpoints. Printouts are used for proofreading, for review of research results, and for inclusion in research reports. Most listings select data from two or more card types; almost all printouts include compound name and number along with other data.

Data on new compounds and use test results are printed semimonthly on the Document Writer. This printout serves both for verification and after correction of errors, as a chronological record of research activity. Printout (Figure 4) of the punched cards shown in Figure 1 illustrates this listing.

A report is prepared annually which contains names and numbers of all compounds in the system sequenced by quantitative or qualitative molecular formula. The number of any compound can be found quickly with this



COMPD. NO. ___ ___ ___ ALT. NO. ___ ___ ___ ( )

Submitted by ______
Project No. ______
Date ______
Test as ______

NAME ______

REACTANTS AND CONDITIONS (structural equations, moles, yields)

UV curve ref. ______
IR curve ref. ______
GLC ref. ______
Other ______

Molecular formula ______
Appearance ______
Calc. mol. wt. ____ . _
M.p. ____ °C. M.p. method ______ B.p. ____ °C./ ____
Sp. gr. _ . ___ / __ °C.
Refr. index _ . _____
Viscosity _______ . __ (__ °C.) Visc. method ______
Sp.Rot. $[\alpha]_D$ _____ . _ (c = gm/100 ml ______)

SEMI-QUANT. SOLUBILITY (room temp.)

Solvent	mg/mg	w/w %
water		

pH (__ . _ %) __ . _
STABILITY To water ____ Heat ____
Other ______
Hygroscopic: ☐

Anal.	Found %	Calc. %

LIT. SEARCH Beil ☐ Ca ☐ Other ☐
Period covered ______
Lit. reference ______

New compound ☐
Amt. available ______
COMMENTS:

Figure 3. Chemical data sheet

a001	014720370	Glycerol monooleate	
b100	014720370	$C_{21}H_{40}O_4$	
c121	014720370	Molecular wt.	356.2
c124	014720370	Specific gravity	0.95/25° C.
c126	014720370	Viscosity	248. cp/77° F.
d150	014720370	Hydroxyl no.	247.
d152	014720370	Saponification no.	166.5
ga10	014720370	Pharm. tablet lubricant	3
gb10	014720370	Bread softening agent	2

Figure 4. Printout of compound data

$C_6H_4Cl_2$	z06040001	Dowtherm E (treated grade of o-dichlorobenzene)
$C_6H_4N_2O_5$	z06042051	2,4-Dinitrophenol
$C_6H_5Br_1$	z06050001	Bromobenzene
$C_6H_5Cl_1$	z06050002	Chlorobenzene
$C_6H_5N_1O_2$	z06051021	Nitrobenzene
$C_6H_5N_1O_3$	z06051031	*p*-Nitrophenol
C_6H_6	z06060002	Benzene
$C_6H_6Cl_6$	z06060001	Lindane (hexachlorocyclohexane, gamma isomer)
$C_6H_6O_1$	z06060011	Phenol
$C_6H_6O_2$	z06060021	Hydroquinone
$C_6H_7N_1$	z06071001	Aniline
$C_6H_8O_7$	z06080071	Citric acid

Figure 5. Molecular formula listing

012350021 $C_6H_{10}O_4$
Isosorbide
Chem. Novelty- found, references incomplete
Chem. Srch. Beil h12 CA 14-65
Chemical References
JACS 67, 1042 (1945)
JACS 68, 927 (1946)
Carbohydrate Res. 2, 122 (1966)

Figure 6. Example of novelty search printout

report, and the presence or absence of a compound in the file is easily determined. A conventional molecular formula listing of non-Atlas compounds is shown in Figure 5. The molecular formula report also contains lists sequenced by quantity of oxygen, nitrogen, or other atom to facilitate access to compounds containing these elements. Molecular formulas containing elements with but one atom are printed by the Document Writer with the subscript one following the atomic symbol (*e.g.*, $C_{10}H_{15}N_1O_1$), but this redundant symbol has caused no difficulty in interpretation. Compounds with qualitative molecular formulas are grouped according to elements present.

Novelty results are printed alone, or in combination with other data. Printout of a novelty search, without other data, is illustrated by the results of a limited search in the literature for isosorbide, a known compound (Figure 6).

All the chemical research information on magnetic tape is printed at about 3-month intervals. Since the tape file is organized by compound (the cards are filed by card code), this printout is a quick reference to all data recorded for a particular compound.

Many lists combine structural characteristics with other data. For example, to consider the range and nature of available research compounds with active hydrogen atoms, a report was prepared of the names of all such compounds in the file arranged by number and type of active hydrogen, together with experimental and calculated hydroxyl numbers. Other reports include physical data, such as solubilities, melting points, or viscosities, along with analytical data, such as saponification and acid number values.

One of the more important reports is an annual listing of all use test screening results in the file, organized both by test and by compound. These listings were prepared by the Document Writer in the first years of operation, but are now done by computer because of their volume. The program for the report organized by compound scans the fragmentation code and molecular formula of each compound and places its name and use test data in up to 18 chemical classes. Each of these 18 chemical classes

F09630212	Urethane rigid foam, prepolymer					
F09630212	mach a 73° b 100° box				13,140 g	
F09630212	a 41.0 %		t 41.0 %	z09062022	TDI (toluene diisocyanate) 80/20	
F09630212	a 15.2 %	b 42.6 %	t 57.8 %	1394215821	oxypropylene (10) sorbitol	
F09630212		b 0.2 %	t 0.2 %	z08202001	TMBDA (tetramethyl-1,3-butanediamine)	
F09630212		b 0.5 %	t 0.5 %	i 00020011	water	
F09630212		b 0.5 %	t 0.5 %	x11100431	DC-199	
F09630212	good foam				2	
F09630212	density-core					5.16
F09630212	moist vapor transm					.18
F09630212	dimensional stab				100° F 100% rh	2v%/ 1d
F09630212	dimensional stab				100° F 100% rh	2v%/ 7d
F09630212	dimensional stab				158° F 100% rh	1v%/ 1d
F09630212	dimensional stab				158° F 100% rh	1v%/ 7d
F09630212	dimensional stab				158° F 100% rh	−1v%/14d
F09630212	dimensional stab				158° F 100% rh	−1v%/28d
F09630212	tensile strength					158.5 psi
F09630212	compres strength x					104.0 psi
F09630212	heat distortion					279° F

Figure 7. Polyurethane formulation printout

is determined by logical combinations of structural fragments or elements in the molecular formula. These chemical classes and the necessary logic for retrieval are specified on input cards queried by the computer. A final miscellaneous class consists of those compounds in the file not included in any of the designated chemical classes. The report is then printed, grouped by chemical class, with each compound number and name followed by the appropriate test names and data. The program for the report organized by use test inverts the file and prints under each test name, the names, numbers, and test results of the compounds tested. Under each test the compounds are sequenced by qualitative test results, best-to-worst.

Multicomponent systems are described by an extension of the compound system. Each multicomponent system is identified by a formulation number pertaining to the whole composition. This formulation number is carried by each card with data on the multicomponent system. Figure 7 shows an example of a printout of data on polyurethane foam. Composition, size of preparation, preparative conditions, and characteristics of the resulting multicomponent system are listed. Subgroups of components used in different stages of preparation of the foam are identified as "a" and "b" etc. It has been found useful to record the quantity of each component as a per cent of the total system. Total percentages of individual compounds are identified in the printouts by the letter "t." The presence of a component in more than one subgroup can be indicated.

By having information that pertains both to whole formulations (tied together by formulation number) as well as information pertaining to individual components (tied together by compound number) and the relative quantities of the components, performance characteristics of whole systems and their parts can be related. For example, lists can be made of all foams prepared from polyethers having a hydroxyl number over 200. Besides polyurethane foams, emulsions and industrial formulations are recorded by this aspect of the system.

LITERATURE CITED

(1) Haefele, C. R., and J. E. Tinker, J. CHEM. DOC. 4, 112–5 (1964).
(2) Reference Manual IBM 870, Document Writer System, Edition A24-1036-3, June 1963.
(3) Skolnik, H., and R. E. Curtiss, J. CHEM. DOC. 8, 41–5 (1968).

The Literature of Food Science and Technology*

M. M. PISKUR
Swift & Company, Oak Brook, Ill. 60521

Received January 2, 1968

Some of the information sources used to accommodate scientists and technologists involved in the study and the production of food are discussed.

Information on food chemistry and technology is found in languages from Albanian to Zulu as well as in forgotten languages carved on tombs and ancient edifices. Also, tools and residues left by prehistoric man are clues to food technology before writing was invented. Because food is life, its literature has a long history, which has been written by people of diverse nationalities and disciplines. As with all literature, the literature of food science and technology is accumulating at a very rapid rate. The dispersed, voluminous, and expanding nature of the literature poses problems for its efficient use in research, development, and business.

My assignment in this symposium is to examine the literature of food science and technology with respect to requirements and availability in an industrial research environment, such as that of Swift & Co., as observed through the eyes of a literature scientist.

In the meat business, standard textbooks, handbooks, reference books, and current and backfiles of scientific and specialty journals and magazines must be available for the research staff which comprises: analytical through physical chemists; agriculture school graduates, such as: animal husbandmen, meat scientists, dairy technologists, home economists, dietitians; and other professionals, such as various engineers, microbiologists, physiologists, physicists, toxicologists, parasitologists, zoologists, geneticists, statisticians, and veterinary pathologists. There are also chefs, bakers, and candy makers. In addition to research and development activities, the scientific information needs of business and public relations are accommodated.

In current awareness activities, fundamental information is garnered from scientific and technical journals, but application information and considerable technological information are scattered throughout technical and trade magazines, the press, trade literature, and patents. A major food company's technical information staff may scan 200 to 400 periodicals to issue daily, weekly, or monthly citations or abstracts for its clientele. About 50% of the journals are scientific. The others are trade journals and magazines which occasionally contain information of fundamental or technical usefulness to the research and development staff. These two sources may furnish 70 to 80% of the current awareness citations intrinsic to the interests concerned. Further items are garnered from Chemical Abstracts Service which stretches the limits of sources where the food information has chemi-

*Presented before the Divisions of Agricultural and Food Chemistry and Chemical Literature, Symposium on the Literature of Agriculture and Food Sciences, 155th National Meeting of the American Chemical Society, San Francisco, Calif., April 2, 1968.