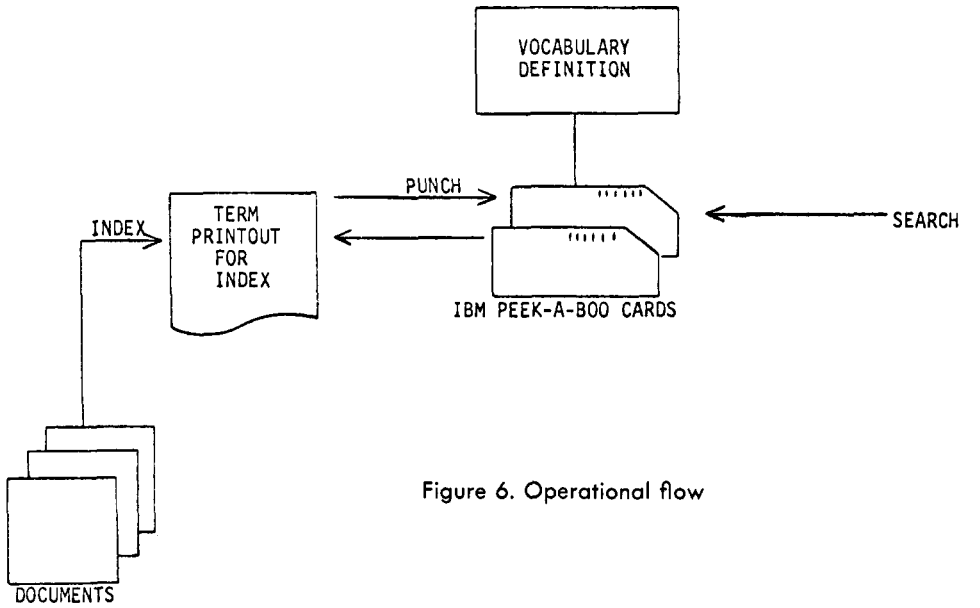


Figure 6. Operational flow

Visual Collation Variable Term–Field cards, and therefore every search would have to be carried through all eight decks.

We have not attempted to establish any definitive guidelines as to when a system of this type can be expected to become too cumbersome for efficient handling. A good rule of thumb seems to be that when the number of daily searches times the number of decks exceeds ten, thought should be given to a new, more effective search tool.

Paper II of this series will describe the techniques that we have developed for converting retrieval systems based on the Variable Term–Field cards to systems that are more appropriate for larger collections and/or more frequent searching.

LITERATURE CITED

(1) Batten, W. E., "Punched Cards," R. S. Casey and J. W. Perry, Eds., pp. 169–181, Reinhold Publishing Co., New York, N. Y., 1951.

(2) Wildhack, W. A., and Stern, J., "Punched Cards," R. S. Casey, J. W. Perry, M. M. Berry, and A Kent, Eds. pp. 147–150, Reinhold Publishing Co., New York, N. Y., 1958.

Computer Indexing of Polymer Patents*

M. M. DUFFEY, I. M. KLANBERG, S. C. MAHR, L. L. MEIER, and J. L. ROMSTAD
E. I. du Pont de Nemours and Co., Experimental Station, Wilmington, Del. 19898

Received March 7, 1968

A computerized system is described for indexing polymers using link-role numbers to describe the polymer backbone (structure), to distinguish between homopolymers and copolymers, to distinguish between unmodified and modified polymers; and to differentiate in addition copolymers between comonomers that must be present and a set of alternate comonomers—e.g., in the case where *x* is copolymerized with *y* or *z*. In all categories the roles also indicate whether the polymer is a product, a reactant, or present. Generic polymer terms are added to cut costs in generic searches.

The Central Research Department in Du Pont has a patent library which houses 600,000 to 700,000 indexed patents—both U.S. and foreign. Included is a comprehensive collection of British, French, German, and U.S. patents which deal with the preparation and use of polymers.

In the early thirties our polymer indexing system depended on buying extra copies of patents and filing these under the name of each monomer of an addition polymer or under the name of each repeating group of a condensation polymer. In 1952 we started making abstract cards for each patent and filing these cards in alphabetical order by headings listed on them. The number of abstract cards per patent varied depending on the number of headings needed to describe the subject material in the patent. Abstract cards relating to addition polymerization had headings listing each monomer in the polymer. The order of monomers was changed on subsequent

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

cards so each monomer appeared as the first one listed, as shown in Figure 1. Thus we could search for specific copolymers or all copolymers of a particular monomer. Abstract cards on condensation polymers were filed under type of repeating unit formed but otherwise handled like the addition polymer cards. We also instituted the use of comma entries after the main headings to indicate

Fr. RESINS, ACRYLIC(ACIDS AND Esters)/VINYL(ACETATE)—MANUF.

SCOTT, JOSEPH D. 1,250,866/60

Prepn. of vinyl acetate-acrylate or methacrylate copolymers by dispn. polymn. in the presence of a free radical initiator
2/26/60 (2/26/59).

TOY, WALTER W.
MADGE, DOROTHY E.
RÖHM AND HAAS CO.

Fr. RESINS, VINYL(ACETATE)/ACRYLIC(ACIDS AND ESTERS)—MANUF.

SCOTT, JOSEPH D. 1,250,866/60

Prepn. of vinyl acetate-acrylate or methacrylate copolymers by dispn. polymn. in the presence of a free radical initiator.
2/26/60 (2/26/59).

TOY, WALTER W.
MADGE, DOROTHY E.
RÖHM AND HAAS CO.

Figure 1. Abstract cards for addition polymers

Brit. RESINS, POLYAMIDE(LATERALLY SUBSTD.)—MANUF.

CRAUBNER, HANS 899,567/62

A graft copolymer is made by nitrosating a polycondensate contg. —CONH— groups at the amide N's. 1/5/61 (1/7/60).

HRUBESCH, ADOLPH
BADISCHE ANILIN- UND SODA-FABRIK A.G.

Brit. RESINS, POLYAMIDE(RING INTERRUPTED)—MANUF.

EDWARDS, WALTER M. 898,651/62

Prepn. of fiber- and film-forming polyamide acids by reacting a diamine such as *p*-phenylenediamine, hexamethylenediamine, or 4,4′-diaminodiphenyl sulfide with at least 1 tetracarboxylic acid dianhydride such as pyromellitic dianhydride or bis-(3,4-dicarboxyphenyl)-sulfone dianhydride in an org. solvent under anhyd. conditions while maintaining the temp. below 60° C.
9/18/59 (9/19/58).

DU PONT

Figure 2. Abstract cards for condensation polymers

chemically after-treated polymers. Figure 2 illustrates the indexing of condensation polymers.

In 1964, we began indexing U. S. patents by a computer-based system. This system is very similar to that described earlier by Barbara Montague (*1*). A year later it was decided that foreign patents should also be indexed by the computer-based system.

Our computer-based indexing system generates four files which can be searched at two levels.

First level general term file containing an alphabetic listing of terms to which patent numbers are posted. Our system is programmed so that when searching under a given term, all related narrower terms are automatically searched.

First level general compound file containing a listing of compound numbers to which patent numbers are posted.

Second level fragment file containing a listing of functional groups and numbers of the compounds containing these groups posted to them.

Second level topological chemical structure file similar to the Compound Registry developed by Chemical Abstracts Service.

In the computer-based system used to index U. S. patents of interest to Du Pont, polymers were indexed

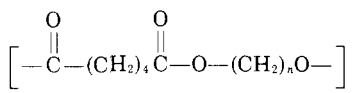
Bis(trifluoro-methyl)ketene	link-role 36	("Must" comonomer in a product)
Ethylene	link-role 38	("Alternate" comonomer in a product)
Styrene	link-role 38	("Alternate" comonomer in a product)
Butadiene/1,3-/	link-role 38	("Alternate" comonomer in a product)
Polycarbons	role 30	(Product, nonmodified)
Copolymers	role 30	(Product)
Vinyl resins	role 30	(Product)

Figure 3. Index terms and link-roles for a copolymer of bis(trifluoromethyl)ketene with ethylene, styrene, or butadiene

Ethylene	link-roles 26, 37	("Must" monomer in a copolymer reactant) ("Must" monomer in a modified copolymer product)
Propylene	link-roles 26, 37	(Same as above)
1,5-Hexadiene	link-roles 28, 39	("Alternate" monomer in a copolymer reactant) ("Alternate" monomer in a modified copolymer product)
Dicyclopenta-diene	link-roles 28, 39	(Same as above)
Polycarbons	roles 20, 31	(Reactant) (Modified product)
Terpolymers	roles 20, 31	(Same as above)
Curing		

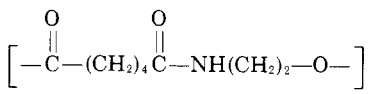
Figure 4. Index terms and link-roles for cured rubber-like terpolymer of ethylene, propylene, and 1,5-hexadiene or dicyclopentadiene

according to prescribed starting materials using roles: to classify the polymers as "Present" (a passive substance), as a "Reactant" and/or as a "Product;" to indicate whether the polymer is a homopolymer or a copolymer; and to indicate the substructure in the backbone for six selected most common polymer classes—i.e., polycarbons, polyethers, polyesters, polycarbonates, polyamides, and polyurethanes; all other polymers, including chemically modified or crosslinked polymers of these common polymer classes, were lumped together in one special class. In



$n = 2$ or 3

and



Adipic acid	roles 94, 95	(Polyamide copolymer product) (Polyester copolymer product)
Ethylene glycol	role 95	(Polyester copolymer product)
Propylene glycol/1,3-/	role 95	(Polyester copolymer product)
Aminoethanol	roles 94, 95	(Polyamide copolymer product) (Polyester copolymer product)
Polyesteramides	role 30	(Nonmodified product)

Figure 5. Indexing of a polymer containing the units:

Adipic acid	role 95	(Polyester copolymer product)
Ethylene glycol	role 95	(Same as above)
Propylene glycol/1,3-/	role 95	(Same as above)
Polyesters	role 30	(Nonmodified product)

Figure 6. Polyesters from adipic acid, ethylene glycol and propylene glycol

Formaldehyde	role 20	(Reactant-actual starting material)
Methylene glycol	roles 56, 66	(Prescribed starting material for a polyether reactant and a polyether product)
Acetic acid	roles 20, 66	(End-capping agent reactant; polyether product)
Polyethers	roles 30, 20, 31	(Preparation of a nonmodified polyether; reaction of a nonmodified polyether; preparation of a modified polyether)

Figure 7. Polymerization of HCHO and end-capping the product with acetic acid

addition, for each polymer an indexing term describing the polymer classification—e.g., "polycarbon" or "polyether"—was used. Polymers containing unusual repeating units not obviously related to starting materials were indexed by fragmentation.

After some study and experimentation, we decided to index polymers by a somewhat modified system which incorporates three new principles:

> Vinyl-type polymers and "condensation" polymers are handled differently.
>
> Combination link-roles are used to obtain more discrimination in indexing and retrieving information on vinyl-type copolymers.
>
> Specific roles are used to index chemically after-treated polymers so that knowledge of the structure of the polymer backbone is retained.

Vinyl-type polymers—i.e., polymers prepared from compounds containing a C=C, C≡C, or combinations of these—are indexed in terms of their monomers. Each monomer is assigned a link-role number. We have provided two sets of link-role numbers, one for nonmodified polymers and one for chemically modified polymers. Specific link-role numbers are used to indicate whether the monomer forms a homopolymer or functions as a "must" or "alternate" starting material for a copolymer. For instance, in a copolymer of A with B, C, or D, A is a "must" starting material and B, C, and D are alternate starting materials. A is given a different link-role than B, C, and D, and thus retrieval of copolymers of B + D, B + C, or C + D is blocked. We use the term "link-role number" because the number links combinations of monomers used to make up vinyl-type copolymers and/or links the polymer to an after-treating agent.

Additionally we use generic indexing terms such as "polycarbons," "homopolymer," "copolymer," "terpolymer" or "tetrapolymer," etc., to facilitate retrieval of polymer classes. Figures 3 and 4 illustrate the indexing of vinyl-type polymers.

The patent illustrated in Figure 3 would not be retrieved falsely in a search for an ethylene-styrene copolymer since the two compounds would be searched in all combinations of 36 and 38 roles and the combination of monomers with 38 roles would be negated.

Our system for indexing nonvinyl-type polymers, which we shall call "Condensation" polymers, differs only slightly from that outlined for the original Du Pont System. The compounds being condensed or polymerized are assigned roles which indicate the type of recurring group that is formed. Provision is made for nine different recurring groups. As indicated previously, the roles indicate the mode (present, reactant, or product) and whether the material forms a homopolymer or a copolymer.

Some of the common condensation polymers are indexed in terms of prescribed starting materials as well as actual starting materials—e.g., polyethylene terephthalate prepared from ethylene glycol and terephthaloyl chloride is indexed under the actual starting materials and under the prescribed starting materials which are ethylene glycol and terephthalic acid. This permits us to index or to search for polymers having a special structure without concern for the actual starting materials.

Polymer terms which describe the backbone structure are always used in indexing condensation polymers.

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	notebook no.	notebook 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