MECHANIZED SEARCHING OF PHOSPHORUS COMPOUNDS

I. RANDOM ACCESS MECHANIZATION*

By J. FROME

U. S. Department of Commerce, Patent Office, Washington, D. C.

INTRODUCTION

The U. S. Patent Office is now conducting mechanized searches in the field of organic phosphorus compounds. Various techniques for mechanized searching have been tried previously in the U. S. Patent Office [1,2,3,4] using the serial or sequential file. Quite recently another experiment was conducted with an "inverted file" technique using parallel access searching.[6] The present system borrows many of the concepts from that experiment but adds several features. One of the advantages of a random access inverted file technique is that the search time increases very little as the number of documents increases. It is felt that the use of an electronic computer having a large random access memory may be one of the answers to the problem of mechanized searching for organic chemical compounds.

Some of the objectives of the present system, called "RAMP" (Random Access Mechanization of Phosphorus), are (1) to test and evaluate the automatic file preparation techniques previously developed; (2) to test, develop and evaluate a random access system for searching organic chemical compounds using disclosures of the phosphate ester art for preliminary data preparation; (3) to aid the patent examiner and the research worker to more effectively and quickly search the chemical literature and (4) to develop principles from the problems encountered in this system which may be applicable to a universal system for searching chemical compounds.

Arts Selected. — The RAMP project involves all the patents in Class 260, subclass 461, as classified in the U. S. Patent Office. It also includes those organic phosphorus patents in other subclasses which are obtainable under our present classification system.[7] There are approximately 1200 patents classified officially in class 260, subclass 461, including original and cross references. The present phosphate project has added to these 1200 patents about 500-700 additional patents containing organic phosphorus compounds, making a total of approximately 1900 patents. It is hoped eventually to include other additional subclasses containing phosphorus and perhaps all the literature which contains phosphorus compounds, but at the present the project is limited to 1900 patents already mentioned.

System. -- Several important features of the present system are (a) use of "inverted file technique," as has already been described [5,6]; (b) the use of groups of atoms as building blocks (NO_2, SO_3H), described in the VS_3 System[3]; (c) use of nodes to show relationships and generate combination terms. H. P. Luhn[9] has suggested that some relationship could be indicated by considering the connection between two elements as a node and this idea has been further extended in this system. A node is a collection of at least two fragments or building blocks. (d) The use of a "compound" digit, which is the fifth digit of a five digit code used in the search. The compound digit enables us to distinguish between several compounds in the same document and helps avoid any searching "noise"; (e) use of combination terms to show certain types of connections, i.e., chlorine-ring (Cl-R) and chlorine chain (Cl-CH) represents that Cl is attached to a ring and chain, respectively; (f) the use of several levels of genericity (see (6); (g) generation of an alphabetical index to each patent and alphabetic index of specifically named compounds 8.

It is believed that the system can best be described or illustrated by dividing it into three phases, namely: (1) the file preparation, (2) the file organization and (3) the search strategy.

1. File Preparation. — Some of the techniques of file preparation used herein have been described in the Patent Office Reports.[8] The initial stage of the RAMP project consists of (a) underlining in each patent all the organic phosphorus compounds contained therein (Fig. 1). This job was done by organic chemists who underlined each and every occurrence of organic phosphorus compounds. The underlining was checked by a patent examiner who is an authority in the phosphorus art in the U. S. Patent Office. From the patents, the name of each underlined compound was punched into a separate card. If there were fifty compounds, fifty cards resulted. After verification, the punched cards then were put through an electronic computer to eliminate the duplicates and to obtain a printed alphabetic listing of all the organic phosphorus compounds named in each patent (Fig. 2). The compounds illustrated by structural formulas in the patent were not included on this list. These printed listings then were attached to the patents and given to the organic chemists.

(b) The organic chemists were instructed to write the structural formula for every organic phosphorus compound on the list. After all the structural formulas were written, the patents

*Presented before Division of Chemical Literature, ACS National Meeting, New York, N. Y., September 13, 1960.

FIG. 1. A COPY OF AN UNDERLINED PHOSPHORUS PATENT

United States Patent Office 2,875,229
Patented Feb. 24, 1959

2,875,229

PREPARATION OF NEUTRAL MIXED PHOSPHATES FROM TRIALKYL PHOSPHATES AND CARBOXYLIC ACID ESTERS

Harry W. Coover, Jr., and Richard L. McConnell, Kingsport, Tenn., assignors to Eastman Kodak Company, Rochester, N. Y., a corporation of New Jersey

No Drawing. Application February 14, 1956
Serial No. 565,300

6 Claims. (Cl. 260—461)

This invention relates to a process for the preparation of neutral mixed phosphates. In a specific aspect this invention relates to a process for preparing neutral mixed phosphates having the structural formula:

$$[RX][R'X][R'X]P{=}X$$

wherein R and R' are radicals selected from the group consisting of alkyl, substituted alkyl, aryl and substituted aryl and wherein X is either oxygen or sulfur.

Neutral mixed phosphates have been prepared by various prior art procedures. For example, such phosphates have been prepared by the reaction of a primary alcohol with phosphorus oxychloride and the resulting phosphoryl dichloride is reacted with sodium phenoxide. Also such phosphates have been prepared by reacting a trialkyl phosphate with an alkyl chlorosulfinate or chloroformate. Another procedure involves the reaction of a chlorophosphate with an alcohol, a phenol or an alkali metal alkoxide. Each of the procedures suffers from some disadvantage. For example, the latter procedure requires the preparation of the chlorophosphate intermediate, which must be isolated and when the chlorophosphate is reacted with an alcohol, a tertiary organic base is required to remove the liberated hydrogen chloride. The only alternative is to react an alkali metal alkoxide with the chlorophosphate but the yields from this procedure are poor due to competing reactions.

In accordance with this invention, it has been found that neutral mixed phosphates can be produced economically in excellent yields by reacting a phosphate selected from the group consisting of trialkyl phosphates and trialkyl thiophosphates, wherein the alkyl radicals contain up to 8 carbon atoms per alkyl radical, with a lower carboxylic acid ester. The products of this invention have the structural formula:

$$[RX][R'X][R'X]P{=}X$$

wherein R and R' are radicals selected from the group consisting of alkyl containing up to 8 carbon atoms, such as methyl, ethyl, butyl, octyl and the like, haloalkyl, containing up to 8 carbon atoms, such as chloropropyl, bromobutyl, and the like, and aryl and substituted aryl, such as phenyl, cresyl, chlorophenyl, nitrophenyl, and the like. In these products R and R' are different and at least one of R and R' is an alkyl radical. X is either oxygen or sulfur. In the reaction, the trialkyl phosphate is reacted with an ester having the structural formula:

$$R'''\overset{\overset{O}{\|}}{C}{-}O{-}R''$$

In this carboxylic acid ester, R'' is the same as R and R' but different from the alkyl radicals in the trialkyl phosphate or thiophosphate reactant. R''' is a lower alkyl radical such as methyl, ethyl, propyl, butyl and the like. When triethyl phosphate is reacted with butyl acetate in accordance with this invention, the reaction can be expressed by the following equation:

$$(C_2H_5O)_3P{=}O + CH_3\overset{\overset{O}{\|}}{C}OC_4H_9 \longrightarrow (C_2H_5O)_2\overset{\overset{O}{\|}}{P}(OC_4H_9) + CH_3\overset{\overset{O}{\|}}{C}OC_2H_5$$

The reaction is carried out for a period of 1 to 24 hours and at a temperature of 100 to 275° C. The preferred temperature range is from 125 to 240° C. for a period of time ranging from 4 to 16 hours depending upon the reactants employed. When no catalyst is used in the reaction, temperatures, in excess of 200° C., for example about 250° C., are employed. The reaction can, however, be carried out in the presence of a catalyst and when a catalyst is used, substantially lower temperatures are suitable. For example, when a lead oxide catalyst such as litharge is used, a temperature of about 130° C. is satisfactory. When a catalyst is used in the reaction, an amount within the range of 0.5 to 5% by weight and higher is usually employed.

Varying the reactants has an effect upon the mixed phosphate esters produced in the reaction. For example, in the reaction of butyl acetate with triethyl phosphate, an excess of the butyl acetate results in the production of dibutyl ethyl phospate. It is actually preferable to use an excess of the carboxylic acid ester since the excess accelerates the reaction and produces better yields of the mixed phosphates. However, equimolar proportions of the reactants can be used and, if desired, an excess of the trialkyl phosphate can be employed.

The products of this invention are useful as plasticizers, solvents, pesticides, synthetic lubricants and intermediates for the preparation of other organophosphorus compounds. Cellulose esters plasticized with these mixed phosphates are self-extinguishing.

The following examples wherein the parts are in parts-by-weight illustrate the invention:

Example 1.—Mixed butyl ethyl phosphates

A mixture of 36.4 parts of triethyl phosphate, 139.0 parts of n-butyl acetate and 5.0 parts of yellow plumbous oxide (litharge) was heated under total reflux for 2 hours with a pot temperature of 130° C. Then low boiling materials were removed from the top of the distillation column within the 75–120° C. range for 12–14 hours. This distillate consisted of a mixture of ethyl and butyl acetate. The remainder of the excess butyl acetate was removed by distillation at atmospheric pressure. The reaction mixture was then filtered to remove the catalyst residue and vacuum distilled. After removing 13.4 parts of unreacted triethyl phosphate, 16.1 parts of butyl diethyl phosphate (B. P. 84–89° C. at 2.5 mm.) and 7.0 parts of dibutyl ethyl phosphate (B. P. 95–98° C. at 2.5 mm.) were collected. It is unnecessary to fractionate the two phosphates and the entire product boiling within the range 84–98° C. at 2.5 mm. can be used since this mixture makes an excellent plasticizer for cellulose esters. This mixture of butyl ethyl phosphates can be used alone or in combination with other conventional plasticizers to give any desired flow. Cellulose esters containing 15–20 parts of this mixture of phosphates are self-extinguishing.

Example 2.—Mixed methyl octyl phosphates

A mixture of 28.0 parts of trimethyl phosphate, 206.0 parts of octyl acetate and 5.0 parts of litharge was reacted according to the procedure in Example 1 to give a mixture of dimethyl octyl and dioctyl methyl phosphate.

Example 3.—Mixed ethyl phenyl phosphates

A mixture of 36.4 parts of triethyl phosphate, 164.0 parts of phenyl acetate and 3.0 parts of litharge was reacted according to the procedure in Example 1 to pro-

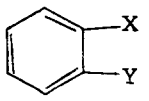
FIG. 2. PRINTED LIST OF COMPOUNDS IN PATENT

2875229 BUTYL DIETHYL PHOSPHATE
2875229 BUTYL ETHYL PHOSPHATES
2875229 CHLOROPHOSPHATE
2875229 DIBUTYL ETHYL PHOSPHATE
2875229 DIETHYL ISOBUTYL THIOPHOSPHATES
2875229 DIETHYL O NITROPHENYL THIOPHOSPHATE
2875229 DIETHYL PHENYL PHOSPHATE
2875229 DIISOBUTYL ETHYL THIOPHOSPHATES
2875229 DIMETHYL OCTYL PHOSPHATE
2875229 DIOCTYL METHYL PHOSPHATE
2875229 DIPHENYL ETHYL PHOSPHATE
2875229 MIXED BUTYL ETHYL PHOSPHATES
2875229 MIXED PHOSPHATES
2875229 MIXED PHOSPHATE ESTERS
2875229 MIXED ETHYL ISOBUTYL THIOPHOSPHATES
2875229 MIXED ETHYL O NITROPHENYL THIOPHOSPHATES
2875229 MIXED ETHYL PHENYL PHOSPHATES
2875229 MIXED METHYL OCTYL PHOSPHATES
2875229 NEUTRAL MIXED PHOSPHATES
2875229 NEUTRAL MIXED THIOPHOSPHATES
2875229 ORGANOPHOSPHORUS COMPOUNDS
2875229 PHOSPHATE
2875229 PHOSPHATES
2875229 PHORPHORUS OXYCHLORIDE
2875229 PHOSPHORYL DICHLORIDE
2875229 TRIALKYL PHOSPHATE
2875229 TRIALKYL PHOSPHATES

2875229 TRIALKYL THIOPHOSPHATES
2875229 TRIETHYL PHOSPHATE
2875229 TRIETHYL PHOSPHOROTRITHIOATE
2875229 TRIETHYL THIOPHOSPHATE
2875229 TRIMETHYL PHOSPHATE
2875229 THIOPHOSPHATES

were then checked to see whether there were any other organic structural formulas in the patent which had been written already, and such formulas were included with the formulas on the list. The patent then was read carefully to make sure that all the reaction products containing organic phosphorus also were represented by structural formulas. Thus there was obtained a series of structural formulas for every organic phosphorus compound either specifically named or shown by structural formula or resulting from any reaction product in the patent. In this list not only were structural formulas drawn for the specific compounds listed but also those generic formulas for which no species had been written. Upon completion by the organic chemist of the structural formulas, these formulas then were thoroughly checked and verified by an organic chemist skilled in the phosphorus art.

(c) Markush. — This term is used in the U. S. Patent Office to designate an artificial genus. One of the most common ways of representing a Markush group is by use of a structural formula such as:



X may be a member of the group consisting of OH, NH_2, SO_3H and Y may be a member of the group consisting of methyl, ethyl and phenyl.

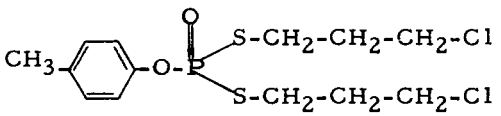
Structural formulas which were in the form of the Markush group were also included in the above list and were later treated in a separate way (see infra).

(d) Encoding is a very important aspect of this system and in order to understand more thoroughly how it is done, we give a description of the method of fragmenting and nodalizing and the terms used therein.

A fragment is a chemical element or a collection of chemical elements treated as a unit for chemical or information retrieval purposes. It is a component part of a structural formula and a series of designated fragments will constitute a structural formula.

A node is a collection of at least two fragments. For the purpose of this system every node will contain a phosphorus atom as one of the fragments.

An example of a compound to be fragmented and nodalized is the formula



GENERAL RULES

Explanation of Fragmentation and Nodalization. — The fragmentation and nodalization of every phosphorus structural formula begins with and revolves around the phosphorus atom in the formula.

(1) The first step is to designate the element or elements which are directly connected to the phosphorus atom and also the number of occurrences of these connections. The phosphorus atom and all the elements or fragments directly connected to it will hereafter be designated as the phosphorus nucleus.

Example: P=O-1
P-S-2
P-O-1

(2) The second step is to designate the first node. The first node comprises the phosphorus nucleus and the fragment or fragments which are directly connected to the phosphorus nucleus and their number of occurrences and also the appropriate symbols in the case of a carbon chain, metals, halogens, etc.

Example: P-S-C3-NT-SAT-ST-CH-2 (Fig. 3)
P-O-PHENYL-1

(3) The third step is to designate the fragment or fragments which are directly connected to the phosphorus nucleus and also the number of times they occur.

Example: C3-NT-SAT-ST-CH-2 (Fig. 3)
PHENYL-1

(4) The fourth step is to designate all of the fragments which are connected either directly or indirectly to the first node as described in (2) above, using the appropriate symbols and recording the number of occurrences of each of the fragments.

Example: C1-CH-2
C-T-R-1

(5) The fifth step is to designate the position of all the fragments on the ring, that is, whether they are ortho, meta or para, and also the total number of fragments in the structural formula as well as the valence of the phosphorus atom in the compound.

Example: P (Para)
12 Frag
Phos 5

FIG. 3. DEFINITION OF TERMS USED IN FRAGMENTING

T = Terminal
NT = Non-Terminal
SAT = Saturated
UNS = Unsaturated
ST = Straight
BRAN = Branched
CH = Chain
R = Ring
O = Ortho – two substituents on ring side by side (vicinal)
M = Meta – two substituents on ring one space removed
P = Para – two substituents on ring two spaces removed
NS = No species used, only with generic terms
CYCL = Cyclic
PHOS = Phosphorus
FRAG = Fragments
AMINE SALT-P = Primary amine salt
AMINE SALT-S = Secondary amine salt
AMINE SALT-T = Tertiary amine salt
AMINE SALT-Q = Quaternary amine salt
POLYMER P = Polymer of phosphorus

The above structural formula therefore will result in these fragments and nodes:

1. P=O-1
   P-O-1
   P-S-2

2. P-S-C3-NT-SAT-ST-CH-2
   P-O-Phenyl-1

3. C3-NT-SAT-ST-CH-2
   Phenyl-1

4. C-T-R-1
   C1-CH-2

5. P (Para)
   12 Frag
   Phos 5

Detailed instructions on fragmenting and nodalization will be available from U. S. Patent Office, R. & D.

(e) Loading into the File. — After the compounds have been fragmented and checked, each fragment is punched into a separate card. Thus if a compound contains ten fragments it will have ten fragment cards.

The punched card (Fig. 4), besides containing the identity of the fragment, also contains the patent number, the accession number and the compound number.

In the system the patent numbers are designated by accession numbers which are four digit numbers. When a patent contains more than 36

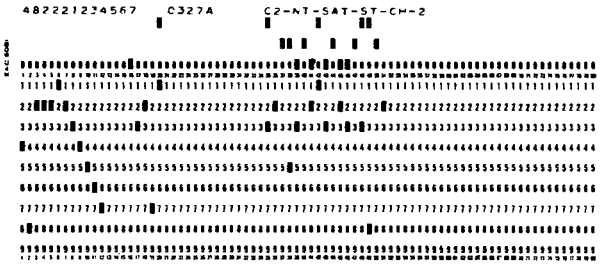


FIG. 4. SHOWING THE FORMAT OF THE PUNCHED CARD USED IN LOADING AND DESCRIBING THE FRAGMENTS AND NODES

Cols. 1-5	III Level Address
Cols. 6-12	Patent Number
Cols. 16-19	Accession Number
Col. 20	Compound Number
Cols. 35-80	Fragment or Node

compounds, the patent is represented by more than one accession number. Thus, for example, a patent which contains 46 compounds is represented by two accession numbers and a patent which contains 80 compounds is represented by three accession numbers. The compound number is the fifth digit, a number or letter, added to the accession number. This gives us the possibility of 36 compounds per accession number. As will be seen later on there is no possibility of confusion between the various compound numbers. The cards, each containing a fragment, derived from the structural formulas, then are put through a tabulating machine and the fragments are listed compound by compound. In the case of the Markush formulation, each Markush group is represented as a single compound even though it may in fact be a representation of many, many compounds. Separate compounds represented by separate Markush groups are given separate compound numbers. After corrections have been made, all the punched cards represented by all the patents then are sorted alphabetically and arranged in alphabetical sequence. These cards next are processed by the RAMAC, which automatically assigns an address for each fragment and automatically loads it into the RAMAC memory. A list of the fragments and their address numbers and the number of times the fragment occurs also is obtained. This list of fragments and their addresses and frequencies constitutes our searching dictionary (Fig. 5).

File Organization. — The file is organized in the same general manner as most random access or inverted files. Under each address or descriptor is listed the combined accession number and compound number of each compound

which contains the descriptor. This is contrary to the normal situation as presented in the serial file. The organization of this file is substantially the same as the organization of the file in our previous work.[6] However, there are two main differences. The specifically named compounds are not individually recorded within the computer but only the fragments composing those compounds are so recorded. The computer contains three levels, all of which are more or less generic to the fragment. In effect the file is organized into four main groups: levels 1, 2 and 3 in the computer and level 4 which is in fact a printed list of all the compounds in the patents exclusive of Markush and reaction products. Level 1 is considered the most generic, level 2 is less generic and level 3 is the most specific. A specific example of the organization of the file can be seen in Fig. 6.

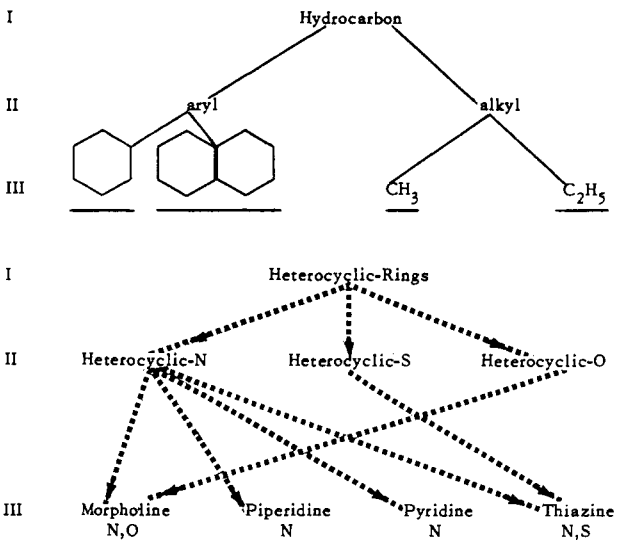
FIG. 5. THESE LISTED FRAGMENTS ARE EXCERPTS FROM THE III LEVEL, PHOSPHORUS DICTIONARY

Address	Name	No. of occurrences in Compound	No. of occurrences in File
20017	ACRIDINE-1		1
20018	ACRIDINE-2		1
20019	ACRIDINE-3		1
20020	AL-CH-1		1
20021	AG-CH-1		1
20022	AL-CH-1		4
20023	AG-CH-1		1
20024	AG-CH-2		1
20706	C-O-C-NT-R-7		1
20707	C-O-C-NT-R-8		1
20708	C-O-C-NT-R-9		1
20709	C-O-C-T-CH-1		18
20710	C-O-C-T-CH-10		1
20711	C-O-C-T-CH-11		1
20712	C-O-C-T-CH-12		1
20713	C-O-C-T-CH-2		4
20714	C-O-C-T-CH-3		2
20715	C-O-C-T-CH-4		1
25882	C2-T-SAT-ST-R-2		
25883	C2-T-SAT-ST-R-3		18
25884	C2-T-SAT-ST-R-4		10
25885	C2-T-SAT-ST-R-5		4
25886	C2-T-SAT-ST-R-6		5
25887	C2-T-SAT-ST-R-7		4
25888	C2-T-SAT-ST-R-8		2
25889	C2-T-SAT-ST-R-9		2
25890	C2-T-SAT-5T-CH-1		1
25891	C2-T-ST-CH-2		2
25892	C2-T-UNS-ST-CH		1
25893	C2-T-UNS-ST-CH-1		47

SEARCH STRATEGY

To identify a chemical compound for patent searching purposes it is believed that it is necessary for a system to be able to do several things, especially in the phosphorus art: (1) to be able to identify each of the fragments comprising a compound, (2) to be able to identify the number of times each different fragment occurs in the compound; (3) to be able to find the relationship

FIG. 6. LEVEL OF GENERICITY



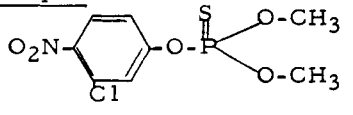
between these various fragments in the compound; (4) to be able to ask the search question either very specifically or generically, that is, with any degree of genericity desired. If a system can do these, it usually will serve the purpose of a patent search. The fragments and the number of occurrences in a compound are found easily by looking in the dictionary, which will have the name of the fragment as well as the number of times it occurs in a compound.

Relationships. — Relationships are obtained by three devices, namely, (a) combination terms, (b) grouping terms, i.e., ring or chain, and (c) the fifth digit number in the address which shows all the fragments are in the same compound.

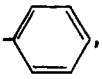
Genus and Species. — Since the computer search file is organized in three levels, it is possible to ask for the compound by asking for a combination of fragments either specifically or generically. Some of the fragments may be asked for by entering the file in the generic first level and others in the more specific third level. Thus by the use of the computer we are able to achieve as much genericity or specificity as required.

3. Asking the Question. — In asking the search question, the examiner studies the compound which is being claimed. He then decides what the essence of the invention is and by using the search dictionary selects the combination of fragments and the relationships which will give him those documents that will meet the claim. A specific illustration of an actual search will be illustrative.

Example: It is desired to find the compound



The compound would contain these fragments

P=S-1	(P=S, one occurrence)
P-O-3	(P-O, three occurrences)
P-O-C-T-CH-2	(P-O-C [terminal C chain], two occurrences)
C-T-CH-2	(C [terminal C chain], two occurrences)
P-O-Phenyl-1	P-O-  , one occurrence
Phenyl-1	 —, one occurrence
Cl-R-1	(Cl Ring attached, one occurrence)
NO_2-R-1	(NO_2 Ring attached, one occurrence
O	Ortho
M	Meta
P	Para
12 Frag	Total fragment count
Phos 5	Valence of P

One could of course ask for all the fragments and their relationship and obtain the compound. However, a more generic question could be asked, i.e., "for all chlorine and NO_2-containing compounds attached to a ring and containing the configuration P-O-Phenyl." In the computer search, there are obtained 34 patents which answer the question from a file of 693 patents.

However, to limit the question and reduce the number of documents retrieved you may insert the descriptor "12 fragments" which means only those compounds containing 12 fragments are acceptable.

When this was done only eight patents were retrieved. The use of the descriptor "No. of fragments" in a compound is a powerful tool. It enables a searcher to obtain the compound exactly and eliminates those compounds which contain additional groups. Furthermore, it distinguishes the specifically named compounds from those which are only included by the virtue of being in a Markush group.

Upon using all the possible descriptors for the compound, only seven patents were retrieved.

It should be noted that for most compound questions it is sufficient to ask for three to five of the most unusual descriptors. The results obtained will be almost as good as if all of the descriptors had been asked. In practice the average search took about 3-5 minutes of machine time.

Of course one could ask the computer a more generic question "For a halogen containing compound having the halogen attached to an aromatic ring and containing the (P-O) group three times," and the above compound would be retrieved.

The table gives a graphic representation of what might be in the computer's memory:

III Level

20495	20782	28796	29847
Cl-R-1	NO_2-R-1	P-O- 	12 Fragments
12340	12340	12340	12340
12356	12356	12354	12351
12452	12452	12452	12452
13332	13332	13332	13456
14587	14568	14569	14584

When the question is asked "Find a compound containing Cl and NO_2 attached to a ring and containing a P-O-Phenyl group," the computer goes to the address 20495 (Cl-R-1) and to address 20782 (NO_2-R-1) and compares the five digit-numbers and stores the numbers that are the same. Thus

12340 12356 12452 13332

It then goes to 28796 (P-O-Phenyl) and compares the stored numbers with those under P-O-Phenyl. This results in

12340 12452 13332

The last digit, which is the compound number, then is dropped and the four digit number then is transformed into the patent number by a dictionary look-up.

If the question included the limitation that the compound should contain only twelve fragments, then the computer goes through the same steps but when it takes the stored

12340 12452 13332

and compares these numbers with those under address 29847 (12 frag), the resulting comparison will only give two answers

12340 12452

These are then converted to patent numbers by the computer.

In Figure 7 there appears an actual search question sheet.

This system is in actual use in the U. S. Patent Office, and some actual statistics from its use are

PHOSPHORUS MACHINE SEARCHES (RAMAC)

1. Applications Searched	52
2. Total Searches Made	200
3. Searches per Application	3.8
4. Time to Prepare Search Question	2-3 min.
5. Machine Time per Search	3-4 min.
6. Patents Retrieved per Search	10

PHOSPHORUS MACHINE SEARCH FILE

Patents Completed (6-15-60)	693
Chemical Structures (compounds)	17,500
Chemical Compounds per Patent	25
Chemical Fragments	256,350
Chemical Fragments per Structure	15
Chemical Fragments per Patent	370
Fragment Terms in Dictionary	23,775
Fragments per Dictionary Term (frequency)	11
Punch Cards Used	350,000

We have made approximately 200 satisfactory searches.

CONCLUSIONS

It is felt that the method of file preparation, file organization and search strategy offers a solution to mechanized searching in the phosphorus art. Several general principles may be gained from our experience in using this system.

1. It is important to use either semi-automatic or automatic techniques in file preparation.

2. It is extremely important to verify the accuracy of the file and analysis at every stage of operation.

3. The random access method of searching for chemical compounds offers a promising approach.

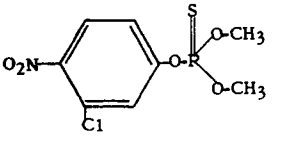
4. It appears that this system can be extended to other areas of organic chemical compounds in which there is a central atom or group of atoms serving as a nucleus such as "silicon" compounds, "boron compounds", etc. Of course it is obvious that it can be used for certain specific arts if one wanted to use such a grouping as "pyridine" ring as a central nucleus for the pyridine art.

Acknowledgement. — Acknowledgement is expressed for the contributions of Martin A. Gannon of R. & D. and Frank M. Sikora of the Examining Corps of the U. S. Patent Office, for their aid in the analysis and encoding operations.

FIG. 7. – SEARCH QUESTION SHEET

EXAMINER	APPLICATION NUMBER	STATUS	DATE

ESSENCE OF INVENTION



R&D PHOSPHORUS (J.F.) (3/60)

	QUESTIONS	ADDRESS	FREQ.	TOTAL # OF DOCU.	(CITED) ANSWERS
1					
	Cl-R-1	24878	461		
	NO_2-R-1	28229	574		
	P-O-Phenyl-1	41130	1009		
2					
	Cl-R-1	24878	461		
	NO_2-R-1	28229	574		
	P-O-Phenyl-1	41130	1009		
	12 Frag.	44345	772		

REFERENCES

[1]Bailey, M. F., B. E. Lanham and J. Leibowitz, "Mechanized Searching in the U. S. Patent Office," U. S. Department of Commerce, Patent Office, Washington 25, D. C., 1950.

[2]Frome, Julius, and Jacob Leibowitz, "A Punched Card System for Searching Steroid Compounds," Patent Office Research and Development Report No. 7, Department of Commerce, Washington 25, D. C., 1957.

[3]Leibowitz, J., J. Frome and D. D. Andrews, "Variable Scope Search System: VS_3," preprints of papers for the International Conference on Scientific Information, National Academy of Sciences-National Research Council, Washington, D. C., 1958, pp. 291-316.

[4]Frome, Julius, Jacob Leibowitz and Don D. Andrews, "A System of Retrieval—Compounds, Compositions, Processes and Polymers," Patent Office Research and Development Report No. 13, Department of Commerce, Washington 25, D. C., 1958.

[5]Nolan, John L., American Documentation, January, 27-35 (1959).

[6]Leibowitz, J., J. Frome and D. D. Andrews, "Variable Scope Patent Searching by an Inverted File Technique," Patent Office Research and Development Report No. 14, Department of Commerce, Washington 25, D. C., November, 1958.

[7]Classification Bulletin of the United States Patent Office, Class 260—Chemistry, Carbon Compounds, Bulletin No. 200, Revision 1.

[8]Frome, Julius, "Semi-Automatic Indexing and Encoding," Patent Office Research and Development Report No. 17, Department of Commerce, Washington 25, D. C., 1960.

[9]Luhn, H. P., "Nodal Index for Branched Structure," IBM Research Laboratory, Poughkeepsie, N. Y., December, 1955.

[10]Mooers, Calvin N., Zator Technical Bulletin No. 64, 1951.