Patent Office
Research and Development
Reports . . . . . No. 18

MECHANIZED SEARCHING OF PHOSPHORUS COMPOUNDS

Prepared by

Julius Frome

Patent Research Specialist

Office of Research and Development

Patent Office

January 3, 1961

U. S. DEPARTMENT OF COMMERCE
Frederick H. Mueller
Secretary

PATENT OFFICE
Robert C. Watson
Commissioner

For sale by the U. S. Department of Commerce, Washington 25, D. C. Price 25 cents.

Acknowledgment

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

CONTENTS

PART I

RANDOM ACCESS MECHANIZATION OF PHOSPHORUS COMPOUNDS (RAMP)

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 previously used in the U. S. Patent Office (1), (2), (3), (4). These techniques used the serial or sequential file. Quite recently another experiment was conducted using an "inverted file" technique with parallel access searching (6). The present system borrows many of the concepts applied in 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. This includes the 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 contain organic phosphorus compounds, but at present the project is limited to 1900 patents mentioned supra.

SYSTEM

Several important features of the present system are:

(a) Use of "inverted file technique" which has been described in (5,6).

(b) The use of groups of atoms as building blocks (NO_2, SO_3H). This technique has been described in the VS_3 System (3).

(c) Use of nodes to show relationships and generate combination terms. H. P. Luhn in (9) suggested some relationship which could be indicated by considering the connection between two elements as a node. This idea has been further extended in this system (infra). A node is a collection of at least two fragments or building blocks.

(d) The use of a "compound" digit. A compound digit is the fifth digit of a five digit code used in the search. The compound digit enables the user to distinguish between several compounds in the same document and helps avoid any "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 (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 3 phases, namely:

(1) the file preparation
(2) the file organization and
(3) the search strategy.

FILE PREPARATION

Some of the techniques of file preparation used herein had been described in the Patent Office Reports (8).

Items Identified

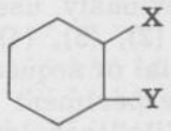
The initial stage of the RAMP project consists of underlining in each patent all the organic phosphorus compounds contained therein. (Appendix 1.) This job was done by organic chemists who underlined each and every occurrence of organic phosphorus compounds. The underlining was checked by a patent examiner 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 were then 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. (Appendix 5.) The compounds illustrated by structural formulas in the patent were not included on this list. These printed listings were then attached to the patents and given to the organic chemists.

Structural Formulas Written

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 (for all the organic phosphorus compounds on this list) the patents were then checked to see if there were any other organic structural formulas in the patent which had not already been written. Those formulas which had not already been written were included with the formulas on the list. The patent was then carefully read to make sure that all the reaction products containing organic phosphorus were also 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 above 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 were then thoroughly checked and verified by an organic chemist skilled in the phosphorus art.

Markush Formulas Included

This term refers to a term 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 structural formula, as follows:



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).

Formula Components Encoded

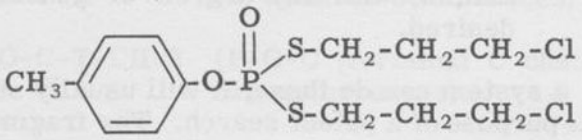
Encoding is a very important aspect of this system. In order to more thoroughly understand how this is done, the following is a description of the method of fragmenting and nodalizing and the terms used therein.

FRAGMENT — 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.

NODE — 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.

The following is an example of a compound to be fragmented and nodalized.

Formula:



General Rules of Fragmentation and Nodalization Explained — 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 the connection in the compound, using the phosphorus atom as the nucleus. This represents the first node.

Example: P=0-1
P-S-2
P-0-1

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

Example: (a) P-S-C_3-NT-SAT-ST-CH-2 (Appendix 3)
(b) P-O-Phenyl-1

This represents the second node.

(3) The third step is to designate the fragments of all the first node connections and their frequencies.

Example: (a) C_3-NT-SAT-ST-CH-2 (Appendix 3)
(b) Phenyl-1

(4) The fourth step is to designate the fragments which are directly connected to the second node, described above in (2) (supra).

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

(5) The fifth step is to designate the position of the groups on the ring, i.e., ortho, meta and para, the total number of fragments in the structural formula, and also the valence of phosphorus in the compound.

Example: P (Para)
13 Frag
Phos 5

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

1. P=0-1
P-0-1
P-S-2
2. P-S-C_3-NT-SAT-ST-CH-2
P-O-Phenyl-1
3. C_3-NT-SAT-ST-CH-2
Phenyl-1
4. C-T-R-1
Cl-CH-2
5. P (Para)
13 Frag
Phos 5

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

File Loaded

After the compounds have been fragmented and checked, each fragment is punched into a separate card. Thus, if a compound contains 10 fragments, it will have 10 fragment cards.

The punched card, (Appendix 2), 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 compounds, the patent is represented by more than one accession number. For example, a patent which contains 46 compounds is represented by 2 accession numbers, and a patent which contains 80 compounds is represented by 3 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, are then 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 are then sorted alphabetically and arranged in alphabetical sequence. These cards are then 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 is also obtained. This list of fragments and their addresses and frequencies constitutes our searching dictionary. (Appendix 6.)

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 said 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 Appendix 4.

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 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 will usually serve the purpose of a patent search. The fragments and the frequency of occurrence in a compound are easily found 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

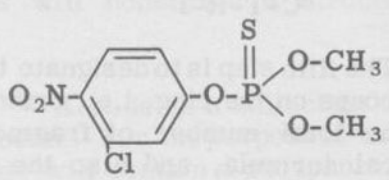
Since the computer search file is organized in three levels it is possible to ask for the compound by asking for a combination of fragments either specifically or generically. Some of the fragments may be asked for by entering the file in the generic first level and others in the more specific third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

Asking the Question

The procedure for asking the search question is as follows: The examiner studies the compound which is claimed in the application for patent. 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 would be as follows:

Example

It is desired to find the following compound:



The compound would contain the following 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 of the fragments and their relationship and obtain the compound. However, a more generic request could be made, e.g., "for all chlorine and NO_2 containing compounds attached to a ring and containing the configuration "P-O-Phenyl." In such a search, 34 patents in a file of 693 patents would be responsive to the query.

A more specific question would reduce the number of documents retrieved. By introducing the descriptor "12 fragments," for instance, only those compounds containing 12 fragments are acceptable. When this was done only 8 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 and those which are only included by the virtue of being in a Markush group. Upon using all the possible descriptors for the compound, only 7 patents were retrieved.

It should be noted that for most compound questions it is sufficient to ask for 3 to 5 of the most unusual descriptors. The results obtained will be almost as good as if all of the descriptors had been requested. In practice, the average search took about 3-5 minutes of machine time. If the computer were asked the more generic question "For a halogen containing compound having the halogen attached to an aromatic ring and containing the (P-O) group three times" the above compound would be retrieved.

The following is a graphic representation of what might be in the computers 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 instructed to "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 the 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 is then dropped and the four digit number is then transformed into the patent number by a dictionary look-up.

If the instruction is limited to including compounds that contain only 12 fragments, 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

They are then converted to patent numbers by the computer.

An actual search question sheet appears in Appendix 8.

SYSTEM USE

The following statistics reflect experience of this system which is in use in the Patent Office.

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

CONCLUSIONS

Each of the 200 searches made has proved satisfactory.

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 your 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.

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 #7. Washington 25, D. C. Department of Commerce, 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. Washington, D. C., National Academy of Sciences-National Research Council, 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 #13, Washington 25, D. C., Department of Commerce, 1958.

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

(6) Leibowitz, J., J. Frome and D. D. Andrews. *Variable Scope Patent Searching by an Inverted File Technique*, Patent Office Research and Development Report #14, Washington 25, D. C., Department of Commerce, 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 #17. Washington 25, D. C., Department of Commerce, 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 #64, 1951.

APPENDIX 1

260-461.1/2

33 EXAMINER

63

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:

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

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:

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

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-

duce a mixture of diethyl phenyl and diphenyl ethyl phosphate.

Example 4.—Mixed ethyl o-nitrophenyl thiophosphates

A mixture of 39.6 parts of triethyl thiophosphate, 181.1 parts of o-nitrophenyl acetate and 5.0 parts of litharge was reacted according to the procedure in Example 1. After removing excess o-nitrophenyl acetate and a small amount of unreacted triethyl thiophosphate, the residue consisted mainly of diethyl o-nitrophenyl thiophosphate. This crude reaction product without further purification is a valuable pesticide.

Example 5.—Mixed ethyl isobutyl thiophosphates

A mixture of 46.0 parts of triethyl phosphorotrithioate [$(C_2H_5S)_3P{=}O$], 139.0 parts of isobutyl acetate, and 4.0 parts of litharge was reacted according to the procedure in Example 1 to produce a mixture of diethyl isobutyl and diisobutyl ethyl thiophosphates.

Example 6.—Mixed butyl ethyl phosphates

A mixture of 36.4 parts of triethyl phosphate and 139.0 parts of n-butyl acetate was placed in a rocking autoclave and heated at 250° C. for 8 hours. Fractionation of the reaction mixture after removal of ethyl and butyl acetate gave the following fractions:

Cut	Boiling point, °C.	n_D^{30}	Yield in parts
1	75–87 (3.0 mm.)		3.7
2	85–108 (2.8–3.0 mm.)	1.4147	21.0
3	Distillation residue		10.0

Cut 2 contained butyl diethyl phosphate and dibutyl ethyl phosphate.

We claim:

1. The process of preparing neutral mixed phosphates and thiophosphates having 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, haloalkyl, containing up to 8 carbon atoms, phenyl and nitrophenyl, said R and R' being different and at least one of said R and R' being alkyl and wherein X is selected from the group consisting of oxygen and sulfur which comprises reacting a trialkyl phosphate wherein the alkyl radicals contain up to 8 carbon atoms per

alkyl radical with a lower carboxylic acid ester selected from the group consisting of

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

and

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

wherein R and R' are defined the same as above and R''' is a lower alkyl radical at a temperature within the range of 100 to 275° C. and in the presence of a lead oxide catalyst.

2. The process for producing neutral mixed phosphates which comprises reacting triethyl phosphate with n-butyl acetate at a temperature within the range of 100 to 275° C. and in the presence of a lead oxide catalyst.

3. The process for producing neutral mixed phosphates which comprises reacting trimethyl phosphate with octyl acetate at a temperature within the range of 100 to 275° C. and in the presence of a lead oxide catalyst.

4. The process for producing neutral mixed phosphates which comprises reacting triethyl phosphate with phenyl acetate at a temperature within the range of 100 to 275° C. and in the presence of a lead oxide catalyst.

5. The process for producing neutral mixed thiophosphates which comprises reacting triethyl thiophosphate with o-nitrophenyl acetate at a temperature within the range of 100 to 275° C. and in the presence of a lead oxide catalyst.

6. The process for producing neutral mixed thiophosphates which comprises reacting triethyl phosphorotrithioate with isobutyl acetate at a temperature within the range of 100 to 275° C. and in the presence of a lead oxide catalyst.

References Cited in the file of this patent

UNITED STATES PATENTS

1,873,513 Van Loon ---------- Aug. 23, 1932

OTHER REFERENCES

Ralson: Fatty Acids and Their Derivatives, 1947, p. 544, lines 5–9.

Rueggeberg et al.: J. A. C. S., vol. 70, pp. 1802–1804 (1948).

Kosalopoff: Organo-Phosphorus Compounds, 1950, p. 227, sect. VIII.

Groggins: Unit Processes in Organic Syntheses, 1952, pp. 597–8, 609 and 618–9.

APPENDIX 2

The following is 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

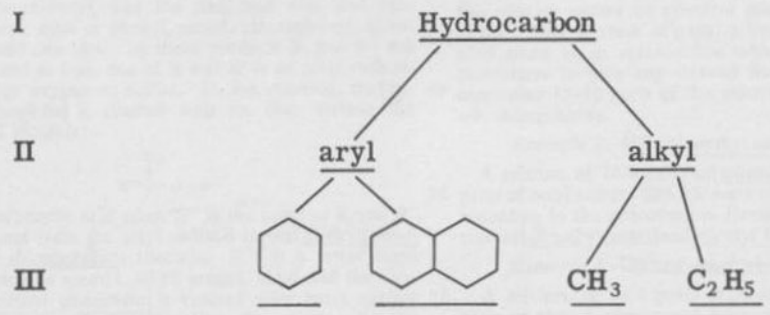
APPENDIX 3

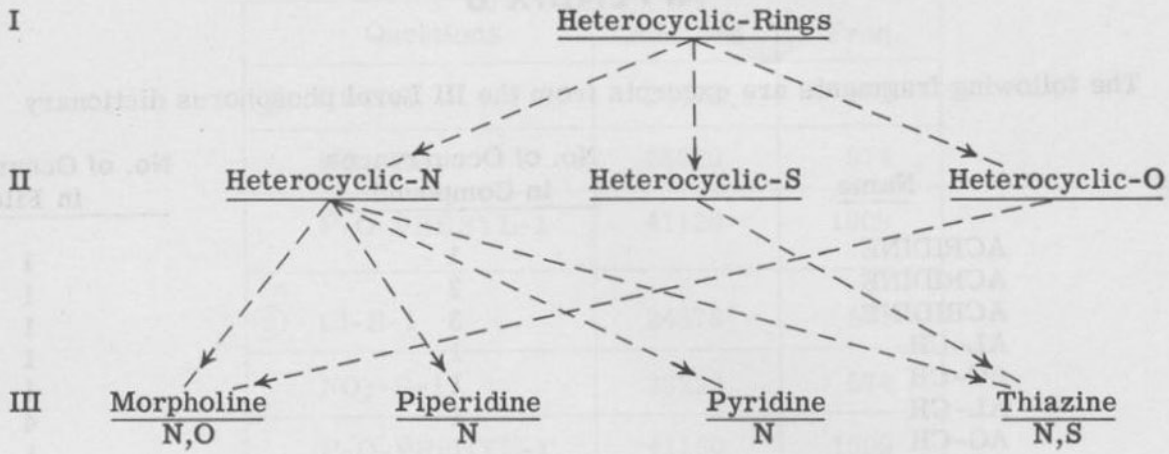
Definition of Terms Used in Fragmenting

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

APPENDIX 4

(Level of Genericity)





APPENDIX 5

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 PHOSPHORUS OXYCHLORIDE
2875229 PHOSPHORYL DICHLORIDE
2875229 TRIALKYL PHOSPHATE
2875229 TRIALKYL PHOSPHATES
2875229 TRIALKYL THIOPHOSPHATE

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

APPENDIX 6

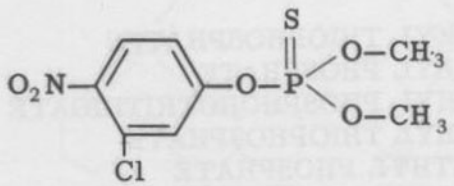
The following fragments are excerpts from the III Level phosphorus dictionary

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-ST-CH	1	1
25891	C2-T-SAT-ST-CH	2	2
25892	C2-T-UNS-ST-CH		1
25893	C2-T-UNS-ST-CH	1	47

APPENDIX 7

EXAMINER	APPLICATION NUMBER	STATUS	DATE

Essence of Invention



	Questions	Address	Freq.
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

APPENDIX 8

	1		1		1		1	1st Node	2	2	3	3	4	4	5	5		6		6		7
									O_1	O_2	O_3	S_1	S_2	S_3	N	Misc A	N		Misc		Mis	
				Cycl	0	Alk	5	Alkyl	12	4	12	4	12	4	12	4			Met	2	C=X	12
				≡	1	=	6	Alkenyl	11	5	11	5	11	5	11	5	Pri	12	Si	3	-X-X	11
				Poly	2	Mono	7	Aryl	0	6	0	6	0	6	0	6	Sec	11	β	4	Cycl P	0
Misc	12	0	11	S	3	N	8	Hetero	1	7	1	7	1	7	1	7	Ter	0	Se	5	Mix	1
				Met	4	H	9	H or Met	2	8	2	8	2	8	2	8	=N	1	Te	6	Sa	2
								Misc	3	9	3	9	3	9	3	9			$C\overset{X}{\underset{X}{=}}$	7	Mis	3

	8		8		9		10	II NODE	11	12	13	14	15	16	17	18	19	20	21	22	23
									O_1	O_2	O_3	S_1	S_2	S_3	N	Alk	Aryl	Het	Misc	CH	R
				Cycl	12	Alk	12	Alkyl	12	12	12	12	12	12	12	12	12	12	12	12	12
				≡	11	=	11	Alkenyl	11	11	11	11	11	11	11	11	11	11	11	11	11
				Poly	0	Mono	0	Aryl	0	0	0	0	0	0	0	0	0	0	0	0	0
Mis	12	0	11	S	1	N	1	Hetero	1	1	1	1	1	1	1	1	1	1	1	1	1
				Acyl	2	OH	2	OH	2	2	2	2	2	2	2	2	2	2	2	2	2
		Met	0	O-O	3	OR	3	-O-	3	3	3	3	3	3	3	3	3	3	3	3	3
		Met	1	Acyl	4	SH	4	SH	4	4	4	4	4	4	4	4	4	4	4	4	4
				S-S	5	SR	5	-S-	5	5	5	5	5	5	5	5	5	5	5	5	5
				=S	6	=O	6	=O, =S	6	6	6	6	6	6	6	6	6	6	6	6	6
		Met	2	R	7	H	7	COOR, MET	7	7	7	7	7	7	7	7	7	7	7	7	7
		N◯	3	R	8	H	8	$CONR_2$	8	8	8	8	8	8	8	8	8	8	8	8	8
				R	9	H	9	NR_2	9	9	9	9	9	9	9	9	9	9	9	9	9
	24		24		24		24		25	25	25	26	26	26	27	27	27	28	28	28	29
				NO_2	2	CN	5	NO_2,CN	12	2	6	12	2	6	12	2	6	12	2	6	12
I	12	F	0	BR	3	Cl	6	Halo	11	3	7	11	3	7	11	3	7	11	3	7	11
R	11	N◯	1	SO_2	4	SO	7	$SO,SO_2,SON\overset{R}{\underset{R}{}}$	0	4	8	0	4	8	0	4	8	0	4	8	0
								Misc.	1	5	9	1	5	9	1	5	9	1	5	9	1

	30		30		31		32	Ter. NODE	33	34	35	36	37	38	39	40	41	42	43	44	45
									O_1	O_2	O_3	S_1	S_2	S_3	N	Alk	Aryl	Het	Misc	CH	R
				Cycl	12	Alk	12	Alkyl	12	12	12	12	12	12	12	12	12	12	12	12	12
				≡	11	=	11	Alkenyl	11	11	11	11	11	11	11	11	11	11	11	11	11
				Poly	0	Mono	0	Aryl	0	0	0	0	0	0	0	0	0	0	0	0	0
Mis	12	0	11	S	1	N	1	Hetero	1	1	1	1	1	1	1	1	1	1	1	1	1
				Acyl	2	OH	2	OH	2	2	2	2	2	2	2	2	2	2	2	2	2
		Met	0	O-O	3	OR	3	-O-	3	3	3	3	3	3	3	3	3	3	3	3	3
		Met	1	Acyl	4	SH	4	SH	4	4	4	4	4	4	4	4	4	4	4	4	4
				S-S	5	SR	5	-S-	5	5	5	5	5	5	5	5	5	5	5	5	5
				=S	6	=O	6	=O, =S	6	6	6	6	6	6	6	6	6	6	6	6	6
		Met	2	R	7	H	7	COOR, MET	7	7	7	7	7	7	7	7	7	7	7	7	7
		N◯	3	R	8	H	8	$CONR_2$	8	8	8	8	8	8	8	8	8	8	8	8	8
				R	9	H	9	NR_2	9	9	9	9	9	9	9	9	9	9	9	9	9
	46		46		46		46		47	47	47	48	48	48	49	49	49	50	50	50	51
				NO_2	2	CN	5	NO_2,CN	12	2	6	12	2	6	12	2	6	12	2	6	12
I	12	E	0	BR	3	Cl	6	Halo	11	3	7	11	3	7	11	3	7	11	3	7	11
R	11	N◯	1	SO_2	4	SO	7	$SO,SO_2,SON\overset{R}{\underset{R}{}}$	0	4	8	0	4	8	0	4	8	0	4	8	0
						·		Misc	1	5	9	1	5	9	1	5	9	1	5	9	1

	52	II NODE -52			52		53	Ter. NODE -53	53							
≡	12	=	0	Cycl	2	Alk	5	Alkyl	12	Alk	1	Cycl	4	=	6	≡
				Poly	3	Mono	6	Aryl	11	Mono	2	Poly				
Mis	11	0	1	S	4	N	7	Hetero	0	N	3	S	5	0	7	Misc

54		55		56		57	
12	X=P-(X_3)	12			R	12	Mono Ester
11	$O\diagdown P\diagup X$ (O, X)	11		12	3AR	11	Di-Ester
0		0	1+S	11	2AR	0	Tri Ester
1	=X,O	1	4S	0	1AR	1	Mixed Ester
2	=S,X	2	=S;2S;0	1	3 Acyclic	2	Cellulose
3	=O,X	3	=S;S;2-0	2	2 Acyclic	3	Resin
		4	=S;3-0	3	1 Acyclic	4	Poly-P
		5	=0;3S	4	3 Alicyclic	5	Salt
		6	=0;30	5	2 Alicyclic	6	Mono Acid
		7	=0;2S;0	6	1 Alicyclic	7	Di-acid
		8	=0;S;2-0	7	3 Hetero	8	Misc
		9	4-0	8	2 Hetero		
				9	1 Hetero		

Other Phosphorus Nucleus

58		59		60		61		62		63		64	
12	P-N	12	P-Hal	12	P-R	12	P-S	12	P-O	12	P=Misc	12	P-(X)-P
11	1	11	1	11	1	11	1	11	1	11	1	11	1
0	2	0	2	0	2	0	2	0	2	0	2	0	2
1	3	1	3	1	3	1	3	1	3	1	3	1	3
2	4	2	4	2	4+	2	4	2	4	2	4	2	4
3	5	3	F	3	P=C	3	5	3	5	3	5	3	5
4	P=N	4	BR							4	Met		
		5	I	4	Cyclic-P	4	P-XR	4	P=O	5	H	4	Phos 3
5	Poly-P	6	Cl	5	3	5'	Mono Ester	5	1	6	Se	5	Phos 5
6	2			6	4	6	Di Ester	6	2+	7	Te	6	
7	3	7	P-P	7	5	7	Tri Ester			8	Cellulose	7	2+
8	4	8	Polymer	8	6+	8	Mono Acid	7	P=S	9	Misc	8	
9	5+	9	Resin	9	All C	9	Di Acid	8	1			9	
								9	2+				

Fragments

65											
12	Alkyl	66		68		70		71	
11	L	5	Heterocyclic	2	C,O	12	O-Cont		
0	H	6	Sat.-Het.	3	C=O,CHO	11	-O-	6	N Misc
1	Alkylene	7	5M	4	O=C-OH,Met	0	=O	7	O Misc
2	1	8	6M	5	O=C-OR	1	-O-O-	8	S Misc
3	2	9	Other	6	X=C-X-X=Het	2	OH	9	Misc
4	3+ 67	12	Unsat.-Het.	7	X=C-X-X-IS	3	O-MET		
5	Alkenyl	11	5-M	8	O=C-N=				
6	C=C	0	6-M	9	C=S,CHS	4	Halogen		
7	Alkinyl	1	Other 69	12	S=C-N=	5	Cl		
8	C≡C	2	Unsat-1	11	C≡N,ISO	6	Br		
66		3	Unsat-2+			7	I		
12	Aryl			0	S- Cont.	8	F		
11	Benz	4	N-Hetero	1	-S-	9	Polyhalo		
0	Naph	5	S-Hetero	2	=S	71			
1	Cycloalkyl	6	O-Hetero	3	-S-S-	12	Amine		
2	Cyclohexyl	7	Mis-Hetero	4	SH	11	NH_2		
3	Cycloalkenyl	8	Hetero-1	5	S-MET	0	N-Sec		
4	Cyclohexene	9	Hetero-2+	6	SO_3H,MET	1	N-Ter		
	68	12	Same	7	SO_2N=	2	N-Quat		
		11	Diff	8	SO_4	3	NO_2		
		0	Mono R			4	=N imine		
		1	Ring Sys.			5	Amine Salt		

R & D, U. S. Patent Office "CAMP" - J. Frome - 6/22/60

Chemist_____ Patent No._____

PART II

A PUNCHED CARD SYSTEM FOR PHOSPHORUS COMPOUNDS (CARD MECHANIZATION OF PHOSPHORUS-CAMP)

INTRODUCTION

A low cost mechanized system for searching phosphorus compounds is described. This system was developed with the important factors of machine cost, machine availability, and search capability in a narrow field of organic chemistry predominantly in view. Punched cards and widely available card sorting machinery are used to provide an effective and economical system which serves the same ends as described in RAMP, Part I. The machine used in the application described is a Census Multicolumn sorter which is comparable in most respects to an IBM 101, which could be adapted to this use.

ART SELECTED

The art selected for "CAMP" includes all the patents in class 260, subclass 461 including official and unofficial cross references as classified in the U. S. Patent Office. This is the same art which was used in "RAMP" above.

SYSTEM

A system for retrieval of organic compounds should have at least two features:

(1) the division of the compounds into building blocks (i.e., NH_2, COOH, SO_3H)

(2) a method of showing relationship between the building blocks.

The building blocks used in this system are those ordinarily recognized by chemists and for this particular system are listed infra and are called the fragment dictionary.

Relationships are extremely important and are shown in this system by the use of a matrix, i.e.,

	O_1	O_2	O_3	S_1	
alkyl	X				
aryl					

Thus if an alkyl group is attached to a single oxygen group (O_1) this relationship is indicated in the matrix by a cross in the box at the intersection. This box is given a column and row number. If an aryl group is connected to a sulfur atom (S_1), the box corresponding to the intersection of the horizontal row representing aryl and the vertical row representing S_1 is then marked, i.e.,

	O_1	O_2	O_3	S_1	
alkyl					
aryl				X	

This box is also given a column and row number. By a further extension of this device, we are able to show further relationships; see the 2nd and terminal nodes infra.

DEFINITION OF TERMS

FRAGMENT—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.

NODE—A node is a collection of at least two fragments. For the purpose of this system there will be three nodes: 1st, 2nd and terminal node.

GENERAL CODING PRINCIPLES

"CAMP" relies on a matrix to show relationships, in addition it describes fragments contained in the organic phosphorus compounds disclosed in the documents. Two coding sheets were used in most cases for every document analyzed. The first coding sheet was used for specific formulas disclosed, while the second coding sheet was used for "Markush" formulations only.

It is important to note that all specific formulas disclosed in each document are composited and coded on one coding sheet, and only one punch card is used to describe all the specific formulas disclosed in a document. The same is true for all Markush formulas.

The "CAMP" coding sheet at present is divided into five (5) sections. The first three sections include the I, II and terminal nodes, the fourth section is the fragment dictionary and the fifth section shows both the specific and generic combinations and permutations of the organic phosphorus nucleus which at present encompass the organic phosphates and thiophosphates.

NODAL DESIGNATION AND DESCRIPTION

The 1st node has ten generic chemical terms and a miscellaneous in the "Y" or horizontal axis as shown by cols. 1-8, and twelve descriptive characters in the "X" or vertical axis as shown in rows 12-6.

PAT. NO. _____

1st NODE	12 O_1	11 O_2	0 O_3	1 S_1	2 S_2	3 S_3	4 X_1	5 X_2	6 X_3	7 SY	8 USY	9 HIGH
Alkyl 1												Hi
Sub. Alkyl 2												Hi
Alkenyl Alkinyl 3												— —
Sub. Alkenyl Sub. Alkinyl 4												— —
Unsub. Aryl 5												Poly
Sub. Aryl 6												Poly
H or Metal 7												Poly
Misc. 8												HET.

A

The descriptive characters in rows 12 thru 6 refer directly to the generic chemical terms in cols. 1-8. In rows 12, 11 and 0, O_1, O_2, O_3 represent the oxygen atoms or atoms directly attached to the phosphorus atom either once, twice or three times. In rows 1, 2 and 3, S_1, S_2, and S_3 represent the sulphur atom or atoms directly attached to the phosphorus atom either once, twice or three times. In rows 4, 5 and 6, X_1, X_2, and X_3 are generic characters and represent either an oxygen or sulphur atom or atoms directly attached to the phosphorus atom either once, twice or three times.

The rows 12 thru 6 in the "X" or vertical axis in the 1st node are designed to show the nodal relationship with the generic chemical terms in cols. 1 thru 8.

In row 7 "SY" represents symmetrical and in row 8 "USY" represents unsymmetrical. These descriptive characters are used only in the 1st node and are designed to further distinguish each of the three 1st nodal connections.

Row 9 is used to further define the generic chemical terms in cols. 1 thru 8.

The 2nd node has 25 specific and generic chemical fragments and terms in the "Y" or horizontal axis (cols. 9-28). Col. 29 covers every other possible chemical term or fragment not included in Cols. 9-28. There are in the 2nd node, 12 descriptive characters in the "X" or vertical axis. Each chemical term in the "Y" axis is further defined by row 9 of the "X" axis.

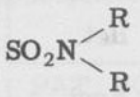
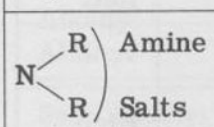
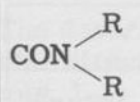
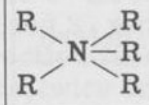
The rows 12 thru 6 in the II Node are designed to show the nodal relationship with the generic chemical terms in Col. 9-28 and are the same as the 1st node. In row 7, the abbreviation "CH" represents the attachment of the fragment being coded to a chain, and in row 8, "R" represents the attachment of the fragment being coded to a ring.

The terminal node is almost the same as the II node. The only difference being there are less generic and specific terms and fragments included within it.

2nd NODE		O_1	O_2	O_3	S_1	S_2	S_3	X_1	X_2	X_3	CH	Ri	
SH; SR	9												R
SO; SO_2	10												SO
SO_2(NR)(NR)	11												Het.
OH-(Acyl-O-)	12												Acyl
Alkenyl Alkinyl	13												Akyl-inyl
N(R)(R) Amine / Salt	14												Het.
NO_2	15												
Alkyl	16												High
Aryl	17												Poly
Alkylene	18												High
Arylene	19												Poly
Halogen	20												IF
-S-	21												
-O-	22												
COOR; H	23												Ester
CON(R)(R)	24												Hetero
CN	25												
OR	26												R
R(R)N(R)(R)R	27												Hetero
=O; =S	28												=S
Misc.	29												HET.

A

Examiner

LAST NODE		O_1	O_2	O_3	S_1	S_2	S_3	X_1	X_2	X_3	CH	Ring	
-S-H; S-R	30												R
SO; SO_2	31												SO
	32												Het.
Alkenyl Alkinyl	33												Alky-enyl
OH; O-Acyl	34												O-Acyl
=O; =S	35												=S
	36												Het.
NO_2	37												
Alkyl	38												Higher
Aryl	39												Poly
Halogen	40												If
COOR; H	41												Ester
	42												Het.
CN	43												
	44												Het.
OR	45												Aryl.
Misc.	46												Hetero

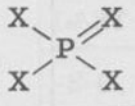
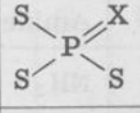
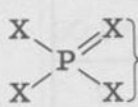
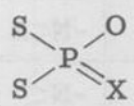
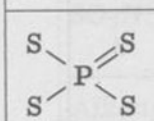
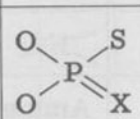
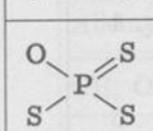
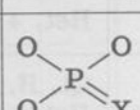
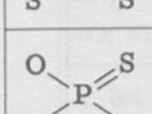
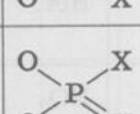
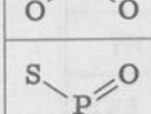
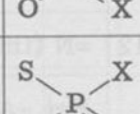
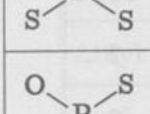
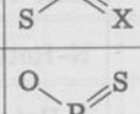
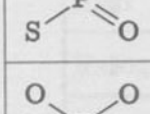
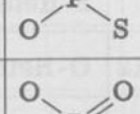
A

The fourth section of the coding sheet is the fragment dictionary, which is used to further define the fragments in the II and terminal nodes. In the situation where the structural formula being coded has more than III nodes, the fragments in the formula not contained in either the II or terminal nodes will be checked in the fragment dictionary also.

	50		51		52
Alkyl	12	-S-	2	Amine	11
Low Alkyl	11	=S	3	NH_2	0
High Alkyl	0	$SO_2N<$	4	-N<H	1
Alkylene	1	-SO-	5	-N<	2
1	2	$-SO_2-$	6	>N<	3
2	3	SO_3; H; Me	7	Amine R Het. + N—R	4
3+	4	SH; SMe	8	Qt. H. + AN R Het. N R R	5
Alkenyl	5	-S-S-	9	NO_2	6
C=C	6	SO_4	52/12	=N (imine)	7
Alkinyl	7		52	Het.-A	53
C≡C	8	-O-	8	N-Hetero	2
Aryl	9	=O	9	S-Hetero	3
Benzene	51/12	OH	53/12	O-Hetero	4
Naphthalene	11	OME	11	N-O-Hetero	5
Cycloalkyl	0		0	N-S-Hetero	6
Cyclohexyl	1		1	S-O-Hetero	7
	54		55		
Halogen	0	N-Hetero	12	O-Hetero	8
Cl-1	1	Pyridine	11	Furan	9
Cl-2	2	Piperidine	0	Mis-O-HET.	56/12
Cl-3	3	Pyrrolidine	1		
Cl-4+	4	Morpholine	2	Hetero-Sat-N-6M	11
F	5	Thiazole	3	Misc. Het.	0
Br	6	Misc.-N HET.	4	Quinoline	1
Br +2	7				
I	8	S-Hetero	5	Sat.N-Het.	2
Poly Halo	9	Thiophene	6	3 M	3
		Misc. S HET.	7	4 M	4
				5 M	5
Mis-	56/6				

48

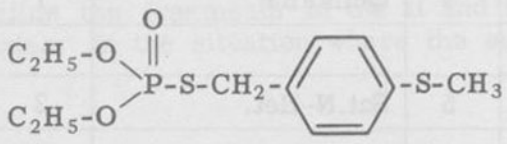
C=O CHO	2
C(=O)OH; ME	3
C(=O)—OR	4
C(=X)(X) I-S	5
C=S CHS	6
C(=O)—N<	7
C(=S)—N<	8
CN ISO.	9

	47		47
	12		6
 I-S NO Triaryl	11		7
	0		8
	1		9
	2		48/12
	3		11
	4		0
	5		1

The fifth section of the coding sheet represents the organic phosphorus nucleus. This section is the starting point in the coding of every structural formula. All of the essential combinations and permutations of the organic phosphorus nucleus both generic and specific are included here. It is well to note that these combinations and permutations only include the organic phosphates and thio phosphates.

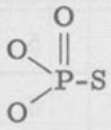
All the nodes, the fragment dictionary and the organic phosphorus nucleus will be described further by a specific example.

EXAMPLE OF ENCODING



The encoding of structural formulas in the "CAMP" system begins by separating the organic phosphorus nucleus from the rest of the compound. The organic phosphorus nucleus is defined as the phosphorus atom and all the elements directly connected to it.

For coding purposes, when a specific formula is disclosed, the analyst will code the formula both specifically and generically in so far as the organic phosphorus nucleus is concerned and also the oxygen and sulphur atoms which are directly attached to the phosphorus atom. In brief, the phosphorus nucleus will always be coded both specifically and generically when a specific formula is disclosed, and the oxygen and sulphur atom or atoms directly attached to the phosphorus atom will also be coded both specifically and generically in the 1st, 2nd and terminal nodes. In the example above it would be done in this manner.

 Col. 47 Row 5

The analyst would code Col. 47 Row 5 on the coding sheet to show the specific organic phosphorus nucleus, and also check Col. 47 Row 12, Col. 47 Row 11, Col. 47 Row 8, and Col. 48 Row 12. The reason the generic phosphorus nucleus is coded in addition to coding the specific configuration is to enable the searcher to ask a more generic question. So in this example, the phosphorus atom has directly connected to it a [=O], [O], [O] and an [S].

To designate the 1st nodal relationship, the fragment or fragments which are directly attached to either the oxygen or sulphur atoms are by the combination of the two (oxygen or sulphur and the fragment) the 1st node. In the example above then, the 1st node would be coded on the coding sheet in this manner.

O_2- (C_2H_5) or [O_2 - (Alkyl)]	Col. 1 Row 11
SY	Col. 1 Row 7
S_1 - CH_2 or [S_1-Alkyl]	Col. 2 Row 1
USY	Col. 1 Row 8

The reason O_2 - (C_2H_5) or O_2-Alkyl was coded (Col. 1 Row 11) is that both oxygen atoms were directly connected to the ethyl or alkyl fragment, SY (symmetrical) Col. 1 Row 7 was coded because the ethyl fragments attached to the oxygen atoms are identical. (Col. 2 Row 1) is self explanatory. USY (unsymmetrical) (Col. 2 Row 8) was coded because the sulphur and the alkyl fragment are not identical to the other 1st node connections. The first node will also be coded generically for reasons stated

above, so Col. 1 Row 5 is coded and also Col. 2 Row 4.

The II node in the example above is the combination of the sulphur or oxygen atom or atoms and the fragment twice removed from it, which in this example would be S-  - only, since there is only one fragment connected to the oxygen atoms. The second node would then be coded specifically as follows:

S-  or [S_1-Sub. Aryl] Col. 17 Row 1

CH Col. 17 Row 7

S-  was coded specifically in Col. 17 Row 1 because the sulphur atom was connected to the phenyl fragment. CH (chain Col. 17 Row 7) was coded since the phenyl is attached to a chain. As stated above in the II and terminal nodes, the fragment will also be coded in the fragment dictionary. So Col. 50 Row 9 and Col. 51 Row 12 will also be coded.

The second node must be coded generically so Col. 17 Rows 4 and 7 must be coded also.

The terminal nodal relationship is a combination of the sulphur or oxygen atom or atoms with the fragment furthest removed in the structural formula. The terminal node is the final fragment, but in order to be terminal it must be further removed than the 2nd node. In this case it would be S_1-[S-CH_3] or S_1-[SR].

S_1-[S-CH_3] or S_1-[SR] Col. 30 Row 1

R Col. 30 Rows 8 and 9

The reason we have combined two fragments here is that in Col. 30 the S-R represents sulphur which has directly attached to it a hydrocarbon radical, whether it is a hydrocarbon ring or chain. Col. 50 Row 11 and Col. 51 Row 2 will also be coded in the fragment dictionary. Col. 30 Rows 4, 8 and 9 are the generic codes for the terminal node.

An example of the finished coding sheet of the above structural formula will be shown below.

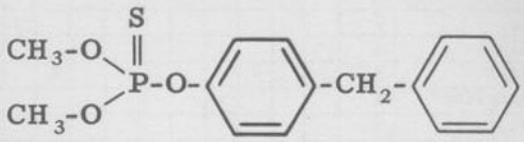
ASKING THE QUESTION:

The procedure for asking the search question is as follows: The examiner, after analyzing the applicant's claims, determines what the essence of invention is and by using the matrix system of nodal relationships previously described, asks the question by checking the appropriate blocks on the coding sheet, depending on whether a specific or generic answer is desired. The board is then wired by the machine operator and the file of cards is fed into the Multicolumn sorter and the cards representing the documents are retrieved.

A specific illustration of an actual search would be as follows:

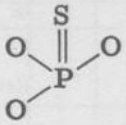
Example

It is desired to find the following compound



The above compound will be coded specifically as follows:

Organic Phosphorus Nucleus

 Col. 47 Row 2

I Node

O_2-(Alkyl)	Col. 1 Row 11
SY	Col. 1 Row 7
O_1-Sub. Aryl	Col. 6 Row 12
USY	Col. 6 Row 8

II Node

O_1-Alkylene	Col. 18 Row 12
Ring	Col. 18 Row 8

Terminal Node

O_1-Aryl	Col. 39 Row 12
Chain	Col. 39 Row 7

It is well to note that when we say the compound is being asked specifically, we mean the organic phosphorus nucleus and those elements directly attached to the phosphorus atom are asked specifically; but all other fragments in

PAT. NO._____

1st NODE		12 01	11 02	0 03	1 S1	2 S2	3 S3	4 X1	5 X2	6 X3	7 SY	8 LSY	9 HIGH	
Alkyl	1		X						X		X		HI	
Sub. Alkyl	2				X			X				X	HI	
Alkenyl Alkinyl	3												═══	
Sub. Alkenyl Sub. Alkinyl	4												═══	
Unsub. Aryl	5												Poly	
Sub. Aryl	6												Poly	
H or Metal	7												Poly	
Misc.	8												HET.	A

2nd NODE		01	02	03	S1	S2	S3	X1	X2	X3	CH	Ri		
SH: SR	9												R	
SO; SO_2	10												SO	
SO_2(NR, NR)	11												Het.	
OH-(Acyl-O-)	12												Acyl	
Alkenyl Alkinyl	13												Alkyl-inyl	
N(R, R) Amine Salt	14												Het.	
NO_2	15													
Alkyl	16												High	
Aryl	17				X			X			X		Poly	
Alkylene	18												High	
Arylene	19												Poly	
Halogen	20												IF	
-S-	21													
-O-	22													
COOR; H	23												Ester	
CON(R, R)	24												Hetero	
CN	25													
OR	26												R	
N(R, R, R, R)	27												Hetero	
=O; =S	28												= S	
Misc.	29												HET.	A

Examiner_____

LAST NODE		01	02	03	S1	S2	S3	X1	X2	X3	CH	Ring		
-S-H; S-R	30				X			X				X	R X	
SO; SO_2	31												SO	
SO_2N(R, R)	32												Het.	
Alkenyl Alkinyl	33												Alky-inyl	
OH; O-Acyl	34												O-Acyl	
=O; =S	35												=S	
N(R, R) Amine Salts	36												Het.	
NO_2														
Alkyl	38												Higher	
Aryl	39												Poly	
Halogen	40												If	
COOR; H	41												Ester	
CON(R, R)	42												Het.	
CN	43													
N(R, R, R, R)	44												Het.	
OR	45												Aryl.	
Misc.	46												Hetero	A

	(47)		(47)
P(X, X, X, X)	(12)	P(S, S, S, X)	6
P(X, X, X, X) 1-S N.O Triaryl	(11)	P(S, S, O, X)	7
P(S, S, S, S)	0	P(O, O, S, X)	(8)
P(O, O, S, S)	1	P(O, O, O, X)	9
P(O, O, O, S)	2	P(O, O, X, X)	(48/12)
P(S, S, O, S)	3	P(S, S, X, X)	11
P(O, O, S, O)	4	P(O, O, S, S)	0
P(O, O, O, O)	(5)	P(O, O, O, O)	1

	48
C=O CHO	2
C(=O)(OH;ME)	3
C(=O)-OR	4
C(X, X) 1-S	5
C=S CHS	6
C(=O)-N	7
C(=S)-N	8
CN ISO.	9

	(50)		(51)		(52)
Alkyl	(12)	-S-	(2)	Amine	11
Low Alkyl	(11)	=S	3	NH_2	0
High Alkyl	0	SO_2N	4	-N=H	1
Alkylene	1	-SO-	5	-N<	2
1	2	$-SO_2-$	6	>N<	3
2	3	SO_3; H; Me	7	Amine Het. + N(R)-R	4
3 +	4	SH; SMe	8	Qt. H. + AN(R)(R) Het.	5
Alkenyl	5	-S-S	9	NO_2	6
C=C	6	SO_4		=N (imine)	7
Alkinyl	7		52	Het. -A	53
C≡C	8	-O-	8	N-Hetero	2
Aryl	(9)	=O	9	S-Hetero	3
Benzene	(51/12)	OH	(52/12)	O-Hetero	4
Napthalene	11	OME	11	N-O-Hetero	5
Cycloalkyl	0		0	N-S-Hetero	6
Cyclohexyl	1		1	S-O-Hetero	7
	54		55		
Halogen	0	N-Hetero	12	O-Hetero	8
Cl-1	1	Pyridine	11	Furan	9
Cl-2	2	Piperidine	0	Mis-O-HET.	(53/12)
Cl-3	3	Pyrrolidine	1		
Cl-4+	4	Morpholine	2	Hetero-Sat-N-6M	11
F	5	Thiazole	3	Misc. Het.	0
Br	6	Misc.-N HET.	4	Quinoline	1
Br +2	7				
I	8	S-Hetero	5	Sat. N-Het.	2
Poly Halo	9	Thiophene	6	3 M	3
		Misc. S HET.	7	4 M	4
				5 M	5
Mis-	(56/6)				

the formula, except in a few instances, are usually generic to those elements insofar as the nodal relationship is concerned.

To ask the above question generically Col. 47 Row 12 would be coded in the organic phosphorus nucleus section. In all the nodes the same column would always be checked but instead of checking the rows O_1 or O_2 or S_1 you would check X_1 and X_2.

SYSTEM USE

This system at present is in experimental use in the U. S. Patent Office where approximately 150 searches have been made to date.

The following are some of the statistics relating to these searches:

1.	Applications Searched	52
2.	Total Searches Made	104
3.	Searches per Application	2
4.	Time to Prepare Search Question	3 minutes
5.	Machine Time per Search	2 minutes
6.	Time to Wire Machine	3 minutes
7.	Patents Retrieved per Search	21

It is well to note that "CAMP" has already been revised in order to take care of all organic phosphorus compounds other than phosphates and thiophosphates in class 260 subclass 461 (appendix 8). (R & D Report #22.)

CONCLUSION

It is felt that the matrix system using nodal relationships on a low cost punched card machine may be for certain chemical compound arts one of the answers to low cost mechanized searching.

27232--U.S.Dept.of Comm--DC--1962