A Notation-Based Fragment Code for Chemical Patents

JOHN A. SILK
Imperial Chemical Industries Limited, Jealott's Hill
Research Station, Bracknell, Berkshire, England

Received May 21,1968

A system of fragment coding for organic compounds is described which uses a notation to construct meaningful code terms for molecular fragments. This approach provides greater flexibility and specificity than a conventional system with a predetermined set of fragments, and it has been applied successfully and economically to the indexing of pesticide and allied patents using fixed field coding on punched cards. A supplementary computer system providing ordered lists of formalized one-line abstracts is also described.

Fragment codes for organic chemical compounds have proved to be the most useful and economical method of recording structural units in a manner permitting retrieval of compounds by class for purposes of analysis and correlation. The major limitations of such codes in their commonest form, namely, as direct codes on punched cards, arise from the comparatively small number of features which can reasonably be incorporated without making the system cumbersome to use and the limited capability for recording modes of linkage of fragments.

This paper describes an approach toward greater specificity by using a notation to construct systematic code terms for molecular fragments and thereby obtaining an open-ended system which is controlled by a relatively simple body of rules. It is, therefore, a pioneer example of an approach now being more fully exploited through the computer processing of chemical ciphers.[1] The particular application to pesticide patents which is reported on was set up in 1961–62, and it operates with fixed field coding on punched cards, so that it requires only a multiposition, pattern-selecting sorter for operation. Although it has more recently been supplemented by a computer-based system, its economy and efficiency favor its continuance as an operational system which currently provides the means of searching close to 20,000 patents relating to pesticides, herbicides, and related chemistry.

Experience with the ICI fragment code,[2] which has about 280 features in it, has shown that, while it was feasible to increase the number to obtain greater specificity, this was accompanied by an increase in complexity which demanded more training of operators and led to a higher error rate, mainly through omissions. It was felt that a chemical notation, on the other hand, provided a systematic terminology with the aid of which a person familiar with a few basic symbols and the rules governing their use could construct systematic codes for a wide variety of molecular units and thus obtain improved differentiation more readily than with a detailed system of predetermined fragments.

Since the body of information to be processed consisted of patent specifications, scope was also needed for covering the alternative structures which are generally embraced by one specification. Use of a notation fulfilled another prerequisite, namely, that the codes were meaningful and so could be read back to obtain information; a fragment code represented by arbitrarily assigned numbers or holes on a punched card lacks this attribute.

With this system seven fields of eight columns each were set up on standard 80-column cards. Four fields were assigned to chemical coding, while the other three recorded patent number, patentee and type of use, and other information. The chemical code was derived mainly from the author's "Linear Notation for Organic Compounds,"[3] although considerable modification was needed to arrive at a system suitable for fixed field coding. One field was allocated to ring systems and one each to characteristic groups on heterocyclic, aromatic, and aliphatic/alicyclic structures.

The manner of coding groups is identical in these latter three fields, and for the purpose of illustration the following symbols will be considered.

Q	—O—(linking oxygen)
O	=O or C=O (terminal oxygen or carbonyl)
M	—N< (amino or amido type nitrogen)
P	P (phosphorus)
T	—S—(divalent linking sulfur)
S	=S (terminal divalent sulfur) or tetra- or hexavalent sulfur
5	lower alkyl substituent (up to C_5)
7	aryl substituent
9	hydrogen (= terminal)

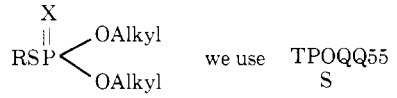
Linking oxygen is present in alcohols and phenols (as ROH) and in ethers and esters (as ROR′ and RCOOR′, respectively). The hydroxyl group is thus coded as Q9 and an ether having a lower alkyl substituent as Q5. All codes are punched in the appropriate field, starting in the first column of that field. For all functions which link two or more rings and chains, the code is assigned to the field for the senior structure, heterocyclic being senior and aliphatic being junior. An aryl lower-alkyl ether must, therefore, be represented by Q5 in the aromatic field, not by Q7 in the aliphatic field.

Terminal oxygen occurs most frequently in the carbonyl group, and, less commonly, attached to N, S, or P. On its own, therefore, the symbol O is understood to represent the C and O of a carbonyl group, but with the other symbols it denotes simply oxygen. The code O9 thus represents an aldehyde, and O5 in the aromatic field would

represent an aryl lower-alkyl ketone, such as acetophenone. The carboxyl group is logically coded as OQ, so that a benzoic acid would be (Ar)OQ9, where (Ar) denotes the field for aromatic substituents. A lower-alkyl benzoate would be (Ar)OQ5 and a phenyl benzoate would be (Ar)OQ7. The phenyl ester of a lower aliphatic acid, on the other hand, is ciphered in the reverse manner as (Ar)QO5; (Alk)OQ7 would be incorrect in not assigning the code to the senior field.

Amides are ciphered similarly with O and M. For an *N*-unsubstituted benzamide, the code is (Ar)OM99 and for an acetanilide it must be reversed, giving (Ar)MO5. Ureas are coded MOM (corresponding to N + CO + N) in the appropriate field, while a carbamate may be QOM or MOQ, depending on whether the senior substituent is attached to the oxygen or the nitrogen atom. If both substituents are the same class, the cipher which comes later in alphabetical order is preferred—i.e., QOM.

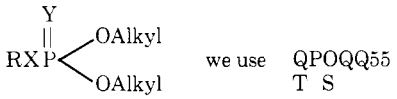
Clearly, this system of ciphering enables the "way-round" of unsymmetrical complex groups to be shown, and this provides a greater degree of specificity than is possible with direct coding, unless many separate terms are allotted to make such distinctions. This approach has been particularly useful in distinguishing the types of phosphorus esters which are important in insecticide chemistry, as is shown in Table I. In these examples, it may be noted also that for complex groups coded by several symbols, all letter symbols for atoms are written as a set before adding the numerals 5–9, which signify the further types of substituents (in addition to the major one determining the field for coding). It is then ruled that the numerals should be cited in the order corresponding to that of substituent atoms in the cipher for the group. Markush formulas of the type commonly encountered, for example, in insecticidal phosphorus esters, are easily encompassed by superimposing codes for the alternatives in appropriate columns of the punched card field. For



we use TPOQQ55 (S)

(where X = O or S)

and for



we use QPOQQ55 (T S)

(where X and Y = O or S)

Where there are alternatives in more than one column, superimposed codes may be imprecise. If in the last example, the patent specified that at least one of X or Y must be sulfur, the coding would still be that shown, although the simple phosphate ester was excluded. In practice, this is rarely a significant disadvantage.

Designed as it was for fixed field coding, code terms for groups assigned to the same field must be superimposed in punching, and this clearly means that false drops may occur. Two steps were taken to minimize it. First, symbols based on the same number in the punch card representation (for example, E, N, and V are combinations of 5 with different zone punches) were allotted one to a frequently occurring unit (in this case, N) and the others were reserved for rarer terms. Second, for certain very common substituents, coding was started in the fifth column of a field instead of the first. This applies for halogens, alkoxy, amino, and nitro groups. The ICT group selection sorter which is used is able to search on any combination of up to eight positions in a card and, as letters are represented by two punches, this means up to four letters. Alternatively, the numerical part only of a larger set of symbols may be sought, depending on the circumstances.

The complete set of symbols for coding characteristic groups is given in Table II and further examples in Table III. These are derived mainly from the Silk notation, which provides a single letter symbol for each of the halogens. Four additional symbols are used, so that certain important units, which would normally be ciphered with two letters can be coded by a single, more specific symbol; among these is W for O_2, as in the Wiswesser notation. The letter X is also used to denote either a positive charge or a bond between two atoms in situations where the two letters alone have a different significance—i.e., peroxide is QXQ, as distinct from acetal, QQ.

The numerals 5 to 9 denote types of substituent, as already exemplified, while 2 and 3 are used to show multiplication of like substituents on different atoms of a chain or ring. Multiplication on the same atom is shown by repeating the symbol—i.e., CCl_3 becomes CLLL.

The notation enables most of the characteristic groups of organic compounds and their attachments to be ciphered in a systematic manner, which an abstractor can readily learn to use with confidence and with only occasional reference to the instructions. A few uncommon groups,

Table I. Coding of Some Phosphorus Acid Esters

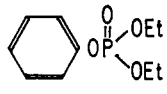
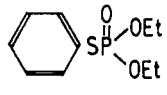
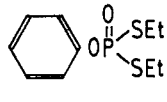
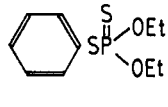
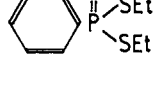
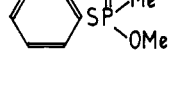
(Ar)QPOQQ55	
TPOQQ55	
QPOTQ55	
TPSQQ55	
POTT55	
TPOQ55	

Table II. Notation for Coding Characteristic Groups

Q	—O— (-ols, ethers, esters)
M	—N< (amines, amides, etc.)
T	—S— (thiols, sulfides)
U	>N—N< (hydrazine)
V	>N—O— (hydroxylamine)
Y	—SS—, —S_n— (di/polysulfide)
O	=O, C=O
N	=N—, >C=N, —N≪
S	=S, S^{IV} or S^{VI}
P	Phosphorus
W	$\overset{=O}{\underset{=O}{}}$
X	Plus
C	Carbon
F	Fluorine
J	Iodine
K	Bromine
L	Chlorine/Halogen
E	Double bond
2	2 like substituents on different atoms of the same chain or ring system
3	3 or more like ditto
5	Lower alkyl substituent (up to C_5)
6	Higher alkyl, cycloalkyl or aralkyl substituent
7	Aromatic substituent
8	Heterocyclic substituent
9	Hydrogen or metal (salt)

Table III. Examples of Ciphers for Characteristic Groups

T5	—SAlk, thioether
T9	—SH, thiol
OT5	—CO. SAlk, carbothiolate
OV	—CO.NH.OH, hydroxamic acid
NM	—C(:NH)NH_2, amidine
EVM	—C(:NOH)NH_2, amide oxime
NO	—NO, nitroso
NW	—NO_2, nitro
NN	—N:N—, azo
NNX	—N_2^+, diazonium
NN	—N(O):N—, azoxy
O	
M99	—NH_2
M55	—$NAlk_2$ } Amines
M59	—NHAlk
M95	Generic code for —NH_2, NHAlk and $NAlk_2$
QOQ	O.CO.O, carbonate
TOQ	S.CO.O, thiolcarbonate
QSQ	O.CS.O, thioncarbonate
TSM	S.CS.N, dithiocarbamate
UNM	N.N.C(:N).N, aminoguanidine
MOUOM	N.CO.N.N.CO.N, biurea
EN	—CN, cyanide
NC9	—NC, isocyanide
NCS	—NCS, isothiocyanate
SEV	>S:N.OH, sulfoxime
TCLLL	—$SCCl_3$, trichloromethylthio
(Ar)EU59	$\overset{Ar}{\underset{Alk}{}}$>C:$NNH_2$
(Ar)UE59	ArNHN:CHAlk } Hydrazones
(Ar)EU09	ArCH:NNHCOAlk

assigned more or less arbitrary codes because of limitations of fixed field coding and searching. Complex composite groups are ciphered both as entities and in terms of their simpler components. A few additional features, which are also desirable to be able to specify in searches, such as benzyl and vinyl units, are covered by simple instructions.

In the notation for ring systems, successive columns record:

(1) Number of rings in a ring system.
(2) Sizes of rings, ignoring six-membered carbocycles.
(3) Heteroatoms other than nitrogen.
(4) Heteronitrogen.
(5) Dispositions of heteroatoms and ring fusions.
(6) Hydrogenation and heteroatom substituents.
(7) Links between ring systems. Details are given in Table IV and examples in Table V.

Codes constructed with this notation are less obviously meaningful than those for functions, since they are mainly numerical, but they can still be readily interpreted by an abstractor who is using them regularly.

In the remaining fields of the cards, mainly nonchemical information is recorded, such as patent number, patentee, and type of use, but changes were made about three years ago when a complementary computer system was introduced. The original system was designed for maximum economy of input, so that a substantial backlog could be dealt with by only one part-time abstractor. Countries were coded by a single letter, which represented both the country and the millions digit of the serial number—e.g., A for U.S. patents from 2,000,000 to 2,999,999 and J for those from 3,000,000. Major companies were assigned two-letter mnemonic codes. Types of use had one-letter codes and other single symbols represented terms like "systemic," "composition claims only," and types of formulation. Another field enabled established products to be recorded by four-letter codes, which again were mainly mnemonic. Finally, a brief note and a formula were written on the card (prior to punching) by carbon copy from the coding sheet. With this system, the average time for an experienced abstractor to read and code a patent was six to seven minutes, and to an appreciable extent this speed was attributable to the economy of writing and coding effort, due to nearly all the information, both chemical and other, being reduced to meaningful alphanumerical codes.

Card punching is relatively complex where codes are superimposed, and a simple hand punch is possibly more suitable than an electric one. However, the division of the card into a number of equal fields provides a series of fixed points from which all codes must start.

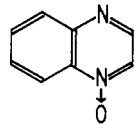
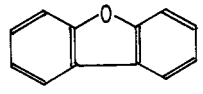
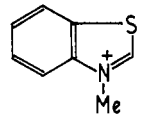
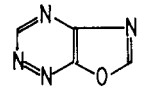
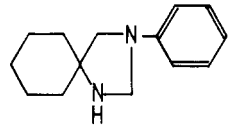
The system has more recently been supplemented with a computer-based one which provides, in particular, comprehensive classified lists of patents. Consequently, the chemical coding on cards is retained in full because it provides searching in depth, so also is the field for established products; the remaining fields have been modified. For each patent, a one-line entry is prepared in a standard format as follows:

NL 6601252 ICI --01FE65 -PI- FURAN, (5-ALKYL)-
2-OXOTETRA-HYDRO-3/4-FURYLPHOSPHON-
(AMIDO)(THIOL)(OTHION)ATE

Table IV. Notation for Coding Ring Systems

Col.		
1	Number of rings in ring system	
	The smallest rings must be chosen	
	The letter instead of the number is used for carbocyclic systems in which all the rings are aromatic	
	A	Benzene
	B	Naphthalene
	C	Anthracene and phenanthrene
	6	All systems with 6 or more rings
	7	For a spiro structure
	8	For a reticular (peri-fused) structure
	9	For a bridged structure
2	Size of each heterocyclic ring, and size of each carbon ring other than six-membered ones (Use 9 for 9 or more)	
3	Heteroatoms other than nitrogen	
4	Nitrogen heteroatom	
	(a) The usual atomic symbols are used; the second letter of a pair goes in col. 4; if it is N, M is punched instead.	
	(b) Multiples are shown by 2, 3 etc. in the same column, except when the use of 1, 2, 3, or 4 in col. 5 suffices.	
	(c) In both columns the absence of heteroatoms must be shown by writing — (a dash) and punching 0 (zero) unless only A is present.	
5	Types of heterocycle	
	1	Only one heteroatom
	2	Two adjacent heteroatoms
	3	Two "other" heteroatoms
	4	Two symmetrical heteroatoms
	5	Three or four adjacent heteroatoms
	6	Three or four symmetrical heteroatoms
	7	Any other number or arrangement of heteroatoms
	8	Heterocyclic end ring
		Use only when three or more rings
	9	Heterocyclic mid ring
6	Hydrogenation; heteroatom substituents	
	9	Heteroatom in iso or angular position—e.g., isoquinoline or quinolizine
	H	Ring system not aromatic
	O	O on heteroatom
	N	N on heteroatom
	X	+ve charge on heteroatom or ± in the ring—e.g., sydnones
	Y	Alkyl, aryl, etc., substituent in heteroatom
	Z	—C—, —C=X or C≡X on heteroatom, where X is any atom, not C or H.
7	Ø	(zero) Two or more ring systems linked directly { like or unlike
	1	Ring systems linked through carbon atom(s)
	2	Two ring systems
		like { systems only—e.g., diphenylmethane, terpyridine
	3	Three or more systems

Table V. Examples of Ring Coding

16-N1	
26-N4O	
35O-1 9	
25SN3X Y	
250N3 6-N7 3	
25-N3HØ 7 A	

The first four parts of this are each of a fixed length (10 spaces for patent number, five for patentee, six for priority date, and four for classification symbols). Following these is a concise, descriptive statement of the contents of the patent, which begins with a keyword expressing a major feature. The statement may contain a maximum of 78 characters, so that a complete entry requires only one line of computer printout. By employing strokes and parentheses to express alternatives and optional features, as shown, an informative expression can usually be obtained in the available space.

Each entry is typed on a punched card, using a Flexowriter, and the card is subsequently punched with both the first four parts of the statement—i.e., excluding the descriptive text—and the chemical fragment and established product codes. The typed statement is useful when cards selected in a search are inspected for relevance, and the Flexowriter tapes, after correction, are fed to a KDF 9 computer and the entries are sorted into alphabetical order based on the keyword at the start of each descriptive text. On each run, new entries are interfiled with previous records, and an updated magnetic tape is produced. The new input is printed out in a single alphabetical list for temporary reference, and at the same time a number of searches are conducted, which may be based on specified sections, part-sections, or combinations of sections of the entries.

Routinely on each run, several searches in a set of about 30 standard ones are conducted to produce comprehensive printed lists of all patents in a major class (Table VI). Typical classes are derived from the class code, in which one symbol in a three-symbol set denotes a broad chemical class of pesticidal interest (for example, L = chlorinated hydrocarbons, P = phosphorus compounds, T = thiocarbamates, V = nitrogen heterocycles), another symbol denotes a broad class of use (for example, F = fungicide, I = insecticide, X = novel chemical compounds not specifically pesticidal), and a third symbol may be for special subdivisions of the use class (for example, IK = acaricide). Thus, we may call for lists covering (1)

Table VI. Printout of a Classified List of Patents

```
JP 10382/7 TOKOA 24MY65 -VN- IMIDAZOLINE-2, 2-ALK(EN)YL AND SALTS + COMPOSITIONS WITH OTHER PESTICIDES
BR 1005900 COMMS 25JA63 -VA- IMIDAZOLINE, 2(1-ALKYL/ARYL-2-HYDROXYALK(EN)YL)-4,4-BIS(H/ALK/ALKOH) + HYDROLYSIS PRODUCT
US 3148211 DU,P- 16JL63 -VH- IMIDAZOLINES 2-OXO 1-PHENYL 3-ALKYL, AND RELATED ACETAL DERIVS OF URON HERBICIDES
US 3173907 AM CY 27NV59 -VX- IMIDAZOLIUMS, 2-AMINO/ALKOXY-ARYL-*AZO-1,3-DI-(*CYANO)ALKYL (DYESTUFFS)
NL 6516360 BAYER 16DE64 -VF- IMIDAZOLONE, (BENZ-), 1(,3-BIS)-DICHLORO FLUOROMETHYLTHIO-(PHARM)
NL 6605131 FISON 15AP65 -VA- IMIDAZOPYRIDINE, 1-ACYL/OXY(THIO)CARBONYL/THIO(THIO)CARBONYL-2-TRIFLUOROMETHYL/PERFLUOROE
NL 6616401 SHELL 24NV65 -VH- IMIDAZOPYRIDINE, 2-TRIFLUOROMETHYL-6-CHLORO-3H-IMIDAZO-
BR 1033775 CSMAK 14AP62 -VA- IMIDE (CYCLIC), OF ETHERIFIED N-HYDROXYMETHYL CARBOXYLIC ACIDS
BR 1033529 MARPH 03DE63 -VOR INDANE-1,3-DIONE, 2(2-PYRIDYL) FROM PHTHALIC ANHYDRIDE * A-PICOLINE, RODENT ICIDE, PHARM
SA 0451/67 RHONE 26JA66 -VF- INDAZOLE, 3-PHENYL AND COMPOSITIONS
NL 6700804 RHONE 26JA66 -VF- INDAZOLE, 3-PHENYL ESP, SCAB, BLIGHT, MILDEW * FABRICS
US 32/7116 SINCL 20DE63 -VHX INDOLE-3-ALKANOIC ACID FROM 1-ISOMER, CAT, WITH METAL SALTS OF STYRENE-MALEIC ANH, RESINS
US 3180875 UPJON 070C63 -VF- INDOLE, 3,3-DITHIOBIS(1-ALKYLINDOLE-2- CARBOHYDRAZIDE) AND DERIVS
BR 0983848 PHILG 12MR60 -VX- INDOLOQUINOLIZINE OCTAHYDRO 3-HYDROXY/ ALKOXYCARBONYLOXY/CYANO-(METHYL) (SEDATIVES)
BG 0653057 AM,CY 110C63 -VBC INDOLOQUINONES, BROADLY COVERED, ANDCORRESPONDING INDOLES
BR 0980997 MIPSI 29JU62 -VY- ISONICOTINIC>=THIOAMIDES,PROCESS FROM NITRILE, H2S AND HYDROSULPHIDES
US 3138523 USRUB 06AP62 -VIK ITACONIMIDE N-M-NITROPHENYL, ACARICIDE
GE 1008739 BAYER 15AP55 -VX- MALEIC HYDRAZIDE MONO-ACYL DERIVS, 4-SUBST, BENZOYL/N-HETEROCYCLIC (CYTOSTATIC)
US 3184475 MONSA 15SE61 -VA- MALEIMIDE, 3-(HALO/ALKYL)PHENYL-THIO/ SULPHINYL/SULPHONYL-N-ALKYL(G,B)
US 3148196 USRUB 17MY62 -VX- MALEIMIDE, N-4-ANILINOPHENYL, ANTIOXIDANT
US 3182072 USRUB 17MY62 -VX- MALEIMIDE, N-(4-ANILINOANILINOMETHYL)- MALEIMIDE
BG 0646321 TOKOA 12AP63 -VEN NITRIFICATION INHIBITORS, AMINO+/CHLORO+/TRICHLORO-METHYL PYRIMIDINES
BR 1023381 TOYAM 11SE62 -VBC OXADIAZOLE-1,3,4 2-AMINO-5(W-(5-NITRO-2- FURYL)-VINYL/1,3-BUTADIENYL
SA 0599/67 BASF- 22FE66 -VH- OXADIAZOLIN-1,2,4 3,5-DIONE-2-ALKYL/BENZYL/ACETYL/ETC-4-PHENYL
BR 1006773 ROCHE 20SE63 -VX- OXAZOLE, N-(4-METHYL-)OXAZOL-5-YL O-(ALKYL/ARYL)CARBAMATES, MEDICINAL
BR 10/2099 MONSA 12SE63 -VH- PHENANTHROLINE-1,10 1,2,3,4-TETRAHYDRO-3,4-DIMETHYL, SELECTIVE
NL 6410620 MONSA 12SE63 -VH- PHENANTHROLINE-1,10 3,4-DIMETHYL-1,2,3,4-TETRAHYDRO, SELECTIVE
NL 6410619 MONSA 12SE63 -VX- PHENANTHROLINE-1,10 3,4-DIMETHYL, PROCESSFROM 8-AMINOQUINOLINE AND METHYL ISOPROPENYL KE
BR 1069489 MONSA 12SE63 -VY- PHENANTHROLINE-1,10 3,4,7,8-TETRAMETHYL
BR 1074889 MONSA 30AU63 -VY- PHENANTHROLINE-1,10 DI/TRI/TETRA-METHYL FROM 8-AMINOQUINOLINES + B-VINYL CARBONYL CPDS
BG 0653001 MONSA 11SE64 -VX- PHENANTHROLINE-3,4-DIMETHYL-1,2,3,4-TETRAHYDRO(H)
BG 0653000 MONSA 11SE64 -VX- PHENANTHROLINE-3,4-DIMETHYL, HERBICIDE INTERMEDIATES
BR 1017844 PLIVA 09AP63 -VY- PHENANTHROLINE(-4,7) 5,6-QUINONE PREPN, ELECTROCHEM, OXIDN OF 6-METHOXY-PHENANTHROLINE
SA 4779/65 SHELL 03SE64 -VA- PHENAZINE, (DI)(N-OXIDE) OPT, HALO/ALKYL/ARYL/ARALKYL ETC, SUBSTD, IK,F,H
NL 6511395 SHELL 03SE64 -VF- PHENAZINE, (MONO/DI-OXIDE)-BZ-H/HAL/ALKYL /ARYL/OR/OH/CN/SCN/NO2/NR2
NL 6501131 FMC-- 03FE64 -VIY PHTHALIMIDE N-OMEGA-ALKYNYL,*SYNERGISTFOR *ALLETHRINE
JP 10384/7 IHARA 29JU65 -VF- PHTHALIMIDE, N-(W-DICHLOROALKENYL)(TETRAHYDRO)PHTHALIMIDE, ESP, RICE BLAST
BR 1038188 FMC-- 03FE64 -VIY PHTHALIMIDE, N-(W-PENTYNYL/HEXYNYL/HEPTYNYL), SYNERGISTS FOR ALLETHRIN/ETC
NL 6600916 SUMIT 25JA66 -VIY PHTHALIMIDE, N-PROPYN-3-YL/BUTYN-3/4-YL/PROPYNYLOXY(M)ETHYL/BUTYNYLOXY(M)ETHYL, SYNERGIST
BR 0987134 MONSA 30MY62 -VH- PHTHALIMIDO-METHYL-(DI)(CYCLO)ALKYL-AMINE, DICOT-SELECTV AND PRE-EM
NL 6502524 ROHM- 04MR64 -VX- PIPERAZIN-2-ONE, 1-VINYL 3-(DI)ALKYL (B,F,OIL ADDITIVES, *COROSINHIB)
NL 6700116 HOECH 21DE62 -VN- PIPERAZINE 1(1,2-BIS-(HALO)PHENYL-/1-BENZYL-2-PHENYL-PROPIONYL)-4-H/(CYCLO/ARYL)ALKYL, AN
NL 6700115 HOECH 21DE62 -VN- PIPERAZINE 1(1,2-BIS-HALOPHENYL-ACRYLOYL)-4-H/(CYCLO/PHENYL)ALKYL, ANTHELMINTIC
US 3152040 FONDR 03JL63 -VI- PIPERAZINE-CARBON DISULPHIDE COMPLEX,FOR KILLING COCKROACHES
US 3186993 ETHYL 07JU62 -VST PIPERAZINE, 1,4-BIS(3,5-DI-ALKYL/ARALKYL-4- HYDROXYPHENYL THIOCARBONYL), ANTOXIDANT
US 3132149 ROHM- 22AP60 -VA- PIPERIDINE/PYRROLIDINE 2-ONES 6/5-ALKOXY 6/5-ALKYL(IK,H)
US 3147267 DU,P- 010C58 -VX- PIPERIDINE-DI*CYANO-2,6 FROM GLUTARALDEHYDE DICYANOHYDRIN (FROM ETHOXYDIHYDROPYRAN)
US 3159639 ABBOT 17SE62 -VY- PIPERIDINE(ALKYL)-CARBOXYLIC ACIDS PROCESS, RH CATALYST + NH3 AVOIDS DECARBOXYLATION
BR 0969197 ICI-- 30AP62 -VY- PIPERIDINE, PROCESS FROM PENTAMETHYLENE GLYCOL USING NICKEL/COBALT ON KIESULGUHR CATALYST
BG 0662762 ROHM- 18AU64 -VH- PIPERIDONE/PYRROLIDINONE, 1-CHLORO+/METHYL-PHENYL-3+/5-METHYL-2-ONE (=CYCLIC*ANILIDE)
FR 1364179 JANSN 17MY62 -VX- PIPERIDYL-PIPERIDINE-2,6-DIONES BY CYCLISATION OF CYANOETHYL-PIPERIDYLMETHYLCYANIDES
US 3310561 ALDRI 05AU65 -VX- PURIDINE, 6-TRI(ALK(EN)YL/CYCLOALKYL)AMMONIUM SALTS, CHELATING AGENTS FOR COBALT
JP 8775/64 TAKED 03NV61 -VH- PURINE 4-AMINO * VITAMIN B1, IMPROVES FRUIT YIELD
BR 1027756 SHELL 17JL63 -VHG PURINE, ADENINE DERIVS, ESP, 6-BENZYL AMINO-9(TETRAHYDROPYRAN-2-YL)-9H, KINETIN ANALOGS
BR 1013520 WELCM 16JL63 -VHX PURINE, BENZYLADENINE, FROM ADENINE + BENZALDEHYDE + FORMIC ACID/METAL FORMATE
```

all phosphorus insecticides (P + I), (2) all other patents on phosphorus compounds (P + *not*-I), and (3) all insecticides other than chloro-hydrocarbon or phosphorus types (*not*-L or *not*-P + I). In the course of a few months, the full set of standard runs is produced, and the cycle starts again to produce an updated set. Lists based on patentees may also be produced.

The format of these lists, consisting of standardized one-line entries in alphabetical order of keywords, makes them immensely scannable, and a good portion of the enquiries, for the period for which they are available, can be answered by on-the-spot inspection, instead of by searching the punched card file. Moreover, their comprehensive nature makes them valuable as surveys of a field, and, being essentially in plain language, they can be studied by interested research personnel. The lists do not permit searching in depth, since only the one keyword at the start of each free-text part of an entry is searchable, and multiple entries are prepared in only a minority of cases. The punched-card fragment code is so economical, however, as the means of storing this detailed information that elaboration does not seem justifiable. The two systems complement each other admirably.

As a further development, the fragment coding could be added to the computer entries, so that either or both forms of record could be searched by computer and a printout obtained as a record of each search. In this case, superimposed coding could be eliminated and false drops reduced. At present, however, remoteness from the computer still leaves some advantage with the punched card system for searching which we can always have immediate access. This situation is likely to change with development of computer facilities.

LITERATURE CITED

(1) Thomson, L. H., E. Hyde, and F. W. Matthews, J. CHEM. Doc. 7, 204 (1967).

(2) "Survey of European Non-conventional Chemical Notation Systems," National Academy of Sciences, National Research Council, Publication 1278, Washington, D. C., 1965.

(3) Silk, J. A., J. CHEM. Doc. 3, 189 (1963).