AN INDEX OF PATENTS PERTAINING TO ORGANIC CHEMICALS AS NEW COMPOSITIONS OF MATTER*

By GERALDINE L. KRUEGER, PAUL L. TIERNAN, MARY B. ABBETT

The Department of Organic Chemistry, Research Laboratories, The Wm. S. Merrell Company, Division of Richardson-Merrell, Inc.

An IBM punched card index to certain United States chemical patents disclosing organic compounds as new compositions of matter has been developed in our laboratories. Such an index was required because our routine searches through the indexes of Beilstein and Chemical Abstracts for particular compounds, their next higher and lower homologs and pertinent isomers failed to cover the most recent chemical literature. Especially we found that available chemical patents indexes did not give a broad view of disclosed structures closely related to the compounds we were searching.

Several years ago, members of the Organic Research Department designed an organic structure code for punched cards which is now in use in our Records Office. This code has been described previously in detail. [1]

Since it was constructed to permit searching for all the members of a structurally related group of compounds, rather than for a single compound, the code seemed well suited to a patent index. In January, 1957, coding of chemical patents was initiated and the index has continued to the present.

As our work progressed we continued to study other patent indexes to gain valuable ideas for our system and to avoid duplication of effort. Such indexes included those to Australian patents,[2] to steroids patented in the United States,[3] and to United States chemical patents.[4] Both the Punched Card System for Searching Steroid Compounds and the Uniterm Index are available to us and are used to extend and verify our searches.

In the preparation of the index, patents are first selected for coding from the Official Gazette. It is difficult to give any hard and fast rules about the selection of patents; they are usually in U. S. Patent Office Classes 167 and 260 and contain one or more claims to organic compounds as new compositions of matter. There is an exception to this rule: when the claim covers the process of treating a disease with a known compound or group of compounds, the patent is coded. Patents disclosing dyes, polymers, special pharmaceutical formulations and processes, except the therapeutic processes just described, are not coded. Patents disclosing new steroid compounds are only partially coded with respect to structure because the index prepared by the Patent Office provides a more exact method of searching steroids. In practice, the number of patents chosen for coding averages about 20-30 per week.

It is important to note that each index card is prepared by scanning the entire patent, which is received and read on Microcard.[5] In addition to the chemical structure of disclosed compounds, the patent number, U. S. Patent Office classification, disclosed therapeutic usefulness and patent assignee are coded. This information is arranged on the card as shown in the accompanying guide (Table I). The patent number is coded in the first seven columns, and in column 8 a letter code is used to designate the country in which the patent is published. This arrangement permits extension of the index to foreign patents and applications. These first eight columns are interpreted and machine-printed in the upper left hand column of the final punched card. Chemical structure is coded in the next section of the card, columns 10-48; a modification of the original code in columns 10-12 permits generic coding of specific antibiotics, their salts and other derivatives as well as terpenes and sugars. In column 18, the presence of non-ionizable halogen is indicated. Columns 19 through 48 are used as they were first described in the Records Office code. In columns 50-58, the Patent Office class and sub-class is coded numerically.

TABLE I GUIDE TO PATENT INDEX CARD

Column	Information
1–7	Patent number
8	Country in which patent is published
9	(Unassigned)
10–48	Structure
10–12	Generic coding (Antibiotics, Sugars, etc.)
13–17	(Unassigned)
18	Non-ionizable halogen
19	Elements other than C,H,N,O,S, halogen
20–22	Number of rings
23	Alicyclic ring size
24–32	Heterocyclic rings, type and size
33–34	Alkylene, terminal chains
35–46	Functional groups
47–48	Miscellaneous structural features
49	(Unassigned)
50–52	Patent Office Class
53–58	Patent Office Subclass
59–70	Therapeutic use
71–72	(Unassigned)
73–75	Assignee
76–78	(Unassigned)
79	Patent type
80	Card Deck Code

The method of coding information in columns 10 to 48 has been described previously.[1]

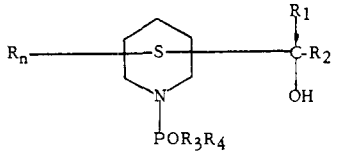
We have attempted to classify therapeutic use in columns 59-70. The most frequently encountered assignees are identified by numbers from 1 to 217 in columns 73-75. All other assignees are coded as 999. Column 79 can be used to sort for all Richardson-Merrell patents, and in column 80 a punch in row 2 identifies the patent deck in the Records Office.

*Presented before the Division of Chemical Literature, American Chemical Society, Chicago, Illinois, Sept., 1958.

From this general survey of the index, we turn to the actual coding operation in terms of a particular patent. The coder uses a 5 x 8 card marked off in 80 blocks with space at the top for indicating the generic formula of the disclosed compound. All variations in this broadly disclosed structure are coded on a single card. To illustrate we will code the chemical and therapeutic disclosures in U. S. 2,832,786 assigned to The Wm. S. Merrell Company (Table II). The patent discloses the

TABLE II EXCERPT FROM U.S. PATENT 2,832,786

The new compounds can be represented by the generic formula



in which S signifies that the ring is saturated; n is an integer from 1 to 4 inclusive; R is selected from the group consisting of the hydrogen atom and the methyl and ethyl radicals; R_1 and R_2 are selected from the group consisting of the methyl, ethyl, phenyl, chlorophenyl, tolyl, methoxyphenyl, benzyl, phenetyl, and cyclohexyl radicals; and R_3 and R_4 are lower alkoxy radicals having from 1 to 8 carbon atoms.

2-, 3- and 4-position isomers of this piperidine carbinol, but our coding does not differentiate position isomers. In Column 18 of the patent index code the presence of non-ionizable bromo-, chloro-, fluoro- and iodo- groups may be indicated in rows 0, 1, 2 and 3, respectively. In this instance, the code is 18/1. Column 19: The presence of phosphorus is indicated at 19/0.[6] Columns 20-22: There may be either one or two non-fused alicyclic rings 20/y,x and one or two non-fused aromatic rings, 21/y,x. There is one non-fused heterocyclic ring, 22/y. Column 23: The possible alicyclic ring is six-membered, 23/3. Column 26: The heterocyclic ring contains one nitrogen, is non-fused and saturated, 26/4. Columns 33-34: There are several possible alkylene chains, one carbon atom in length, 33/1, and terminal chains from one to eight carbons in length, including all the alkyl and alkoxy groups disclosed, 34/1,2,3,4,5,6. Column 36: The hydroxy group bonded to a tertiary carbon, 36/3, and the ether groups, 36/6, are coded here. Column 38: The tertiary nitrogen function, part of a ring, is coded at 38/y3. Columns 45-46: The chloro group bonded to a benzene ring is coded 45/4. The phosphorus functions are coded 46/y,1,3,7.

This completes the structure coding. Finally the therapeutic disclosure is coded. Here the patent reads:

> This invention relates to new chemical compounds which are useful as depressants for the central nervous system....the usefulness of these compounds in many cases, however, lies in the variations in mood which they produce rather than in measurable depression, e.g., some of the compounds cause dogs to become very amenable to command and people less bothered by worries and external annoyances.

Our most difficult problem in designing the patent index has been encountered at this point; the terms used to describe therapeutic utility sometimes seem to defy classification. Columns 59-70 (Table III) represent our attempt to set up categories for the therapeutic

TABLE III COLUMNS 59–70. THERAPEUTIC ACTIVITY

	59 CNS effects	60 PNS effects	61 Cardiovasculars
Y			
X			
0	Sedative	Stimulant	Cardiac stimulant
1	Hypnotic	Spasmolytic	Antifibrillatory
2	Ataractic		Coronary dilator
3		Adrenergic block	Rate effects
4	Analgesic	Ganglionic block	Vasodilator
5	Antipyretic	Local anesthetic	Hypotensive
6	General anesth.	Muscle relaxant	Vasoconstrictor
7	Anticonvulsant		Anti-athero.
8	Stimulant (analept)	Diuretic	
9	Other	Other	Other

	62 Hemic Agents	63 Gastrointestinal	64 Antiinflammatory
Y			
X			
0	Anti-anemic	Antacid	Topical
1	Anticoagulant	Acidifier	Systemic
2	Hemostatic	Digestant	Cortisone-like
3	Plasma modifier	Choleretic	
4	WBC modifier	Laxative	
5		Anti-irritant	
6		Anti-ulcer	
7		Smooth muscle relaxant	
8			
9	Other	Other	Other

	65 Hormonals	66 Antihormonals	67 Chemotherapy
Y			
X			
0	Estrogen	Anti-estrogenic	Anti Gram +
1	Androgen	Anti-androgenic	Anti Gram -
2	Progestational	Antiprogest.	Antirickettsial
3	Glucocorticoid	Antiglucocort.	Antiviral
4	Mineralocorticoid	Antimineralocort.	Antiprotozoal
5	Antihypothyroid	Antipituitary	Antifungal
6	Antihyperthyroid		Anti-Tb, -leprotic
7	Pituitary		Anthelmintic
8			Anti-cancer
9	Other	Other	Other

	68 Nutritionals	69 Antagonists (Misc.)	70 Miscellaneous
Y			
X			
0	Protein anabolic	Antihistamine	Diagnostic aid
1	Growth inhibitor	Antiserotonin	Oxytocic
2	Mineral	Anticholinesterase	Contraceptive
3	Vitamin	Anti-enzyme	Expectorant
4	Lipotropic	Antinausea, -emesis	Antitussive
5	Injectable	Anti-appetite	Proteolytic, muco-
6	Hypoglycemic	Vs. Motion sickness	Counterirritant
7	Hyperglycemic		
8			
9	Other	Other	Other

activities most frequently encountered in patents; this section of the code is certainly

subject to revision. For the patent we are now discussing, the best approximation is 59/2 and the cautious coder will indicate 59/9 as well.

Before leaving this patent, let us note that to search the deck for this type of structural disclosure, not more than two sorts are required: first 46/7, for the P to N bond, then the selected cards are sorted for 46/y, the ring punch which selects those compounds with either phosphorus or nitrogen as a hetero-ring atom. We carried out this search, using the Records Office IBM Type 082 sorter on our present deck of about 2000 cards and the first sort yielded 17 cards. Since, in the process of checking the punched cards when they return from the key-punch operator in the Accounting Department, the coders draw the generic formula on the reverse of the punched card, it was a simple matter to hand-sort the 17 cards for the correct patent.

Our coding team consists of two chemists who spend a total of three days a week on the patent index. In this time, they perform the functions of coding, checking, verifying the punched card, drawing structures on the reverse side of the card, abstracting important patents for Scientific Division members and assisting in preparing write-ups of machine searches. During the first year of operation, they indexed over 1300 patents. In January, 1958, the index was extended to cover certain patent applications from Australia and Belgium which we select from the Derwent Belgian Report and Fine Chemicals Patent Journal and receive in photo-copy or micro-film form.

By the methods we have described, selected United States patents have been indexed and are available for machine searching within a month to six weeks of the date of issue. In our experience, this type of index is of inestimable help in extending our coverage of organic chemical literature to current patents.

REFERENCES

[1] K. W. Wheeler, et al., "A Structure Code for Organic Compounds," American Documentation, 9, 198-207 (1958).

[2] Commonwealth of Australia, Patent Office; private communication.

[3] J. Frome and J. Leibowitz, "A Punched Card System for Searching Steroid Compounds," Patent Office Research and Development Reports, No. 7, July 8, 1957.

[4] Uniterm Index to Chemical Patents, Information for Industry, Inc., Washington, D. C.

[5] U. S. Chemical Patents on Microcards, Microcard Foundation, Ann Arbor, Mich.

[6] The reader who wishes an exact description of the structure code must consult to reference 2 for information contained in columns 19-48, omitted in the present paper to avoid duplication of previously published material.