Some Observations Concerning *Chemical Abstracts* Formula Indexes*

JAMES H. SCHWARTZ†
Technical Information Services, Olin Mathieson
Chemical Corp., New Haven, Conn. 06504

Received February 13, 1969

A small study was made to determine the certainty of locating organic compounds by use of the molecular formula indexes of *Chemical Abstracts*. A random sample of 12 English language references (1953 to 1959) to 37 compounds was selected. The tabulated results show a 92% retrieval for the journal literature, but a 33% retrieval for the patent literature.

Although the formats and advantages of use of different kinds of formula indexes have been discussed[2, 3, 4, 7, 8], only one appraisal has been made regarding the completeness of formula indexes. Watson[9] showed, in his search for references to cycloaliphatic amine derivatives, an average of 30% additional references per compound retrieved from indexes of *Chemisches Zentralblatt (CZ)* which were not obtained from a search of *Chemical Abstracts (CA)*. By using author and patent number indexes, Watson later found that 75% of the references retrieved from *CZ* but not obtained from *CA* were abstracted in *CA*. Watson believed the omissions were due to errors in human judgment applied to a basic set of rules for including chemical compounds in *CA* formula indexes.

Although Mellon[5] reported that procedures for naming compounds and finding names of compounds in *CA* indexes have been published, no basic set of instructions used in the selection of material for inclusion in the formula index has been made known to users of *CA*. Introductions to *CA* formula indexes, however, contain the statement that all new compounds and all compounds for which new information is given are indexed.

In a general discussion of *CA*'s philosophy of subject indexing, Bernier and Crane[1] stated that *CA* takes novelty and significance of subject matter as a basis for selection of information for indexing. Common solvents or reagents and well known starting materials are indexed only if they constitute the theme of the paper. The authors admitted that these compounds should often be indexed, but felt that they would clutter a subject index if considered.

This limited study was made to establish the effect of *CA*'s policy for selecting compounds for inclusion in formula indexes.

METHOD

As shown in Table I, 37 compounds used as starting materials (S.M.) or intermediates (I) were selected from 12 English language original references (seven journal articles and five U.S. patents) published between 1953 and 1959, a period frequently encountered in retrospective literature searches. The journal articles were randomly chosen from *J. Org. Chem.* (examples 4 and 9), *J. Am. Chem. Soc.* (examples 5 and 10), *J. Chem. Soc.* (example 8), and *Tetrahedron* (examples 11 and 12).

The selection of U.S. patents was limited to patents relating to the synthesis of organic compounds. The patents chosen were U.S. Patents **2,838,153** (example one), **2,791,599** (example two), **2,894,028** (example three), **2,671,085** (example six), and **2,802,769** (example seven).

In order to learn whether or not the 37 compounds were included in the *CA* formula indexes, the location in *CA* of the selected references containing each of the compounds was determined by using *CA* author indexes.

RESULTS

By matching the molecular formulas of the compounds with the corresponding *CA* entries for the selected references, 27 of the 37 compounds were located in the *CA* formula indexes (Table II). Of the 25 from journal articles, 23 had entries which correlate with the *CA* reference to the journal article. Of the 12 from patents, only four had formula index entries which match the *CA* reference to the patent.

Although the sample used in this paper is very small and statistically insignificant, the poor retrieval of compounds from the selected patents is surprising.

Table I indicates that starting materials and intermediates are not indexed in examples one, two, and seven. In example three (U.S. Patent **2,894,028**), the selected compounds, both starting materials and intermediates, are indexed. The starting material is not indexed in example six (U.S. Patent **2,671,085**), but the intermediates are. It is difficult to understand this irregular pattern. In a paper on the documentation of patents by *Chemical Abstracts*, Platau[6] described patent acquisition, selection, duplication, and abstracting, but did not report details of indexing procedure. His only mention of indexing was that all claimed compounds are identified in the indexes by the designation "PC" in order to show that these entries represent claims.

* Presented at the 4th Middle Atlantic Regional Meeting, ACS, Washington, D. C., February 12-15, 1969.

† Present address: Celanese Research Co., Box 1000, Summit, N. J. 07901.

Table I. Results—Selected References and Compounds

	Example No., Reference, and *CA* Reference	Compounds and Molecular Formulas	Starting Material (S.M.) or Intermediate (I.)	Found in Formula Index
1.	U. S. Patent **2,838,513** *CA* **52**, 13807*f* (1958)	3,6-Dichloropyridazine $C_4H_2Cl_2N_2$	I.	No
		3,6-Pyridazinedithiol $C_4H_4N_2S_2$	S.M.	No
2.	U. S. Patent **2,791,599** *CA* **51**, 13307*b* (1957)	*S*-(Ethylthio)methyl *O,O*-dimethylphosphorothioic acid, *O,O*-dimethyl ester $C_5H_{13}O_3PS_2$	S.M.	No
		Diethyl(ethylthio)-phosphorothioic acid, ethyl ester $C_8H_{19}O_3PS_2$	S.M.	No
3.	U. S. Patent **2,894,028** *CA* **53**, 19924*g* (1959)	*N*-Chlorocyclohexanimine $C_6H_{10}ClN$	S.M. and I.	Yes
		Cyclohexanone hydrazone $C_6H_{12}N_2$	S.M. and I.	Yes
4.	*J. Org. Chem.* **22**, 1076 (1957) *CA* **52**, 10886*a* (1958)	Methyl *n*-propyl ketone cyanohydrin $C_6H_{11}NO$	S.M.	Yes
		Methyl *n*-hexyl ketone cyanohydrin $C_9H_{17}NO$	S.M.	Yes
		α-Acetoxy-α-methylvaleronitrile $C_8H_{13}NO_2$	I.	Yes
		α-Acetoxy-α-methylcaprylonitrile $C_{11}H_{19}NO_2$	I.	Yes
5.	*J. Am. Chem. Soc.* **77**, 2256 (1955) *CA* **50**, 2604*g* (1956)	2,4-Dichloropyrido(2,3-*d*)-pyrimidine, $C_7H_3Cl_2N_3$	I.	Yes
		2-Chloro-4-hydroxypyrido-(2,3-*d*)pyrimidine, $C_7H_4ClN_3O$	I.	Yes
		2,4-Dihydroxypyrido(2,3-*d*)-pyrimidine, $C_7H_5N_3O_2$	I.	Yes
6.	U. S. Patent **2,671,085** *CA* **49**, 3271*g* (1955)	4-Chlorobenzonitrile, C_7H_4ClN	S.M.	No
		4-Chlorobenzamidine, $C_7H_7ClN_2$	I.	Yes
		4-Chlorobenzimidic acid, ethyl ester, $C_9H_{10}ClNO$	I.	Yes
7.	U. S. Patent **2,802,769** *CA* **52**, 6708*i* (1958)	4-Chlorobenzyl chloride $C_7H_6Cl_2$	S.M.	No
		4-Chlorobenzyl mercaptan C_7H_7ClS	I.	No
		4-Chlorobenzyl 2-chloro-ethyl sulfide, $C_9H_{10}Cl_2S$	I.	No
8.	*J. Chem. Soc.* **1958**, p. 1750–4 *CA* **52**, 13725*g* (1958)	1,2,3,4-Tetrahydro-1-oxoquinolizinium bromide $C_9H_{10}BrNO$	I.	Yes
		1,2,3,4-Tetrahydro-1-oxobenzo(b)quinolizinium bromide $C_{13}H_{13}BrNO$	I.	Yes
		3-Ethoxybutyrylisoquinoline $C_{15}H_{17}NO_2$	S.M.	Yes
		2-Ethoxybutyrylpyridine $C_{11}H_{15}NO_2$	S.M.	No
9.	*J. Org. Chem.* **24**, 888–91 (1959) *CA* **54**, 348*e* (1960)	Ethyl isopropylidene malonate $C_{10}H_{16}O_4$	S.M.	Yes
		4-Phenyl-3,3-dimethyl-butanoic acid, $C_{12}H_{16}O_2$	I.	Yes
		4-Bromo-1-phenyl-2,2-dimethylbutane, $C_{12}H_{17}Br$	I.	Yes
		4-Phenyl-3,3-dimethyl-1-butanol, $C_{12}H_{18}O$	I.	Yes
		Ethyl 4-phenyl-3,3-dimethyl-2-carbethoxybutanoate $C_{17}H_{24}O_4$	I.	Yes
10.	*J. Am. Chem. Soc.* **75**, 1889–91 (1953) *CA* **48**, 4490*a* (1954)	Ethyl (3,4,5-trimethoxyphenyl) propionate, $C_{14}H_{20}O_5$	I.	Yes
		Ethyl formyl(3,4,5-trimethoxyphenyl)propionate $C_{15}H_{20}O_6$	I.	Yes

(Continued)

Table I. *Continued*

	Example No., Reference, and *CA* Reference	Compounds and Molecular Formulas	Starting Material (S.M.) or Intermediate (I.)	Found in Formula Index
		Ethyl 3,4,5-trimethoxy-benzoylacetate, $C_{14}H_{18}O_6$	S.M.	No
11.	*Tetrahedron* 7, 130–7 (1959) *CA* **54**, 4681*c* (1960)	16 α-Cyano-3β-hydroxy-5-pregnen-20-one $C_{22}H_{31}NO_2$	I.	Yes
		16 α-Cyano-5-pregnene-3β,20β-diol $C_{22}H_{33}NO_2$	I.	Yes
		16β-Carboxy-5-pregnene-3β,20β-diol 16,20-lactone $C_{22}H_{32}O_3$	I.	Yes
		3β-Acetoxy-20β-hydroxy-5-pregnen-16β-carboxylic acid 16,20-lactone, $C_{24}H_{34}O_4$	I.	Yes
12.	*Tetrahedron* 5, 15–26 (1959) *CA* **53**, 14144*g* (1959)	Δ^4-Pregnene-3β,20β,21-triol-21-monoacetate $C_{23}H_{36}O_4$	I.	Yes
		Δ^4-Pregnene-3β, 21-diol-20-one diacetate, $C_{25}H_{36}O_5$	I.	Yes

Table II. Number of Selected Starting Materials and Intermediates

	From Original Journal Articles	From Original U. S. Patents	Total
Number of compounds	25	12	37
Number of starting materials	6	7	13
Number of intermediates	19	5	24
Number of starting materials found in formula index	4	2	6
Number of intermediates found in formula index	19	2	21

FUTURE PLANS

An attempt will be made to obtain financial support for a more definitive and statistically significant study of formula indexing procedures, especially for the years 1962 to 1968.

LITERATURE CITED

(1) Bernier, C. L., and E. J. Crane, *Ind. Eng. Chem.* **40**, 725–30 (1948).
(2) Dyson, G. M., *Chem. Ind.* **1952**, 676–84.
(3) Fletcher, J. H., and D. S. Dubbs, *Chem. Eng. News* **34**, 5888–91 (1956).
(4) Garfield, E., J. CHEM. Doc. **3**, 97–103 (1963).
(5) Mellon, M. G., "Chemical Publications, Their Nature and Use," p. 135 McGraw-Hill, New York, 1965.
(6) Platau, G. O., J. CHEM. Doc. **7**, 250–5 (1967).
(7) Schmerling, L. S., J. CHEM. Doc. **1** (1), 45–51 (1961).
(8) Skolnik, H., and J. K. Hopkins, *J. Chem. Educ.* **35** (3), 150–2 (1958).
(9) Watson, C. E., Division of Chemical Literature, 152nd Meeting, ACS, New York, N. Y., September 1966.

Literature Retrieval in the Aluminum Industry*

RALPH A. MAGOWAN
Olin Mathieson Chemical Corp., New Haven, Conn. 06504

Received January 30, 1969

A review of the major sources of information pertinent to the aluminum industry, with emphasis on *Chemical Abstracts* and the *ASM Review of Metal Literature*. An evaluation is made of these sources, including coverage and indexing.

The objective of this paper is to serve as a guide to the literature of the aluminum industry.

The sources evaluated include *Chemical Abstracts, ASM Review of Metal Literature, Official Gazette of the U.S. Patent Office, Uniterm Index of Chemical Patents, Engineering Index, Electrical Engineering Abstracts, Metals Abstracts, Aluminum Technical Information Service,* and *Aluminum Abstracts.* Most of the paper is devoted to the first two, as they are the major sources for a retrospective search. Some mention should be made here of the U.S. Government sources, but these are in turn searched and abstracted by both *Chemical Abstracts* and the American Society for Metals; one exception is *TAB (Technical Abstract Bulletin)*, which is a restricted document, but can still be searched "in-house." NASA's *STAR* contains such a wealth of information in the areas of cryogenics and welding that it is often worth using directly.

* Presented before the Division of Chemical Literature, 156th Meeting, ACS, Atlantic City, N. J., September 13, 1968.