

Time Studies in Producing Subject Indexes for *Chemical Abstracts**

F. R. WETSEL

Chemical Abstracts Service, Ohio State University, Columbus, Ohio 43210

Received July 21, 1965

The purpose of this paper is to present a breakdown of the time, professional and clerical, that goes into the preparation of *Chemical Abstracts*' (CA) Subject and Formula Indexes. The figures presented in this paper are for Volume 60 (January–June 1964) of CA, the last index period for which complete information was available. This volume contained 78,800 journal abstracts and 14,200 patent abstracts, for a total of 93,000 abstracts. The time reported for each operation is in terms of productive hours and not total hours. This means that time for sick leave, vacation, administration, special internal file input, and maintenance, etc., is not included in the hours reported for each operation.

Much of CA indexing is done not from the abstract alone but with the aid of the original article or patent. In the indexing of chemical compounds, the primary papers are most often consulted. Original articles and patents corresponding to almost all other abstracts are available, subject to the indexers' needs to give in-depth indexing when the printed abstracts lack sufficient detail. Library, secretarial, and clerical services, which organize material prior to indexing, perform their operations in such a way that index support is not normally identified separately from the support for abstracting, abstract editing, etc. Therefore, this type of clerical support for producing indexes is not included in the figures given. Table I summarizes the time spent in each indexing operation, and the following text describes these operations in the order given in the table.

The first indexing operation is the drawing of structural diagrams by B.S. degree chemists. All compounds which, according to the author, are reported for the first time or have new associated data are indexed regardless of whether they are specifically mentioned in the abstract. Many of the structures are diagrammed to aid in the determination of molecular formulas and to aid in the later assignment of index names according to CAS nomenclature rules. For Volume 60, structures were drawn for 153,300 compounds. The 21,700 hours used for this operation include the time required to review the original paper or patent along with the abstract to select the compounds which are to be included in the index.

Additional structures or partial structures may also be prepared in later indexing operations as part of the naming of compounds. The time required for preparing these additional structures by the subject indexers is included in the indexing hours. Therefore, the number of

structures drawn per six-month volume in the indexing process is closer to 200,000 than 150,000. Note that these 200,000 structures do include multiple entries for a single compound. Therefore, the net number of different compounds structured is less than 200,000. CA indexes cite an even higher number of compounds than those represented by the 200,000 structures when account is taken of the inorganic and simple organic compounds which do not require structuring in the derivation of their molecular formulas and index names.

After structuring, the next operation is preparing the index entries by chemists who vary in background from people with B.S. degrees coupled with work experience in speciality areas to chemists with Ph.D. degrees, who may also have previous work experience. An index heading is a word or group of words that indicates a concept under which all material relating to that same concept is indexed. is a word or group of words that indicate a concept under which all material relating to that same concept is indexed. In Figure 1, subconcepts, often termed modifications, are indented in light-face type under the bold face **Diphenic acid** heading. Diphenic acid reads directly into each of these modifications. A modification may be a reference to more than one entry, as in the case of the "diester with 4'-bromo-2-hydroxyacetophenone" modification under the **Diphenic acid** heading in Figure 1.

The initial preparation of 580,700 Subject and 175,000 Formula Index entries for the 93,000 abstracts took 33,700 technical hours. Often discussions of indexing address themselves only to this point of initial indexing and not the total index production. However, it is important to note that the total technical staff time for the complete production of the volume Formula and Subject Indexes is 80,900 hours—more than twice the amount of technical time required for the initial selection of index entries. This is a major point in this presentation.

Indexers dictate the 755,700 entries and transcribers type the individual index entries on 3 × 5 manuscript cards. The cards, which serve as the index manuscript, are checked and later arranged and edited. The indexer on the average produces an entry in less than three minutes, with the typist being responsible for the completion of the index cards according to a fully detailed, predescribed format that relieves the indexer of a number of routine operations.

After typing, indexers or index editors give a quality control check of both indexing and typing to selected samples of index cards. For indexers in training, *all* cards are technically checked immediately by a chemist on the indexing staff. This provides a quality control check as well as an aid to training. Time for corrections is included in the total technical hours for indexing and editing.

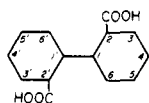
* Presented before the Division of Chemical Literature, Symposium on Work and Time Studies in Technical Information, 149th National Meeting of the American Chemical Society, Detroit, Mich., April 1965.

TIME STUDIES FOR SUBJECT INDEXES OF CA

Table I. Subject Indexes Time Studies of *Chemical Abstracts* Volume 60

Operation	Quantity	Hours (rate)	
		Clerical	Professional
Journal and patent prepn.	93,000 abstracts		
Structuring	153,300 compounds		21,700 (7 compd./hr.)
Indexing	580,700 subj. entr. 175,000 form. entr.		33,700 (22 entries/hr.)
Typing	755,700 entries	11,700 (64 entries/hr.)	
Checking	755,700 entries		9,700 (78 entries/hr.)
Ordering and interfiling	755,700 entries	12,200 (62 entries/hr.)	
Editing	755,700 entries		15,200 (50 entries/hr.)
Composition	1,247,000 lines	24,900 (50 lines/hr.)	
Proof	1,247,000 lines	5,900 (214 lines/hr.)	5,600 (223 lines/hr.)
Graphic arts	Subj., 2812 text pp. (3-col.) Form. 1003 text pp. (2-col.)	1,200	
Total hours		55,900	85,900

Diphenhydramine. See *Ethylamine, 2-(diphenylmethoxy)-N,N-dimethyl-*
Diphenic acid (2,2'-biphenyldicarboxylic acid)



bis(2,2,3,3,4,4,5,5-octafluoropentyl) ester, 60:344e
cobalt complex, redn. of, 60:11865a
compd. with 2-benzyl-2-thiopseudourea, spectrum of, 60:12787d
compd. with quinine, optical rotatory dispersion of, 60:5383h
derivs., asymmetric isomerization by, 60:5383g
detn. of, 60:11383a
diester with 4'-bromo-2-hydroxyacetophenone, 60:5387d, 12787d
formation in oxidn. of phenanthrene, 60:13205c
manuf. from phenanthrene ozonization, 60:P 10606e
manuf. of, 60:P 482d
oxidn. with H₂O₂, 2'-hydroxy-2-biphenyldicarboxylic acid 5-lactone by, 60:P 1657c
prepn. of, 60:13176h
thermal analysis and thermogravimetry of, 60:4803e
Diphenic acid, 3,3'-dimethyl-, 60:5383h
compd. with benzene and quinine, 60:5384h
with brucine, 60:5384h
with quinine, 60:5384a
—, **3,3'-dimethyl-4,4'-dinitro-**, 60:5383h
compd. with quinine, 60:5384a
—, **4,4'-dinitro-**
compd. with quinine, optical rotatory dispersion of, 60:5383g
—, **dodecahydro-**. See [Bicyclohexyl]-2,2'-dicarboxylic acid
—, **5-[(O-glucosylglucosyl)oxy]-4,4',5',6',6'-pentahydroxy-**
di-5-lactone*, in guava, 60:4458h
—, **4,4',5,5',6,6'-hexahydroxy-**
di-5-lactone*, in *Agrimonia eupatoria* rhizomes, 60:5591f
in *Eucalyptus*, 60:9596d
in *Eucalyptus* spp., reaction with Fe, 60:2049b
from *Lithospermum officinale* fruit, 60:11042c
Diphenicillin*
Staphylococcus aureus inactivation of, 60:3302g
Diphenimide (2,2'-biphenyldicarbamide)
—, **N-[2-(diethylamino)ethyl]-**, 60:449b
Diphenolic acid. See *Valeric acid, 4,4-bis(p-*

As soon as the index cards become available, the clerical staff starts to check and organize the entries. This job is completed shortly after the indexers finish their indexing. At this time the index entries are very nearly in the order of the final printed index. Changes in order from this point onward in the operations dealing with 3 × 5 manuscript cards result from editing by the professional staff.

In filing over 750,000 index entries, large numbers of very closely related entries are prepared. This complicates the clerical ordering procedure. An average rate of approximately a minute per entry is required to bring index entries into the final index order. The crossing off of duplicate heading and modification information is also done by the clerical staff and is included in this time figure.

The technical editors, with backgrounds similar to those of indexers, check the interrelationships of related index entries and the nature and the wording of each entry. Nomenclature and structural details are reviewed during the editing of the entries for compounds. Each index heading and its modifications are checked by a chemist as a unit. Technical Subject Index editing operations for Volume 60 required 15,200 hours.

The final printed text in proof form is clerically checked for entry order and proofed letter by letter and number by number against manuscript copy to verify information which cannot be checked in the technical proofreading. This operation required 5,800 hours.

It required 5,600 hours for technical proofreading by chemists, who review only the technical sense of index entries and headings and make no check against the manuscript index cards except in areas which require major changes.

The 1,200 hours in the graphic arts area is that time required for producing the positive paper proof and final

Figure 1. Sample from page of Volume 60 Subject Index.

film negatives that were used in preparing the printing plates. This time also includes the attachment of negatives of two-dimensional structures for the ring systems which are included in the Subject Indexes. CA's Subject Indexes currently contain over 3,000 such structures per volume. (About 1,000 nonring structures also appear in each volume index; these are typeset as line formulas and require no attachment of structure negatives.)

Later index preparation operations, which include film stripping, plate preparation, press runs, collating, binding, and distribution, are performed by the printer, and are therefore not included in this time study.

The Subject Index for Volume 60 was the first Subject Index composed entirely by a line-at-a-time system (1) completely within CAS. Previous Subject Indexes were set in "hot-type" using monotype composers. With this cold-type system, each printed index line is composed on a single, blank, data-processing card. The cards are then photographed one at a time using a sequential card camera. The result is a strip of film which corresponds to a column of the index. Printing plates are prepared from the film. A major advantage of this system is that collective indexes can be prepared by merging the cards used in volume indexes without major recomposition expenses. Prior to Volume 60, Formula Indexes were converted to the line-at-a-time process starting with Volume 53, and Author Indexes were converted starting with Volume 55 (2). The Subject Indexes complete the conversion of all of CA indexes to cold-type.

We expect the changes made in going to line-at-a-time composition to be the first step in putting all index preparation on a computer base. A shift to the computer would allow us to produce the several indexes through a single keyboarding. This means that an entry for a compound would be put into the system only once and be manipulated by the computer to compose the Formula and Subject Indexes. This does not mean that we expect to use a computer directly in CA index production within the next several years; we have not yet found the necessary computer-driven composing equipment. However, we do feel that this step is not far off and we want to be ready

to make the transition for the next collective index period (1967-1971) if equipment becomes available.

The present shift to cold-type procedures has allowed us to change our editing procedures to a continuous, even-flow operation that will make the shift to a computer operation relatively easy.

As noted above, the cold-type method allows for the physical storage and re-use of most printed lines. This is not possible in hot-type procedures. Therefore, we expect that collective indexes covering five years or ten volumes can be produced without the clerical proofchecking since this has already been accomplished in the volume index production. By eliminating the proofchecking, we can combine the editing with the technical proofreading operation. The time for these combined operations will be more than proofreading time but will be considerably less than the editing time. Volume index experience shows that the time required for technical editing is almost six times that of technical proofreading. Collective index operation figures correspond closely to these. With the use of the cold-type method production of the estimated 17 subject and formula volumes of the Seventh Collective Index will be accomplished with approximately the same number of man hours that have been required for the ten-volume Sixth Collective Subject and Formula Indexes.

The initial aim of this paper was to present data for the production of CA Subject and Formula Indexes for Volume 60. Discussions and production statistics allowed the opportunity to discuss briefly index composition changes either started or planned for the future.

LITERATURE CITED

- (1) Becker, J., Hayes, R. M., "Information Storage and Retrieval," John Wiley and Sons, Inc., New York, N. Y., 1963, pp. 80-83.
- (2) Angel, W. D., Wetsel, F. R., in "Automation and Scientific Communication," Short Papers, Part 1, H. P. Luhn, Ed., American Documentation Institute, Washington, D. C., 1963, p. 95.