Punched Card Literature Retrieval System for Gas Chromatography*

By O. F. FOLMER, Jr.
Research and Development Department, Continental Oil Co., Ponca City, Okla.
Received December 13, 1962

Convenient and rapid access to the information contained in the literature is as necessary in gas chromatography as in any analytical technique. For gas chromatography, such information retrieval is complicated, not so much by the large number of references available but by the tremendous rate at which new references appear. As Fig. 1 shows, this rate was approximately exponential for the first nine years. Although this rate of increase has currently leveled off, still many papers are published by authors new to the field; and much information appears in journals in which papers on gas chromatography are not ordinarily expected. Furthermore, as gas chromatography becomes more widely employed as a tool, many references of importance are only briefly mentioned in experimental detail and escape notice in the standard abstract publications. While this makes literature searching difficult for the chromatographer, there is no cause for complaint when a researcher publishes his work in a journal specializing in his particular field. This merely reflects the widespread application of gas chromatography to many different and somewhat unrelated fields.

Since gas chromatography is a complex field, necessarily many papers are quite complex, covering a wide range of different aspects of gas chromatography. Classification of papers as to intent, such as theoretical papers, papers dealing with apparatus, or papers concerned with analytical applications, etc., cannot readily be done.

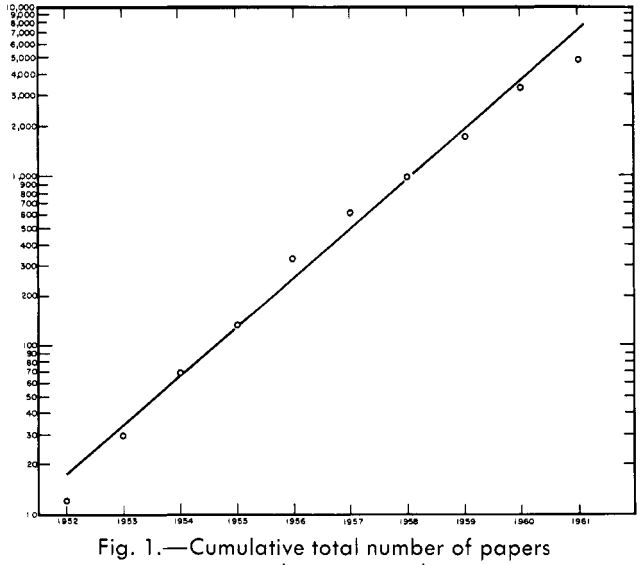

Fig. 1.—Cumulative total number of papers on gas chromatography.

Many papers contain all of these references in varying degree as well as information which does not fit any of these classifications. Good examples of such papers can be found in preprints of the Fourth International Gas Chromatography Symposium.

For the reasons, simple files such as author indexes or more complex searching aids as *Chemical Abstracts* do not provide a sufficiently complete search of the gas chromatographic literature. A systematic method specifically adapted to the requirements of gas chromatography is necessary.

Requirements of any Useful Retrieval System.—The requirements of a method for searching the gas chromatographic literature are:

1. *The search must be complete.* There are very few trivial subjects in gas chromatography. What may be a minor point in one context may be extremely important in another.
2. *The system must be economical of time.* The search must be rapid, and the system should be such that it is not excessively time-consuming to set up or maintain. Such a system gives best results if set up and maintained by a person who is knowledgeable and active in the field. Consequently, maintenance of the system should not be a full-time job.
3. *The system should accommodate a relatively large large number of classes* to enable as detailed a search as is necessary.
4. *The system should not be physically cumbersome nor inordinately expensive.*
5. *The system should be compatible with machine methods of literature searching.*

A number of different methods have been considered in comparison with these requirements. A system using conventional cross-indexed file cards was discarded because of the bulk and the time required for its assembly, use, and maintenance. More than 30,000 cards would be required for the references up to 1962, and this number would increase by 10,000 per year at the present rate of literature increase. Moreover, each card taken from the file would have to be replaced in its original position, or the cards would become disordered and useless.

A hand-sorted, punched-card system most nearly fulfills the requirements listed. An added benefit of hand-sorted punched card systems is the availability of such cards already containing printed abstracts. These abstract cards, which may be obtained from the Preston Technical Abstracts Company, cover approximately ninety per cent of the gas chromatographic literature. Figure 2 shows one of these cards, which is familiar to most chromatographers.

* Presented at the 141st Meeting, American Chemical Society, Division of Analytical Chemistry, Washington, D. C., March, 1962.

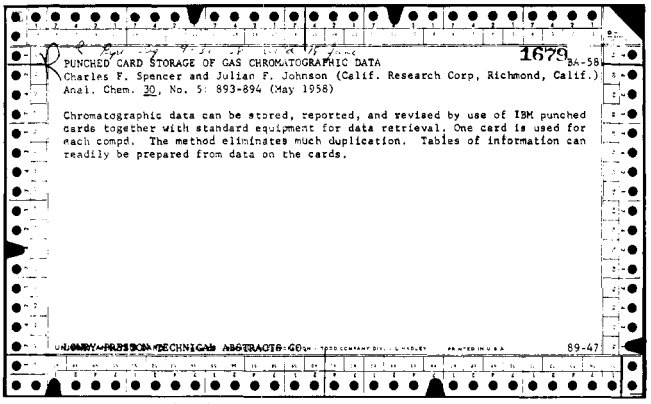
1679 BA-58
PUNCHED CARD STORAGE OF GAS CHROMATOGRAPHIC DATA
Charles F. Spencer and Julian F. Johnson (Calif. Research Corp, Richmond, Calif.)
Anal. Chem. 30, No. 5: 893-894 (May 1958)
Chromatographic data can be stored, reported, and revised by use of IBM punched cards together with standard equipment for data retrieval. One card is used for each compd. The method eliminates much duplication. Tables of information can readily be prepared from data on the cards.
89-47

Fig. 2.—Typical abstract card obtained from the Preston Technical Abstracts Co. (after being numbered and notched by the author).

The simplest method of coding cards would employ a direct code in which each hole represents one class and one class only, and this type of code is recommended by the Preston Technical Abstracts Company. Unfortunately, this is not detailed enough for most purposes. With only 91 holes around the card, only 91 classes can be accommodated. For example, all alcohols are included in one class. If only information about one specific type of alcohol is desired, it is still necessary to scan laboriously the literature on all alcohols since a more detailed subject breakdown is unavailable.

More detailed codes, such as the more commonly used indirect codes in which patterns of holes are used to represent classes, allow only one class to be coded on a card (Robert S. Casey, James W. Perry, Madeline M. Berry, and Allen Kent, "Punched Cards," Second Edition, Reinhold Publishing Corporation, New York, 1958, Chapter 21, "Mathmatical Analyses of Coding Systems" by Carl S. Wise, p. 446). If more than one class is placed on a card, then too many unwanted cards will be produced during a search. Placing only one class on a card again generates the problem of a very bulky file.

Random Number Code Most Useful.—To eliminate these difficulties, a random-number superimposed code was chosen. Such a code allows many classes to be coded on one card and yet produces relatively few unwanted cards during a search ("Punched Cards," pp. 446–457).

The basic theory behind this code is quite simple.[1] To illustrate with a hypothetical example, let ninety of the holes around the card be divided into nine fields of ten holes each. If, in the first field a hole is notched at random, then a needle inserted at random into this field has one chance in ten of being inserted in the notched hole. If a class is represented by a series of nine randomly chosen holes, one in each field, then the chance of a needle being inserted in each field in the notched holes is $1/10 \times 1/10 \times 1/10 \times 1/10 \ldots$ or $(1/10)^9$. If two classes are coded on this card, then the probability of a needle being inserted, in each field, in a notched hole is $(2/10)^9$. For three classes, the probability is $(3/10)^9$. These are the approximate probabilities of receiving unwanted cards when a search is made for some particular class. The actual probability is somewhat smaller than this, since there is overlap of the patterns when more than one class is coded on a card.

According to Wise, the actual fraction of each field notched out is given by the relation

$$\frac{X}{H} = \frac{-\log (1 - f)}{0.434} \qquad \text{(I)}$$

where X is the number of classes per card, H is the number of holes in a field, and f is the fraction of the field actually notched. The dropping fraction F_d, which closely approximates the fraction of unwanted cards obtained in a search, is

$$F_d = f^y \qquad \text{(II)}$$

where y is the number of fields.

Thus, the fraction of unwanted cards produced in a search depends on the number of classes per card and on the manner in which the total number of holes is divided up into fields. Wise shows that this fraction, F_d, will be a minimum when f is 0.37, and consequently (X/H) is 0.46.

A small-scale exploratory sampling of the gas chromatographic literature disclosed that an average of six to seven classes would be necessary to describe each reference adequately. Assuming seven classes per card, and if X/H is to be 0.46, then H is 15.2. This means each field should have 15 holes, and the 91 holes on the card may be divided into six fields of fifteen holes each and one field of one hole. The fraction of unwanted cards calculated with this arrangement is 0.0016.

Classification.—The subject, gas chromatography, has been initially divided into 233 classes. These were arranged in a dictionary in a logical order by subject for easy reference to the necessary code numbers. Each class was assigned a six-digit random number chosen from a table of random numbers. The first digit is in the range 1–15; the second, 16–30; and so on to the sixth which is in the range 76–90.

So far, approximately 4,000 cards have been coded by this method; and 42 complete searches have been made. The average fraction of unwanted cards obtained is 0.0034 ± 0.0028 average deviation. This corresponds to an average of 14 unwanted cards out of the 4,000 cards searched, with an average deviation of 11 cards. From this experimental value f is calculated to be 0.39, and X is 7.4 classes per card. For the last 33 searches, $F_d = 0.0028 \pm 0.0020$ (or 11 unwanted cards ± 8 cards out of 4,000), f is 0.375, and X is 7.05 classes per card.

Thus, it can be seen that experience agrees quite well with theory and that six fields of 15 holes is very close to the optimum arrangement for a gas chromatographic literature code of this type.

Approximately 1% of the references coded requires more than 15 classes per card. If more than 15 classes are coded on a card, then this card will drop out as an unwanted card too frequently. In such cases, a second card is made up, and the classes are divided between the two cards. These duplicate cards are notched in the "extra" hole, number 91, to facilitate correlation when references dealing with two or more classes are sought.

The procedure used in searching is relatively simple. After locating the code number for the class desired, the cards are searched by sorting with a needle for the first

[1] Since there is an excellent theoretical treatment in the reference given, there is no need to repeat it here. Only a brief outline of the theory will be given, and those interested in more detail may consult "Punched Cards," Chapter 21.

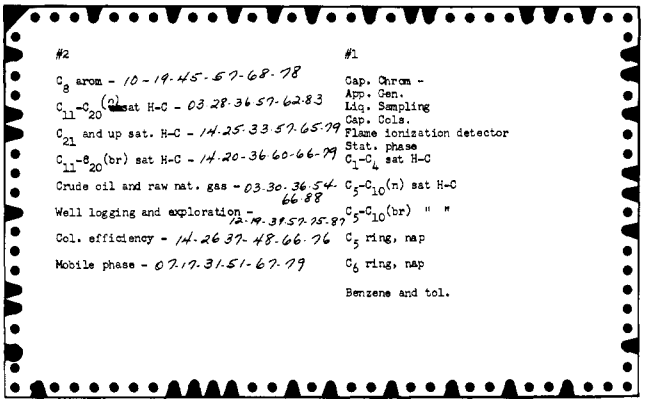

Fig. 3.—Reverse of an abstract card showing the listing of classes and code numbers. Note that this card is the second of a pair of duplicate cards and the classes coded on the first are also listed.

number, and so on until all six numbers have been sorted. The cards obtained are then reversed and their backs scanned so that unwanted cards may be removed. For this reason, as well as to facilitate the notching operation, the classes and code numbers are listed on the back of each card as shown in Fig. 3. In 42 searches, the average total time required both to sort and to remove unwanted cards is 24 minutes, with an average deviation of 6.8 minutes.

In order to prevent loss of the punched cards, they may conveniently be serially numbered and microfilmed in numerical order. With this arrangement, the serial numbers of the cards obtained in a search may be noted; and only this list, not the cards themselves, need be taken from the area of the file. The customer takes the list of numbers and the microfilm to a reader-printer, reads, and makes copies of the cards as he desires.

If further detailed information is needed, recourse can be made to a file of reprints. For the system being described, reprints of about 70% of the reference have been secured and filed by author. In the future it is planned to microfilm the reprints in order to prevent their loss.

Considerable additional information may be had from a supplementary second set of cards coded for author, journal, date, and location. Rather than include this less-needed information in the first code system, it seems easier to set up a second series of cards and make use of a simpler code. The second set of cards may also be coded so as to allow a loss check of the entire system. It is usual to convert to machine operation when the number of cards has increased to such a figure that it is inconvenient and uneconomical to continue with hand searching. The coding system described is quite compatible with the code systems used in machine operation. In fact, it is very similar to several systems already in use with IBM punched cards.

Logograph—Communicating Chemical Procedures*

By KENZO HIRAYAMA
Research Laboratory, Fuji Photo Film Co., Ltd., Minamiashigara, Odawara, Japan

and

AKIRA FUJINO[1]
Faculty of Science, Osaka City University, Sugimotocho, Sumiyoshiku, Osaka, Japan
Received January 3, 1963

1. INTRODUCTION

Hieroglyphs find their source in ancient times but are still used in the modern world. For example, the Chinese character for the sun is pronounced *re* in Chinese, *il* in Korean, and *nichi* or *jits* in Japanese, but represents the sun in each of these languages. This is evidence that the hieroglyphic character is useful as a non-linguistic means of communication, without regard to spoken language.

An attempt to use hieroglyphic symbols for scientific description was made by the physiologist Serge Tchakhotine, but he did not intend to use such symbols as a wordless language. Symbolic descriptions of physiological experiments are given in his book "Organisation rationnelle de la récherche scientifique."[2] He named this system of description "lograph" (lógos, word; gráphos, to write). Tchakhotine used many words other than symbols in his writing, but his idea of using emblematic symbols other than those of a phonetic alphabet for communication is worth a great deal of consideration.

To avoid the complexity of scientific writing in Japanese, the writers and their collaborators[3] have been using such a method of writing, which had been named logograph,[4] for a long time. This is a system of writing formulated by introduction of Tchakhotine's idea of

* Presented at the 142nd National Meeting of the American Chemical Society, Sept. 11, 1962, Atlantic City, N. J.

[1] Chemistry Department, Graduate School, Boston University, Boston 15, Mass.

[2] "Actualités scientifiques et industrielle," Hermann & Cie., Paris, 1938, p. 732.

[3] For the establishment and improvement of this logograph, valuable coöperation has been received from the members of research laboratories of organic chemistry of the Science Faculty, Osaka City University, and of the Research Laboratory, Fuji Photo Film Co., Ltd.

[4] Logograph also means logotype but this seemed a better representation of this idea (a character of sign representing a word) than lograph from the point of word formation.