Henneberg, D., Casper, K., Ziegler, E., Weimann, B., *Angew. Chem., Int. Ed. Engl.* **11**, 347 (1972).
Hites, R. A., Biemann, K., *Anal. Chem.* **39**, 965 (1967).
Hites, R. A., Biemann, K., *Advan. Mass Spectrom.*, 37 (1968).
Knock, B. A., Smith, I. C., Wright, D. E., Ridley, R. G., Kelly, W., *Anal. Chem.* **42**, 1516 (1970).
Robertson, D. H., Cavagnaro, J. F., Holz, J. B., Merritt, C., Jr., Proceedings of the 20th Annual Conference on Mass Spectrometry, and Allied Topics, American Society for Mass Spectrometry, Dallas, Tex., 1972, Paper No. R3.
Smith, D. H., Olsen, R. W., Walls, F. C., Burlingame, A. L., *Anal. Chem.* **43**, 1796 (1971).
Stenhagen, E., Abrahamsson, S., McLafferty, F. W., Ed., "Atlas of Mass Spectral Data," Wiley/Interscience, New York, N. Y., 1969.
Venkataraghavan, R., Klimowski, R. J., McLafferty, F. W., *Accounts Chem. Res.* **3**, 158 (1970).

Received for review December 5, 1973. Accepted April 26, 1974. Presented at the Symposium on Computers in Flavor Chemistry sponsored jointly by the Agricultural and Food Chemistry and Analytical Chemistry Divisions, 166th National Meeting of the American Chemical Society, Chicago, Ill., Aug 1973.

An Information Acquisition and Utilization System for the Flavor Chemist

William G. Galetto* and William H. Stahl

Most workers in the area of flavor chemistry have collected numerous reprints and/or photocopies of journal articles dealing with various techniques and reports of isolation and identification of flavor components found in natural products. The retrieval of information stored in this fashion soon becomes cumbersome and very time consuming. An information retrieval system has been developed to assist in the utilization of this published information. The total system is comprised of a journal reference subsection (file number, author, key word, reference, and year listings) and a chemical data subsection (where found, name, data, Wiswesser Line Notation, number, and functional group listings). These mesh together to permit entrance at a number of different points. The raw data are manipulated *via* the computer, resulting in the preparation of a number of directory-type print-outs which are then used directly by the flavor chemist.

Those currently working in the area of flavor chemistry are well aware of the fact that this is an active area of research. There are many academic, governmental, and industrial laboratories involved in flavor work and publishing results of their investigations.

Keeping up-to-date on the voluminous literature, while necessary, can become very time consuming and thus expensive if steps are not taken to facilitate the task. We have developed in our laboratories a system to manipulate the information found in the open literature and present it in a useful printed format.

The total system (Figure 1) is comprised of two major subsystems or data collections: the Journal reference collection and the chemical data collection. These two subsystems are interconnected such that entry anywhere into one leads directly to any information contained within the entire system. All data are handled *via* low cost programs and in a batch mode, with updating being done two to three times a year. All computer generated reports are in the form of printed directories which can be kept in the laboratories for easy referral or sent to other laboratories if the laboratories are physically separated.

JOURNAL REFERENCE COLLECTION

The journal reference collection subsystem (Figure 2) serves as a means of documenting and retrieving the basic information inherently associated with journal reprints. It is based on a computer program developed by Steed (1971), of Brigham Young University (presently at American River College, Sacramento, Calif.). Data input is *via* punched cards; output is in the form of printed reports whose contents are represented by the five outer circles shown in Figure 2. The interconnecting arrows indicate the cross-referencing that exists between the various print-outs. The listings can be described as follows. D (file number) is a list, in numerical order, of all journal references in the collection, plus cross-references to authors and journal source data. E-1 (first author) is a listing, alphabetically, by first author; second author, year, file number, and journal source data are included. E-2 (all authors) is an alphabetical listing of all authors; year, file number, and journal source data are included. F (year of publication) is a listing of all references in chronological order; also included are the file number, complete title, and key words. G (key words) is an alphabetical order listing of all key words; file number, year, and full title are referenced. H (reference source) is a listing of journal references in alphabetical order by title of journal; year, first and second authors, and file number are also listed.

The flavorist by referring to one or more of the above six "telephone directory" reports can very quickly learn if any specific paper is in the collection, or, if desired, any and all papers dealing with any specific topic such as raspberry flavor components. The key word file is of particular interest because of its grouping together of papers on the same subject.

MECHANICS OF SYSTEM MAINTENANCE

The article to be added to the system can be in the form of a published paper, thesis, summary, or even an internally generated laboratory report. The number of individual references that can be added is unrestricted. A practical limit, due to running time and print-out size, is estimated at 30,000–50,000 references. This is a number seldom, if ever, approached by the practicing flavor chemist. Very large reference collections could be broken down into smaller subunits and processed separately, if desired.

The actual reprints are assigned a sequential file number as acquired, and in our laboratory are bound in book form in groups of 100. This prevents the loss of reprints

Research and Development Laboratories, McCormick & Company, Inc., Hunt Valley, Maryland 21031.

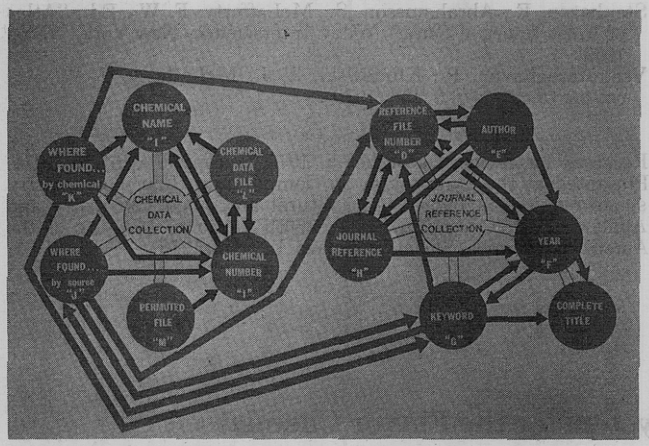

Figure 1. Total system, indicating cross-over between journal reference collection and chemical data collection.

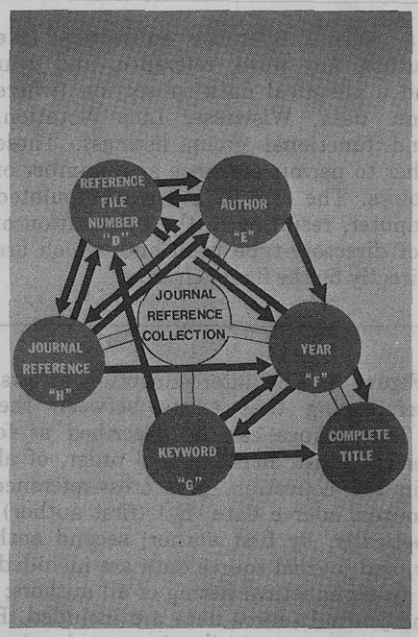

Figure 2. Journal reference collection subsystem.

and permits the easy and rapid referral to any original report in the system. This practice is highly recommended, as it eliminates the aggravating and time-consuming task of locating "strayed" papers.

The newly selected paper is reviewed by a single individual who, being trained and completely familiar with our system, then assigns the appropriate key words. This assures uniformity in the later retrieval of the desired subject material and avoids confusion. The careful and consistent selection of key words is critical to retrieval of the pertinent information. The number of permissible key words per individual paper is 20. However, the total number of key words is unrestricted.

The original paper is then given to a nontechnical person who codes the necessary information (number, title, key words, etc.) onto a coding form which is forwarded to the keypunch operator. An average coding rate is 15–20 references per hour.

The actual updating is done on a batch basis, about once or twice a year. The total computer time to run 1000 file references, *via* punched card input, is about 25 min with an I.B.M. 370-145 system. This results in a print-out of some 203 total pages (about 0.75 in. thick) that is easily distributed and used in the laboratory or office. The original 1000 reprints, by comparison, formed a stack about 20 in. high which is, of course, very cumbersome to sort through to locate the desired paper.

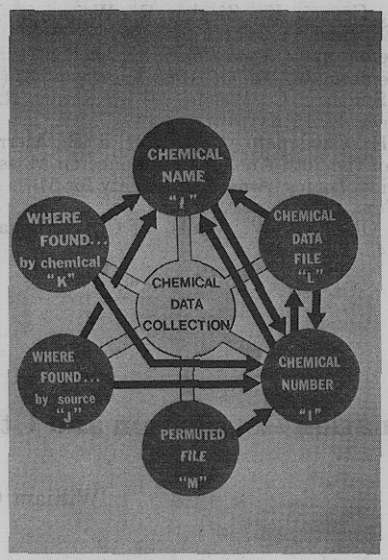

Figure 3. Chemical data collection subsystem.

Table I. Natural Product Chemicals; Where Found, by Source

Source	Chemical no.	Chemical	Journal ref
Allspice	000702	Eugenol	00621
	000810	α-Phellandrene	00621
	001104	Zonarene	00780
Basil	000307	Benzyl alcohol	00795
	000879	*cis*-α-Bergamotene	00795
	000880	*trans*-α-Bergamotene	00795
	001027	β-Bourbonene	00795
	001032	*trans*-Cadinol	00915
	000882	Calamenene	00795
	000883	Δ-Carene	00795
	001037	β-Caryophyllene oxide	00915
	000412	Cinnamic acid	00624
	000889	α-Cubebene	00795
	000906	β-Cubebene	00795
	000749	*p*-Cymene	00561

CHEMICAL DATA COLLECTION

The use of the journal reference collection directories provides for rapid access to any paper in the system on or by any particular subject, author, date, journal, etc. However, it does not allow one to easily answer specific questions frequently raised by the flavor chemist, *i.e.* what trisulfides have been found in food, or, do you have a list of all the compounds reportedly found in peanuts? These questions can only be answered by retrieving the detailed specific information contained within the original papers and not *via* key words which can only describe the type and nature of the paper.

A separate subsystem (Figure 3) has been developed to manipulate this detailed chemical information. It functions in a similar fashion as the system described above. Data are coded and put onto punched cards. Printed reports are periodically generated at the computer center when desired, normally one to two times per year.

Again, the six outer circles of Figure 3 represent the actual computer-generated listings. The interconnecting arrows indicate the cross-referencing that exists between the various printouts. The individual listings can be described

Table II. Natural Product Chemicals; Where Found, by Chemical

Chemical no.	Chemical	C	H	N	O	S	X	Source	Journal ref
000422	Acetaldehyde diethyl acetal	16	14		2			Cocoa	00715
000924	Acetol acetate	5	8		3			Beef	01049
								Coffee	01199
000609	Acetol propionate	6	10		3			Coffee	00789
000167	Acetone	3	6		1			Beef	00735
								Celery	00681
								Chicken	00812
								Chives	00113
								Cocoa	00247
								Coffee	00522
								Coffee	00525
								Coffee	00532
								Ginger	01209
								Ham	00730
								Leek	00909
								Onion	00120
								Parsley	00346
								Sausage	01224
								Spearmint	01155
								Tea	00399
								Tea	00567
000175	1-Acetoxybutan-2-one	6	10		3			Coffee	00525
								Coffee	00532

as follows. Ia (alphabetical listing) is a list of all chemicals in alphabetical order; the chemical's number and molecular formula are also listed. Ib (numerical listing) is a list of the same data as above except the listing is in numerical order. J (where found, by source; see Table I) is a list of each natural product (allspice, black pepper, etc.) in alphabetical order. Under each source is listed, alphabetically, every volatile component reported as being present in that product. The chemical's number and the number of the original journal references are also included. If more than one publication reports the findings of a particular chemical in the same source, then the item is multiple listed, one time for each such journal reference. K (where found, by chemical; see Table II) is a listing of chemicals in alphabetical order; chemical number and molecular formula data are included. Beneath each chemical is printed the source (onion, peanut, etc.) and the file number of the journal reference reporting the findings. L (chemical data) lists along with each chemical's number its name, molecular formula, Wiswesser Line Notation (WLN), molecular weight, boiling point, GRAS ("Generally Recognized as Safe") number, and spectra reference number for ir, mass, uv, and nmr spectra, if available. M (permuted file) is a computer-generated listing which groups together similar functional features. All aldehydes, thiazoles, trisulfides, etc. can be easily located in the alphabetical sorting of the WLN codes. In a typical print-out on some 1140 naturally occurring flavor chemicals the average item was listed 3.2 times, once for each functionality. The Wiswesser Line Notation is the basis for this sorting program; data input is *via* punched cards.

The print-outs or directories can be used alone or in conjunction with the journal reference directories to readily determine what chemicals are in the system, in what products they are found, and what original reference reported the findings. We have found these to be the type of questions most frequently asked by all flavor chemists and flavor compounders.

The first five print-outs described above, when prepared from an input of approximately 1000 chemicals, result in a print-out less than 1.5 in. thick. The permuted index for 1141 chemicals is 74 pages in length (about 0.25 in. thick). Both are thus quite compact and manageable.

COMPLETE SYSTEM

The journal reference and the chemical data collections are interconnected, as shown by the connecting arrows in Figure 1, to form the complete system. It can be seen that the "where found by chemical" and "where found by source" files in the case of the chemical data collection and the "key word" and "file number" files in the journal reference collection serve as the key intermediates. However, once can enter into the system at any point and gain ready access to any other information contained within the entire system.

The system described above is currently functioning in our laboratories and is proving to be extremely useful. Other laboratories with other interests and objectives might prefer a somewhat modified system.

LITERATURE CITED

Steed, S. P., *J. Chem. Educ.* **48**, 375 (1971).

Received for review October 15, 1973. Accepted March 26, 1974. Presented at the Symposium on Computers in Flavor Chemistry, 166th National Meeting of the American Chemical Society, Chicago, Ill., Aug 1973.