the worker and his colleagues would not usually repeat the work. Actual estimates of loss of scientists' time from inadequate use of recorded knowledge have ranged from 30 to 80%.[21] If excellent indexes prevent as little as 1% of unwanted duplication, then their construction and use is easily justified on the basis of economics alone. If other factors are considered, such as the irreplaceable loss of time of scientists and engineers, then even greater increase, when needed, in the budget for indexes are probably justified.

REFERENCES

(1) C. L. Bernier and E. J. Crane, *J. Chem. Doc.*, **2**, 117 (1962).

(2) E. J. Crane, Ed., "CA Today—The Production of Chemical Abstracts," American Chemical Society, Washington, D. C., 1958, p. 38.

(3) U. S. Senate Committee on Government Operations, Subcommittee on Reorganization and International Organizations, "Coordination of Information on Current Research and Development Supported by the U. S. Government," Report 263 of the 87th Congress, 1st Session, May 18, 1961, p. 229, Appendix 1; "Cost of Research per Technical Article Describing Research Results," 1961, U. S. Government Printing Office, Washington, D. C.

(4) Cost data from the Defense Documentation Center and from *Chemical Abstracts*.

(5) Ref. 2, p. 61.

(6) Indexers with ten or more years of experience at *Chemical Abstracts* regularly found this range of uncertainty in number of kinds of headings of nonorganic subject entries chosen per abstract.

(7) Ref. 2, pp. 47-48.

(8) L. A. Schultheiss, D. S. Culbertson, and E. M. Heiliger, "Data Processing in the Library," Scarecrow Press, Inc., 1962, p. 36.

(9) Ref. 8, p. 176.

(10) J. C. Costello, Jr., *J. Chem. Doc.*, **3**, 165 (1963).

(11) Ref. 1, p. 120.

(12) C. L. Bernier, *Am. Doc.*, **8**, 47 (1957).

(13) Ref. 2, p. 49.

(14) Ref. 2, p. 64.

(15) Ref. 1, pp. 120-121.

(16) "ASTIA Thesaurus of Descriptors," Second Ed., The Office of Technical Services, Washington, D. C.

(17) C. L. Bernier, "Correlative Indexes. IX. Vocabulary Control," accepted for publication by *J. Chem. Doc.*

(18) The DDC information-retrieval system using correlation of descriptors by computer for 10^6 reports is not incapacitated by irrelevant retrieval

(19) Ref. 1, pp. 118-119.

(20) C. Burt, "A Psychological Study of Typography," University Press, England, 1959, pp. 8-9.

(21) I. Hirsch, W. Millwitt, and W. J. Oakes, "Increasing the Productivity of Scientists," *Harvard Business Review*, **36**, No. 2, 66 (1958).

Automatic Preparation of Selected Title Lists for Current Awareness Services and as Annual Summaries*

By ROBERT R. FREEMAN**
The Chemical Abstracts Service, Columbus 10, Ohio

JOHN T. GODFREY
E. R. Squibb and Sons Division, Olin Mathieson Chemical Corporation, New Brunswick, New Jersey

ROBERT E. MAIZELL
Olin Mathieson Chemical Corporation, New Haven, Connecticut

CHARLES N. RICE and WILLIAM H. SHEPHERD
Eli Lilly and Company, Indianapolis, Indiana

Received October 17, 1963

This paper constitutes a progress report in which our experience in the use of computers for searching *Chemical Titles* over a period of more than a year is reviewed. The limitations of *Chemical Titles*, and of titles in general, are well recognized by the authors.[1] In the present experiments we have accepted these limitations.

In 1961, Chemical Abstracts Service (CAS) began issuing *Chemical Titles* (*CT*) as a current awareness service for chemists and chemical engineers. The development of *CT* has been described elsewhere.[2] The by-product punched card and magnetic tape records of references to thousands of articles from chemical journals remained unexploited during the first year of publication.

A significant proportion of the users of *CT* are information scientists who search each issue for a group of chemists.[3] In a typical operation, references to papers of interest are copied, pasted, or typed on file cards, and individuals are notified. While effective, this method ties up many hours of the searcher's time.

Early in 1962, however, we realized that the machine records might serve as a useful basis for automatically retrieving references of interest to company research efforts. Time required for conventional searching of *CT* could be reduced or eliminated altogether, and computer output could be easily disseminated to individuals with minimal clerical work.

* Presented before the Division of Chemical Literature, 144th ACS National Meeting, Los Angeles, Calif., April 1, 1963.

** Now with the American Meteorological Society, P. O. Box 1736, Washington 13, D. C.

CAS agreed to cooperate with Eli Lilly and Company and Olin Mathieson Chemical Corporation in experimental programs to test the feasibility of machine searching *CT*'s approximately 3200 titles per issue. These experiments provided an opportunity to test in a field situation, *i.e.*, one in which there would be feedback from a sample of scientists who are representative users of current awareness services.

An Experiment With Olin.—The New Haven Technical Information Services Section of Olin organized a set of terms of relevance to some of the Company's research interests. This set was forwarded to CAS, where, at the time each issue of *CT* was prepared, a separate listing of all titles which contained the terms was created. Each listing was sent to Olin more than a week before it was possible for them to receive a printed copy of *CT*.

More than 25 research scientists in key positions at the Olin Research Center, New Haven, inspected the listings carefully. The participants were organic and inorganic chemists, chemical engineers, and information scientists at levels up through research managers and directors.

These scientists are presently assisted in keeping up-to-date by a weekly patent bulletin, circulation of title pages of journals, and an informal "current awareness" service. These services are offered by the staff of the Technical Information Services Section. A permanent record of pertinent external literature is maintained on a selective basis.

The listings performed a dual function at Olin. As a current awareness tool, they are regarded by the Technical Information Services Section as being of value even if only a few pertinent references are located within a minimum time. Titles are also thought to be useful as retrospective search files. For current awareness purposes, the listings were divided according to request terms and given, through the department heads, to departments whose interests were pertinent to the terms. Evaluation questionnaires accompanied all transmittals.

The Lilly Searching Experiment.—The experimental system developed at Eli Lilly and Co. is similar to that proposed by Luhn.[4] Written for an IBM-1401-705 computer complex, the Lilly program matches interest profiles against a file of key words which occurred in titles. The concept differs from Luhn's in that key words are not derived from the body of articles and no frequency-of-occurrence or other statistical criteria are used.

In January, 1962, 53 employees of the Lilly research, development, and control laboratories were invited to participate in this current literature alerting experiment. Among them were organic chemists, biochemists, pharmacologists, biologists, microbiologists, virologists, bacteriologists, analytical chemists, physical chemists, and plant science research personnel.

Various individuals were invited to participate on different bases. In one group, individuals were asked to indicate their own specific interests in terms of a list of key words. Those in another group were asked to serve as literature alerting members of research project teams, and to provide key words that described the interests of the team. Some whose major responsibility lies in administration of research, and whose interests therefore were expected to be quite broad, were asked to participate.

When the vocabularies of key words had been collected and the request cards had been punched (a typical example in shown in Fig. 1), the combined vocabularies were computer-matched with key words from *CT*. In the early stages of the experiment, the Lilly program could accept only requests expressed as single terms. Later it was modified to accept multi-term requests, involving terms related by "and," "or," and "not." The CAS program has included this capability since it was put into use.

When single-term requests were used at Lilly for an issue of *CT*, 11,000 to 23,000 references were disseminated to the participants from searches with 1900 terms. Twenty-five per cent of the participants revised their interest profiles when the improved program was introduced. Fifty participants remaining in the experiment by January, 1963, had evolved 1681 requests of which 1486 were single-term requests and 195 were multiple-term requests. The multiple-term requests consist of nearly five terms per request. The number of titles disseminated decreased to 7,000–14,000 for each issue of *Chemical Titles*.

A questionnaire, distributed with each listing, was returned for our analysis. The data reported are for the last four months of 1962 (*CT 1962* #17–24). In addition, each individual was invited to comment informally concerning his satisfaction or lack of satisfaction with the service.

Table I summarizes the number of individuals invited to participate in the program. In the same table are listed the number who have reported faithfully the results they obtained. The data will include only those from the latter group.

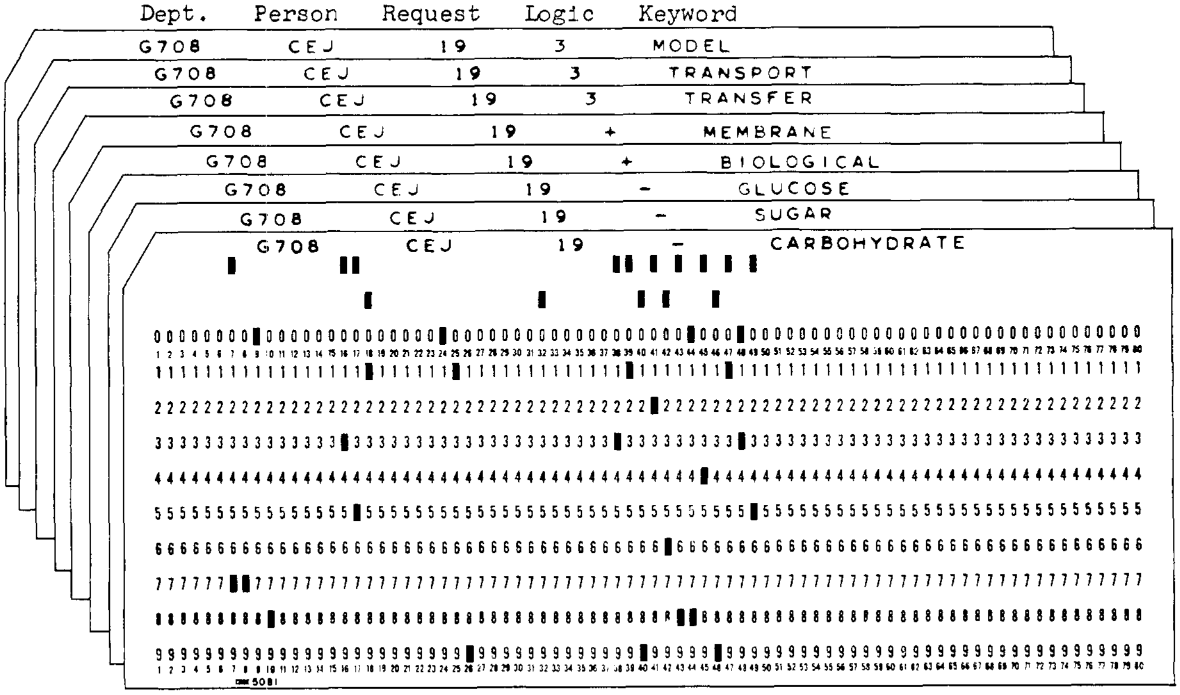
The level of value of the references to the individuals was measured in the following manner. All of the titles printed after a search of *CT 1962* #17–24 with each individual's key word list, are totaled. The number of references among these that the individual reports to be of some value to him are totaled. The latter total, divided by the total number of titles printed, represents the value the reference list has to the individual. Likewise, an average for all 22 individuals is obtained by totaling (1) all papers and (2) all papers reported as being of interest by any individual and dividing the latter by the former figure. The lowest and the highest individual levels for members of each of the various disciplines are shown in Fig. 2.

The number of key words per request are tabulated in Table II and Table III and are assumed to measure the complexity of vocabulary used by each individual. A value of 1.00 applies to a vocabulary made up of single key-word terms. Anything greater than 1.00 indicates that additional key words act for either inclusion or exclusion.

The same order of line entries is maintained in Tables II and III. In Table III the number of titles printed per request is taken to be a measure of the potential suitability of *CT* as a source of information for the individual, whereas the number of titles of interest per request is taken to be the utility of *CT* as a source of information for the individual. One should note that pharmacologists and biologists are among those who find *CT* to have the most titles of interest per request.

The number of single key word and multiple key word requests are shown in Table II. Multiple key word

MEMBRANE • BIOLOGICAL • (MODEL + TRANSFER + TRANSPORT) - (GLUCOSE + SUGAR + CARBOHYDRATE)



logic n number of inclusions $(9 \geq n > 0)$

+ inclusion

- exclusion $(\leq 50; \geq 0)$

Fig. 1.—Example of a search request.

Table I
Summary of Participation at Eli Lilly & Co.

	A[a]	B[b]	C[c]	Total
Chemists				
Organic	1/4[d]	3/6	...	4/10
Biological	1/6	3/4	...	4/11
Other	8/13	...	0/1	8/13
Biologists				
Pharmacologists	2/2	2/3	...	4/5
Other	3/8	0/3	0/1	3/12
Total	15/33	8/16	0/3	23/52

[a] Individuals asked to provide their personal interest profile in key-word form. [b] Individuals asked to provide research project group's interest profile in key-word form. [c] Administrators asked to provide their interests in key-word form. [d] Key to figures: 1/4 means that one individual reported faithfully for *CT 1962* #17-24 from four invited to participate.

requests are divided into those in which either the principle of exclusion or the principle of inclusion was utilized. If, for example, a participant was interested in the subject of *absorption in biological systems*, but not in *light absorption*, he might utilize the word *absorption* as a primary key word in a request, and list a series of key words to be excluded such as: *ultraviolet, infrared, etc.* Any title having the word *absorption* in it would then be inspected for the secondary key words to be excluded. If one of these were found, the title would be discarded. In a request having inclusion key words one might, for ex-

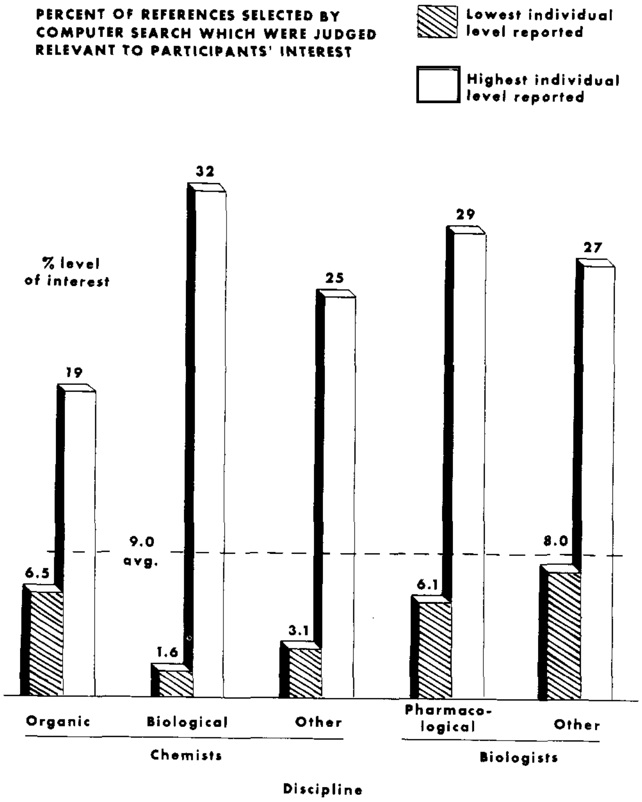


Figure 2.

Table II
Classification of Vocabularies

Discipline	Class[a]	No. of requests	Type of request[b] S	I	E	Key words per request
Chemists						
1 Organic	B	47	47	0	0	1.00
	A	28	28	0	0	1.00
	B	37	36	1	0	1.03
	A	13	10	1	2	1.07
2 Biological[c]	B	25	25	0	0	1.00
	A	67	67	0	0	1.00
	B	114	114	0	0	1.00
	A	86	35	41	10	3.07
	B	33	6	27	0	5.36
3 Other	A	44	43	1	0	1.05
	A	30	25	2	3	1.60
	A	31	26	2	3	1.68
	A	23	14	0	9	1.91
	A	27	15	0	12	1.96
	A	22	18	4	0	2.04
	A	14	8	6	0	2.64
Biologists						
4 Pharmacologist	A	78	78	0	0	1.00
	B	27	27	0	0	1.00
	A	19	15	0	4	3.16
	B	21	13	0	8	4.90
5 Other	A	23	23	0	0	1.00
	A	29	29	0	0	1.00
	A	21	21	4	0	1.95

[a] A = individual with his own interests; B = individual representing a team's interests. [b] S = single key word; I = logical inclusion; E = logical exclusion. [c] One individual provided two sets of requests to give five sets for four individuals.

Table III
Levels of Interest

Key words per request	Titles printed No.	No. per request	Titles of interest No.	%	No. per request
1.00	997	21.2	65	6.5	1.4
1.00	1083	38.7	85	7.8	3.0
1.03	478	12.9	89	19	2.5
1.07	1488	115	119	8.0	9.2
1.00	2877	115	145	5.0	5.8
1.00	7925	118	452	5.7	6.8
1.00	4143	36.3	68	1.6	5.8
3.07	2280	26.5	325	14	3.7
5.36	241	7.30	76	32	2.3
1.05	3369	76.6	178	5.3	4.1
1.60	1848	73.9	58	3.1	2.3
1.60	2838	94.6	98	3.5	3.3
1.68	1338	43.2	175	13	5.6
1.91	3477	151	126	3.6	5.4
1.96	2154	79.7	126	5.9	4.7
2.04	408	18.6	102	25	4.6
2.64	1411	101	105	7.4	7.5
1.00	984	12.6	60	6.1	0.77
1.00	1002	37.1	293	29	11
3.16	1265	66.5	221	18	12
4.90	1129	53.8	70	6.2	3.3
1.00	449	19.5	43	9.6	1.9
1.00	2876	99.2	787	27	27
1.95	1246	59.3	99	8.0	4.8

ample, have a request in which the primary key word is *transfer* and the next key word, which must also be in the title is *charge*. The title would not be chosen unless both words were found in the title, regardless of their relative positions within it. One may also search the title for other key words that must be present as in the foregoing situation, for example, either the word *electron* or the word *proton*. In an inclusion request there is a great danger that titles of interest will be lost because there are alternate ways of stating the information. Many subjects of interest might be excluded. The participants were warned of this possibility

A comparison of simple key word, logical exclusion key word, and logical inclusion key word requests was made by one of the authors. A group of seven requests was searched, utilizing seven single key words, and results of this search for *CT 1962* #17–24 were compared with results of a search utilizing the same key words with 74 additional key words added to limit the type of information retrieved. The results of this comparison are shown in Table IV. Another 20 key words were used as single words for a search of the same numbers of *CT*. Then a vocabulary covering the same subject matter, but involving inclusions, was searched. This latter vocabulary contained 28 requests, because some of the 20 requests were divided into two or more requests to provide for the proper search logic. A total of 77 key words were added as secondary words used for inclusion with the 28 request key words. The results of these searches are shown in Table V. By inspection one sees that retrieval of information of interest is more nearly perfect when one utilizes exclusion rather than inclusion. The number of titles printed is reduced almost tenfold in the latter case, as compared to a little over twofold in the former.

Table IV
Comparison of Single Key-Word Search
vs.
Search with *Exclusion* Terms Added[a]

		Single key word	Multiple
1	Total titles retrieved	2015	910
2	Number of immediate interest in current work	18	17
3	Number of high degree of importance to professional interests	24	22
4	Number of general reference value	7	7
5	Number of some general interest	83	77
6	Total relevant titles retrieved (sum of 2–5)	132	123

[a] Totals for eight issues of *Chemical Titles*.

Table V
Comparison of Single Key-Word Search
vs.
Search Limited by *Additional* Terms
Which *Must* Also Be Present

		Single key word	Multiple
1	Total titles retrieved	4558	479
2	Number of immediate interest in current work	21	11
3	Number of high degree of importance to professional interests	32	16
4	Number of general reference value	17	10
5	Number of some general interest	155	47
6	Total relevant titles retrieved	225	84

[a] Totals for eight issues of *Chemical Titles*.

Annual Summary Bibliographies.—Existence of a retrieval program provided the opportunity to conduct another experiment. Since magnetic tapes for each issue of *Chemical Titles* were readily available, bibliographies for broad areas of chemistry could be prepared and issued within a short time after the end of the year or any other time period. The bibliographies would represent a sort of annual summary of work done in the fields they represented.

To evaluate the usefulness of such bibliographies, the Research and Development Department of the Chemical Abstract Service used the IBM 1401 computer to retrieve from 1962 issues of *Chemical Titles* all titles which refer to (1) chromatographic techniques, (2) catalysis, and (3) the platinum metals. The first two of these collections were subsequently key-word indexed and author indexed to provide publications which resemble *Chemical Titles.* The third was processed by a new program, which produced an index in which each key word is followed by the complete title in which it occurred and the reference code for that title. The latter technique has been called Keyword-out-of-Context, or KWOC indexing. Table VI summarizes the content of the three collections.

Table VI
Summary of Three Special Annual Summary Bibliographies[a]

Term used for retrieval request	Titles retrieved
(1) Cataly-	1371
(2) Chromatogr-	1514
(3) Irida-, iridi-, osma-, osmi-, osmy-, platina-, platini-, platino-, platinu-, pallad-, rhodat-, rhodit-, rhodium, ruthen-	693

[a] Corpus searched: approximately 75,000 titles.

Having retrieved this many titles, we wondered if false drops and failures to retrieve pertinent titles might occur significantly. Examination of the titles in the three collections showed only one false drop; a paper dealing with the flowering plant, *Centaurea ruthenica,* was retrieved with those dealing with ruthenium.

Retrieval failures are more difficult to find, but we located one major source. Journals which define x as their entire sphere of interest are likely to publish titles which do not mention x. For example, the 13 issues of *Journal of Chromatography* which were indexed in *Chemical Titles* during 1962 contained 301 titles, 74 of which did not mention any word beginning with the fragment, *chromatogr-*. An analysis of these titles showed that two strongly associated terms, *electrophor-* and *(ion, anion, cation) exchange* occurred in a total of 35 of the 74 titles. Other clues were *chromatoplates* (2), *gradient* (3), R_f *values* (1), R_m *values* (1), and co-occurrence of terms of list A with those of list B.

List A	List B
Behavior	Dextran gel
Migration	Sephadex
Separation	Alumina columns
Fractionation	Thin-layers
Filtration	Cellulose layers
	On paper

While enlarging the number of request terms to include these *associated terms* would doubtless yield more relevant titles, the number of false drops probably would also show an increase.

Another, more practical approach in compiling bibliographies of the type described here is to modify the program to allow all titles from any given journal to be selected.

Participant Evaluation.—After nearly a year's participation in the experiment, 85% of the Lilly group wished to continue receiving selected references; 65% said that the system significantly reduced their dependence on other means of obtaining current references; while 88% said that the service markedly extended their access to current literature.

The participants at Olin pointed out that the convenience, appearance, and legibility of the reference lists enhanced their utility. They feel that the capability of the service to inform them of even a relatively few papers of interest within the shortest possible time is of great value. Many of the papers retrieved were from journals not commonly read by American chemists because of language difficulties.

Economic Evaluation.—What about the cost of providing a service such as we have discussed? There appear to be four major factors which interact to effect the time and cost of a search. They are: (1) the total number of terms regardless of how the terms are distributed within the requests; (2) the number of references in the file, and the number of index entries which provide access to the file; (3) the number of references in the file which satisfy the terms of the search; and (4) familiarity of the searcher with the vocabulary of the search.

Our experiments included a number of human searches, either parallel to or similar to the machine searches we conducted. Although we do not yet have enough data to draw final conclusions, the most favorable combinations of the four parameters do not appear to justify the present cost of machine search. However, considerations of speed, available staff, and the alternative uses for the time consumed if each scientist were required to make his own searches may justify the cost for machine searching. The cost of operating second-generation systems we contemplate may be less than that of a manual system.

Future Development.—A number of possible improvements which would facilitate selection of relevant titles have been suggested by participants in these experiments. These are detailed in another paper.[5]

CONCLUSION

In summary, we have experimented with automatic retrieval of references from *Chemical Titles*. The results have provided a current-literature alerting system at Eli Lilly and Co., a current awareness system and a retrospective search file at Olin Mathieson Chemical Corporation, and indexed annual bibliographies for various areas of chemistry at the Chemical Abstracts Service. Comparisons with human searches of *Chemical Titles* suggest that the present experimental machine systems are more expensive to use under certain conditions. Advantages gained from speed and utility of the search output, and prospects for improved programs, indicate that systems which are more practical and economical than those now in use will evolve.

ACKNOWLEDGMENTS

The authors wish to express their particular thanks to Mrs. J. B. Haglind of Olin Mathieson Chemical Corporation; and to Mrs. Claudene Frank, Mr. Karl Hardey, and Mr. William Muirhead of Eli Lilly and Co. for their contributions to the experiments described herein.

REFERENCES

(1) See R. E. Maizell, *Rev. Doc.*, **27**, 106 (1960).

(2) R. R. Freeman, and G. M. Dyson, *J. Chem. Doc.*, **3**, 16 (1963).

(3) Unpublished readership survey, July, 1962.

(4) H. P. Luhn, *Am. Doc.*, **12**, 131 (1961).

(5) R. R. Freeman in "Automation and Scientific Communication," H. P. Luhn, Ed., Papers contributed to the Theme Sessions of the 26th Annual Meeting of the American Documentation Institute, Washington, D. C., 1963, part 2, pp. 213–214.

Some Unusual Features of a Chemical Retrieval System Used in the Eastman Kodak Company*

By CARL R. HAEFELE and JOHN F. TINKER
Research Laboratories, Eastman Kodak Company, Rochester, New York

Received April 22, 1963

This paper discusses briefly the mechanical and manipulative aspects of a system used in the Kodak Research Laboratories to index and retrieve chemical information. This system is used to locate the information on specific compounds and to show where samples can be obtained within the Company. Just as the system can uncover a single compound, it can also be used to retrieve all compounds of a given class, *i.e.*, with a given functional group.

Some features of our system are common to other systems, but there are certain aspects which appear to us to be unique. We claim no priority for these innovations, but we have not encountered them elsewhere, and so we present them as a new approach to the handling of chemical compounds.

The large number of organic compounds synthesized and used in the Kodak organization will inevitably require the use of high-speed computers for efficient, economical retrieval of information. However, the system has been developed so that the initial work can be done with simple sorter–collator equipment until the volume of data makes this approach impractical. When this happens,. a conversion will be made to Minicard, computers, or some other type of high-speed hardware. This means that information entered in punched cards for use on the sorter–collator equipment has had to be in a format that could be accepted by a computer. This restriction created some limitations; random superimposed coding, or multiple punches, in columns that would be unintelligible to a computer were precluded. These limitations forced us to abandon the possibility of entering all chemical structural data for a single compound on one card. Once we decided

*Presented before the Division of Chemical Literature, 142nd National ACS Meeting, Atlantic City, N. J., September 12, 1962.