A Multiple Card Abstract Retrieval System

By JOHN W. WEIGL
General Aniline and Film Corporation, Binghamton, New York
Received October 9, 1961

While a number of more or less automated information retrieval systems have been used with success in large industrial and government research establishments, it remains difficult for scientists in smaller laboratories to index their literature abstracts in such a way that they can easily identify and locate them. It is the purpose of this note to describe a simple, inexpensive system which has been found to be very effective in our laboratory of about twenty professional persons. It may prove equally useful, with suitable modifications, to individual researchers.

The abstractor's indexing problem is four-fold:

(1) To prepare a digest of a given document in such a way as to make it physically separable from notes on other references.
(2) To apply to the digest coded marks which bear a one-to-one relationship to preselected items of interest to him (or his readers).
(3) To provide practical means whereby the abstract may be retrieved, upon demand, according to each of the classifications encoded upon it.
(4) To provide simple means, if possible, whereby the information and the code marks may be altered at any time in order to bring the abstract up to date, to add or delete sorting classifications, *etc.*

In our system, multiple photocopies of each abstract are filed, each copy located physically according to its encoded "address," and retrieved by visual scanning through appropriate categories. This approach results in relatively bulky files which can, however, be tolerated easily in small to medium-sized systems (up to perhaps 15-20,000 entries). The user also benefits from his ability to scan entire groups of abstracts visually since he is able to "browse" for poorly encoded information; he can frequently pick up relationships between separately encoded items which would be missed by computer search.

This system has been in use in our laboratory since its inception by C. E. Herrick, Jr., some eight years ago. It was made possible by the advent of relatively inexpensive copying machines and diazo-sensitized filing cards manufactured by Ozalid. The original abstract is typed (or hand-written) on a translucent master form, and encoded according to an arbitrary code. The required number of photocopies is prepared on a standard Ozalid machine, and the librarian files as many of these as are indicated by the subject code. One extra set of copies is circulated to the staff to inform all members of what the others have learned. Additional copies are made on request, for the abstractor, as well as for other individual desk files. A minor drawback is that occasionally two or more persons have abstracted a given document—though not necessarily from the same point of view. We have experienced at most 2% redundancy of this sort. By confining the writing space on the master sheet to a pre-marked 3 × 5 inch area but making the sheet larger (*e.g.*, 4 × 5.5 in.), it is possible to copy the *identical* master onto cards which fit into existing 3 × 5 inch, 4 × 5.5 in., or 8.5 × 11 in. filing systems.

Our subject code is alphanumerical. Blank forms, offset-printed a few thousand at a time onto translucent paper, and pads of "first" and "second-page" sheets are furnished to all staff members who read technical journals and patents. (Double-coated diazo cards are commercially available for two-page abstracts.) After the first few years, abstracts began to pour in too fast for the typist to keep up with them, and it became necessary to switch to hand-written originals. Despite occasional difficulties in deciphering the handwriting of some members of the staff, the system works smoothly even without the services of a typist.

The expense of the entire operation is low. Aside from an Ozalid machine which is likely to be accessible in a nearby office or drafting room, the costs are confined to the master sheets and light sensitive cards, and to the secretary's and librarian's time for duplication and filing (about 6 hours per month for our system of 3000-odd abstract masters).

The new Abstracts of Photographic Science and Engineering Literature (publ. by Columbia Univ., ed. by H. Lester) is available in a special translucent loose-leaf edition, suitable for multiple-card indexing and reproduction.

It is a pleasure to acknowledge the contribution of our librarian, Mrs. Rosemary Frey, who is responsible for the successful operation of the system.

Indexing Physical Chemical Properties of Antibiotic Substances

By A. NEELAMEGHAN
Hindustan Antibiotics Ltd., Pimpri, India
Received July 10, 1961

Chemical compounds usually are identified by comparing their physical and chemical properties with those of known ones. In the case of biologically active compounds such as antibiotics comparison of biological properties is also necessary. For such identification studies we need not only collected information but also facility of approach to the data from various angles.

The physical, chemical and biological properties of antibiotics have been subjects of numerous reviews. They have been studied as groups of different types of chemical compounds (amino acids and peptides, polyenes, polyacetylenes, *etc.*), as compounds active against different species of microorganisms, or through correlating chemical structure with biological activity. Several handbooks and

manuals of collected data on antibiotics have been published. These reviews and compilations are not, however, done specifically from the point of view of identifying newly isolated antibiotics. In a few works [1-3] more detailed classification and indexing of data provide for different approaches to the information recorded. Recently, the Antibiotics Research Institute, Warsaw, issued printed cards to supplement the classification and indexing of data in the book of antibiotics published from that Institute. [1] There is a basic set of cards, one for each antibiotic (726 cards in the file now available), recording physical, chemical and biological data, and the original references. These cards can be rearranged as desired. A separate set of bibliographical reference cards for each antibiotic is also available. Certain punched card systems for retrieval of information in the chemical field include antibiotics. [4,5] Ohrmund [6] describes an IBM punched card system more specifically developed for antibiotic data comparison studies. The scope of application of mechanized systems is indeed vast and their capacity for information retrieval appears to be limited only by our ingenuity to feed appropriately coded data into the system. And yet methods for information retrieval often have to be adapted to meet special requirements and the equipment that can be afforded and operated economically.

At Hindustan Antibiotics information on antibiotics is systematically collected. A permanent classified card index to the literature is maintained and kept up to date with the cards prepared for a weekly literature bulletin. [7] A basic file of antibiotic data similar to that described by Ohrmund [6] was prepared for reference. When antibiotics were isolated in this laboratory it became necessary to compare their properties with those of known ones. Such preliminary comparisons not only help in identifying compounds but also in making further studies on the newly isolated compounds. Provision of indexes to physical and chemical data such as melting point, ultraviolet absorption maxima, optical rotation, empirical formulae, *etc.*, facilitates narrowing the range of search among the known compounds and ascertaining which of the compounds have properties similar to or closely related to those of the compound at hand.

Among the various methods considered for the indexing of the data, punched cards and uniconcept coordinate indexing appeared most feasible. We were also interested, however, in a data index which the scientist could make for himself and have on his table for reference. Such an index needs to be compact. Simple edge-punched cards could not be obtained easily and mechanically operated systems would not be economical to sort a deck of about 1,000 cards; furthermore, the researcher cannot have such an index for ready reference on his table. Uniconcept coordinate indexing would also involve about a thousand cards. As a practical alternative a compact indexing method was worked out, examples of which are described below.

The basic data file may be arranged in any preferred way, let us say, by species of producing organism, and each data sheet serially numbered providing for accommodation of new data sheets in each group:

From bacteria	1 to 199
From actinomycetes	200 to 999
From moulds and fungi	1000 to 1499
From lichen, etc.	1500 to 1599

Data such as ultraviolet maxima, melting points, optical rotation, *etc.*, can be indexed on charts as in Fig. 1, 2, and 3. The charting can be for all of the data sheets or for individual groupings just mentioned. We prefer the latter.

In Fig. 1 for ultraviolet maxima, each square represents one wave length in mμ on the horizontal in units (0 to 9) and the numbers (210, 220, *etc.*) in the first vertical column denote the beginning of the range. Serial numbers of the data sheets (which serve as accession numbers of the antibiotics) are posted in the appropriate squares such that compounds having the same ultraviolet maxima fall in the same square. Thus, for antibiotics with ultraviolet maximum 236 mμ we look up the data sheets numbered 39, 53, 88, 111. If the compound on hand has an additional maximum at 242 mμ we know that only the antibiotic numbered 88 has these two maxima.

In Fig. 2 for the melting point, the serial numbers of the data sheets are posted in the squares headed by the relevant melting point range such that serial numbers of antibiotics having the same melting point range fall in the same square. The melting point range is in steps of 5°. Normally this range is necessary in preliminary comparisons, when the exact melting point of the compound on hand may not have been determined.

λ Max	0	1	2	3	4	5	6	7	8	9
210	39 41	166 207 208 253	29	7			7 129	70 214		
220	121	26			53	12 69	69 129		70 107	
230	9 77 121 165	165	88 89 111		51 165	52 78	39 53 88 111	41 70	161 215	158 161
240	9 230	18 70 75 116	72 88 100 117	104	111	75	29 53 57 75	26	39 75	41
250	30 127 40 194 58 198 69 201 98	104 215	84 173	178 180 199 200 211		37 183		129	84 85 89 110 173 183	60 214 226

Fig. 1.—Chart of ultraviolet absorption maxima.

M.P.	25–30	30–35	35–40	40–45	45–50	50–55	55–60	60–65	65–70	70–75
S.Nos.		26	91		36	70	29	21 32	33	130 218
M.P.	75–80	80–85	85–90	90–95	95–100	100–105	105–110	110–115	115–120	120–125
S.Nos.	144 210	131	21 167 44 149	11 38 57	51	8	34 153 124 125 136	6 185	5 17 121 214	7 123 64 146 78 148 85 209
M.P.	125–130	130–135	135–140	140–145	145–150	150–155	155–160	160–165	165–170	170–175
S.Nos.	7 160 85 120 159	11 89 54 101 65 108 77 137	22 150 111 154 134 155 135	89 169 99 231 117	10 114 52 140 88 166 100 207	87 4 110 157 129	47 221 73 186 220	18 67 23 145 42 53	95 219	3 58 142 15 75 143 30 139 37 141

Fig. 2.—Chart of melting point ranges, ° C.

		0	1	2	3	4	5	6	7	8	9
0	C							73	58 127 126	128	129
	H							126 127		58 76 73 128	
	N		1 75 88 114 2 81 91 3 83 96 4 84 100 19 86 110 52 87 113	51 90 73 101 76 104 77 105 80 106 89	108 115	34 103 111 126 131	97	82 85		92 99 112	
	O			59 126 75 127 76 128 77 129 99	58 80 85 89	34 86 126 78 87 81 91 83 97 84 116	60 61 73 78 82 88	63 131 64 67 96 116	62 65 100	66	18a 104 90 105 92 106 98 101 103
	S			126 129 127 131 128		126					
	C		75	76 78 77	59 80	81 83 82 126	84 87 131 85 88 86 90	60 91		61	
10	H	59 129	77	126	75	78 131			84	60	60 81
	N	98									
	O	26 90	69 112	110 111	27 68	1 3 20 109 2 19 108				51 114 113	4
	C	92		62 96 63 97		64 99 98	65 100	66 103 101	104 105	67 106	
	H	60 82 85 62 83			86 88 87 91	96		80		89 90	98
20	O	52 115	155	115						110	

Fig. 3.—Chart of index of empirical formulas.

In a similar manner other data can be charted out. A specimen empirical formula chart is given in Fig. 3. Probably an arrangement of the empirical formulas in the usual way by carbon, nitrogen, hydrogen, *etc.*, may be easier to consult. But in the chart new formulas can be indexed as and when they become known. It is also easier to find certain types of data, *e.g.*, all antibiotics containing sulfur or chlorine.

We have made charts for different groups of antibiotics and research workers find them more compact and easier to handle than card indexes.

There is a great deal of similarity between the charts and uniterm cards, if we consider each square in the chart as one uniterm card.

REFERENCES

(1) Korzybski, T., and Kurylowicz, W., "Antybiotyki." Warszawa, Panstowowy Zaklad Wydawnictw Lekarskich, 1959.

(2) Spector, W. S., Ed., "Handbook of Toxicology," Saunders Co., Philadelphia, Pa., 1957, Vol. 2, "Antibiotics."

(3) Waksman, S. A., and Lechevalier, H. A., "Guide to the Classification and Identification of Actinomycetes and Their Antibiotics," Williams & Wilkins, Baltimore, Md., 1953, p. 168–184.

(4) Kuentzel, L. E. "Punched Cards as Aids to Qualitative Chemical Analysis by Spectral Methods," "In Punched Cards," ed. by R. S. Casey, J. W. Perry, A. Kent and M. Berry, Reinhold Publishing Corp., New York, N. Y., 1958, p. 175–231.

(5) Wheeler, K. W., Andrews, E. R., Fallon, F., Krueger, G. L., Palopoli, F. F., and Schumann, E. L. "Structure Code for Organic Compounds," *American Documentation*, **9**, 198–207 (July 1958).

(6) Ohrmund, M. "Antibiotic Literature File for Chemists." *Advances in Chemistry Series*, **16**, 143–155 (1956).

(7) Neelameghan, A., "A Bibliographical Bulletin," *IASLIC Bulletin* **2**, 5–8 (1957); "A Library Service for Antibiotics Research and Development," *IASLIC Bulletin*, **3**, 84–90 (1958).