(2) H. P. Luhn, Keyword-in-Context Index for Technical Literature (KWIC Index). IBM Advanced Systems Development Division report RC-127, August 1959. (The first published use of the Luhn KWIC system appears to be in the IBM bibliography, Literature on Information Retrieval and Machine Translation, September 1958.)
(3) J. W. Tukey, "Keeping Research in Contact with the Literature: Citation Indices and Beyond," *J. Chem. Doc.*, 2, 34–37 (1962).

A Correlative Indexing and Retrieval System for the Screening of Biological Data*

By ARTHUR W. ELIAS** and MARSHALL R. WARREN
Warner–Lambert Research Institute, Morris Plains, New Jersey
Received July 13, 1961

This system for retrieval of biological data was developed in response to the need of providing access to a large body of test results on about 10,000 chemical compounds. A further need was the *arrangement* of such results in an order that would enable correlative searching.

We began by analyzing the characteristics of our particular problem. The subject matter was diverse, and covered several biological disciplines. Within each discipline were many testing techniques. Finally, some of the test data were expressed as numbers; others, as judgments (for example, "this compound has *fair* activity"); and still others as descriptions or observations of response ("this compound produced withdrawal and depression in rhesus monkeys").

When the various needs were related to the characteristics of the data, these various requirements were formulated: (1) The system must be flexible, and able to accept data from many disciplines in several forms. (2) These data must be arranged or indexed so that correlations are possible within and between disciplines. (3) The system must be economical as to the time required for the coding of data, and with regard to the physical means employed.

The first step was to devise a general classification of the total body of information so that it might be indexed. The disciplines involved were studied, and seven general test areas were established (Table I). Four of these were pharmacological. The remaining areas reflected the discipline directly (numbers 5 and 7), or the intent of the study (number 6).

Table I. Classification of Subject Matter

No.	General Test Area
1	Toxicology
2	Central Nervous System
3	Muscle
4	Cardiovascular
5	Microbiological
6	Metabolism
7	Endocrine

Table II. Test Classification

400 Cardiovascular

400A Cardiovascular Screening–(B. P.-Biogenic Amine Response)
400B Cardiovascular Screening–(Induced Hypertension)
401A Contractile Force–*in Vivo*
401B Contractile Force–*in Vitro*

*Presented before the Division of Chemical Literature, American Chemical Society, St. Louis, Mo., March, 1961.

**Scientific Information Section, Wyeth Laboratories, Philadelphia, Pa.

Tests within these general areas were assigned three-digit numbers. The alphabetic modifiers indicated the cases in which different techniques are available to test a particular response. Table II shows two procedures, either or both of which may be employed in studying similar cardiovascular effects.

A single test may return data on more than one testing parameter, as shown in Table III. In this case decimal notations were assigned to each unit of data returned. Thus, in test 400A the blood pressure and its response to pressor amines are indicated.

Table III. Test Classification

400A. Cardiovascular Screening (B.P.-Biogenic Amine Response)
400A. 1 Blood Pressure *per se*
400A. 2 Blood Pressure *vs.* epinephrine
400A. 3 Blood Pressure *vs.* acetylcholine
400A. 4 Blood Pressure *vs.* histamine

When all tests were indexed they were seen to represent three levels of complexity. Some showed the general properties of the compounds; for example, the toxicity and cardiovascular activity or some other area of current company interest, such as analgesic efficiency. This type of test was called a screen. Other tests directly confirmed the screening data, or "evaluated" these properties; and the third type provided highly complex, specific information, or "evaluations in depth." To describe this stratification a subclassification based on the level of complexity was devised. This was accomplished by addition of modifying terms, as "screens," "evaluations" and "depth evaluations" (Table IV).

Table IV. Test Classification (Complexity)

400A. Cardiovascular SCREEN
401A. Contractile Force *in vivo* EVALUATION
404A. Peripheral Blood Flow DEPTH EVALUATION

Complexity levels were assigned in accordance with these criteria. A screen should return data of a general nature or area of company interest. It should, in addition, be produced by a relatively simple technique that is economical as to the amount of time required for completion. Finally, a screen must return numerical data.

An evaluation was required to provide direct confirmation of a screen or screening parameter. The data, however, could be expressed in numerical or judgment form. Depth evaluations by definition produced highly specific data, usually by a complex technique. The data in this case were acceptable in numerical, judgment or narrative form.

On the basis of the criteria for complexity levels, 14 tests were assigned to the *screen* category, 28 to *evaluations*, and the remaining 127 to *depth evalutions*. The 14 screens returned data on 36 testing parameters. However, 8 of these were concerned with the toxicity of the compound, and were assigned no corresponding evaluation test. By definition, each *screen parameter* required an evaluation test, hence there are 28 evaluation parameters. The depth evaluations contained 300 parameters.

Forty-two (25 per cent) of the total of 169 tests available were *screens* or *evaluations*, yet this percentage produced 462 of the 522 tests recorded, or almost 90 per cent. The code, therefore, was required to deal with these types in the greatest detail, whereas the depth evaluations could be recorded in a more general fashion.

At this point it was necessary to determine the physical form of the retrieval system, since the data-recording arrangements indicated for center-punched, machine-handled cards were different from those required for manually operated systems. A major factor governing this choice lay in the extent of the collection of chemical compounds. This amounted to less than 10,000 substances, so that manual searching was feasible. However, the collection was of sufficient size to recommend some form of mechanical assistance. These factors led to consideration of an edge-notched type of card. Since much of our "backlog" of information occurred in a variety of physical forms (memos, cards, physiological tracings, reports), the format of the edge-notched card also provided a standard readable unit record for such data.

Since data could originate from a large number of tests, a card with maximum punching space was preferred. Such a card is available from E-Z Sort, Ltd. It measures 5 × 8 inches, and is punched on the periphery four holes deep, affording more than 500 direct punches.

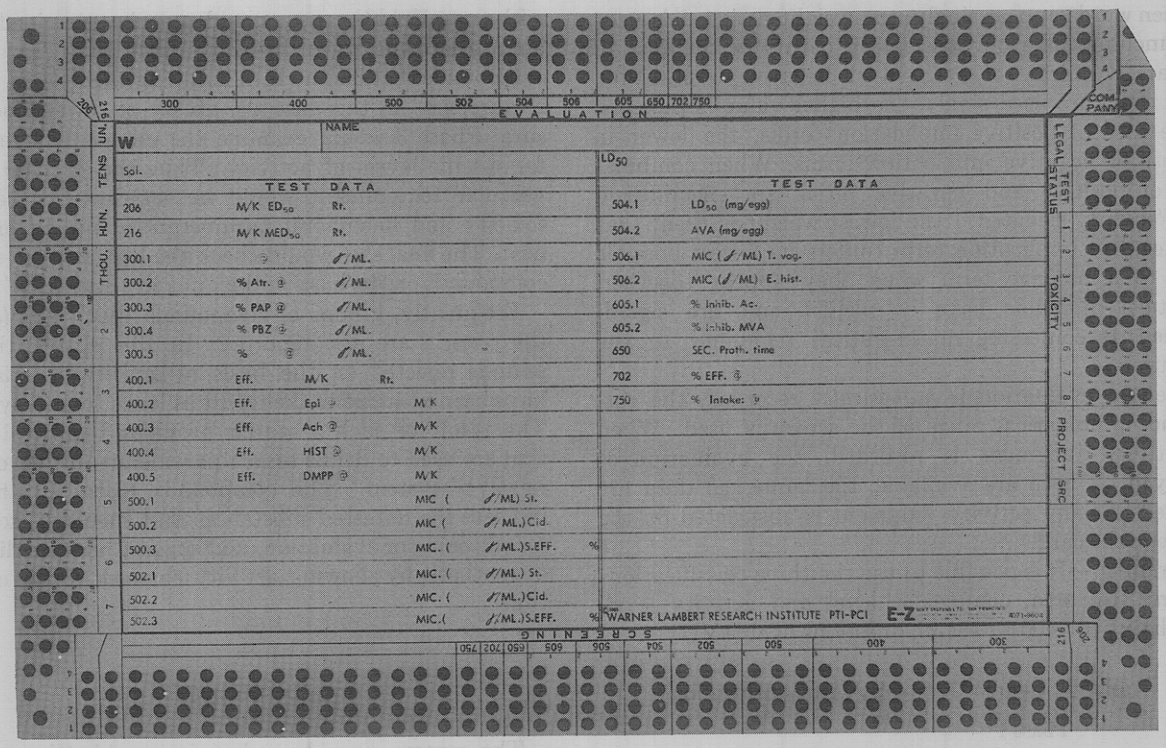
Convenient abbreviations for the screening tests were provided, giving a prewritten abstract of the parameter tested. Appropriate numerical data were then entered. Short abstracts of data dealing with evaluation and depth evaluation were placed on the reverse of the card, with care to keep such entries in test number order.

Field Location.—In the card illustrated, the screen punching field is placed at the *bottom*. Inasmuch as this is the field of heaviest punching, hence of greatest fragility, it is protected from the further abuse of being passed repeatedly through typewriter rollers when test data are entered.

Card Design.—On the card, the screening and evaluation fields are identical, as dictated by the presence of an evaluation test for each screening parameter. The depth evaluation area of the card provides punching locations for all tests of this type performed in each general area, with sufficient locations reserved for expansion (Fig. 1). The other areas of the card are self-explanatory with one exception. The toxicity field (general test area number 1) records oral LD_{50} levels. Other columns in this field describe the symptoms observed in the course of the toxicity determination. Each punch position has a specific symptom equivalent. Each row in this field deals with a pharmacological classification, *e.g.*, symptoms of central nervous system depression, stimulation, and the like.

Depth evaluations are assigned punch locations in the appropriate field of the general test area. An abstract of the results of this type of test is placed on the reverse of the card as numerical, activity, judgment or descriptive data. When more than one depth evaluation is performed in the same area, trailer cards are used for that particular test, with the abstract entered on the trailer card. This is due to punching convention, which permits a single number per field. Since depth evaluations represent only 10 per cent. of testing, trailer cards have increased the deck in that proportion.

Since the screen tests were required to return numerical data only, each of the four holes per row was used to express some *range* of the values on each parameter. These



ranges were assigned by an investigator experienced in the particular test technique and grouped with reference to the activities of standard substances. In the analgesic screen shown in Table V, the unit is the ED_{50}, that is, the dose effective in 50 per cent. of the animals tested. The useful range of activity is set at less than 400 mg. per kilogram, so that doses greater than that value are placed in range 1.

Table V. Data Classification

Punch depth	Screen range	Evaluation term
1	ED_{50} > 400 mg./kg. PO	Inactive
2	ED_{50} 200–399	Slight
3	ED_{50} 26–199	Moderate
4	ED_{50} <26	High

Table VI. Adjective Analysis

Negative Implication		Positive Implication	
No.	Weighted (punch) value	A. Superlative	Weighted (punch) value
Inactive	1	*High*	
Inert		Excellent Good	4
Imperceptible		Optimal Maximal Remarkable Outstanding Ideal Superior	
A. Comparative			
Curtailed Decreased Depressed		B. Equivalent	
Diminished Fair Lessened Little Lowered Limited		Equal Similar Uniform Alike Same	3–4
Mildly		Analogous	(Relative to standard)
Minute Perceptible	2–3	Congruent Identical Standard	
Poor Reduced *Slight*		C. Utility	
Trace Weak		Practical Satisfactory Interesting Desirable	3
Statistical Modifiers		D. Comparative	
Significant Variable Valid Reliable Average		Added Ample Amplified Augmented Conspicuous	Enlarged Increased Intensified Magnified *Moderate* 3–2
		Considerable Effective Elevated Enhanced	Raised Strong Superior

Range 2 is active at 200 to 339 mg. per kilogram; and range 3, at 26 to 199 mg. per kilogram. This range includes the value for aspirin, which by this test is 150 mg. per kilogram. The fourth range, established at less than 26 mg. per kilogram, includes codeine at an ED_{50} of 10 mg. per kilogram. Thus aspirin-like and codeine-like compounds may be differentiated.

Evaluations are coded in much the same manner, except that provision is made to receive data as judgments of activity level. Each evaluation punching range is assigned an evaluation term in addition to the numerical ranges. Such terms were selected as corresponding to the language used in the reports and data analyzed in our study of 500 compounds. A list of weighted synonyms has been prepared to cover variants (Table VI).

When words appear singly, the weight assigned is used for punching. The only exception occurs in the case of comparative terms of positive or negative implication. When these are used singly, the higher value is punched in the case of positive implication terms, the lower in the case of negative implication terms. When combinations of positive and negative terms are encountered (*e.g.*, "slightly elevated") the lowest weight is used; that is, a negative implication term confers its lowest possible value to a positive term when they appear together. Statistical modifiers have no weight and are signal-punched as conferring mathematical reliability to the estimation.

When an evaluation is *numerically* recorded, the corresponding numerical range of the screen is used. When an activity judgment is recorded, the evaluation is employed. If both are available, the numerical data are punched but any activity judgment is abstracted on the reverse of the card.

Searching.—It is possible to use the three classifications (general test area, complexity level and activity level) in outlining correlative search questions (Table VII).

Table VII. Search Outline by Test Classification

Phase 1
General Test Area (1–7)

Toxicology — Microbiology
CNS — Metabolic
Muscle — Endocrine
C-V

↓

Phase 2
Complexity Level (1–3)
Screen
Evaluation
Depth Evaluation

↓

Phase 3
Activity Level (1–4)
Inactive
Slight
Moderate
High

↓

Search Result

An example of such a multiple correlative search might be the following:

Query.—"What compounds gave good activity in lowering blood pressure, and are not adrenolytic? They must be nontoxic ($LD_{50} > 500$ mg. per kilogram *per os*) when given orally and have no effects on the central nervous system. Only novel compounds should be reported."

The first phase of the search outline would examine the general test areas, to select those compounds that *have* tests in the areas of the search. In answering this query, we would select the cardiovascular area, number 4, and the toxicology area, number 1. In phase 2, compounds with tests in both areas would then be examined for the appropriate complexity level, enabling us to discard those compounds that had no tests fulfilling the desired primary requirement, in this case the ability to lower blood pressure. Phase 3 would examine the compounds remaining for specific levels of activity. Thus the primary cardiovascular requirement would be satisfied first, and the toxicity and behavioral parameters would be examined next. The final sort would check the legal field to separate compounds on the basis of their novelty status.

Frequently, groups of compounds that *partially* fulfill the criteria are *also* reported. In this search the compounds rejected on the basis of *toxicity* criteria might have been included in the report as leads to new syntheses. The result of such a search provides biological models that are used to define search parameters for our chemical retrieval system.[1] The compounds recovered then will indicate the untested substances that should be scheduled for screening or evaluation, and suggest possible directions of synthesis by showing activity changes versus chemical type.

DISCUSSION

The system provides what is considered a realistic solution of our particular problem. For our purpose it enables flexibility of input, analysis, correlation and retrieval of data. Especially useful properties are the ability to utilize information in other than numerical form, the wide spectrum of testing recorded, and the availability of readable data.

The encoding of data in numerical form by assignment to one of four ranges requires sharp breaking points at the upper and lower end of each range. Thus, compounds may be placed in different ranges when the difference amounts to only 1 unit of measurement. This is an admitted source of "noise" in the system, which can be surmounted by understanding the specificity of each search presented by the questioner. In the event of a "pin-pointing" requirement, hand sorting of a particular range may be performed.

Our three basic requirements—flexibility for multidisciplinary information, correlation within and between disciplines, and economy of coding time and physical form—were met by the techniques described.

With regard to the susceptibility of this system to machine handling, we recognized at the outset that manual operation of a search system would become less efficient as the size of the compound collection increased. A search outline such as we have presented in a multidisciplinary area presently requires twenty to thirty minutes for completion. The generation of the final report may require two to three hours. The print-out capabilities of machine procedures are obviously desirable, as well as the search speed available as the compound collection expands.

We are now contemplating a code that will transfer the results on screening tests *only* to center-punched cards. The general approach will be the same except that the ranges will be revised in accordance with the 9 hole depth available on tabulating cards. Eighty columns easily handle the 28 screen parameters and permit print-out of compound number, LD_{50} and some other frequently employed parameters. The remaining test data will print

out as the appropriate range number, which can be decoded by reference to legends provided. This is one approach. However, codes covering all three levels of complexity (not just the screen) also can be used, setting up decks reflecting test information by the discipline or general test area concerned.

On the basis of the analysis of test performance by complexity levels we suspect that decks of each complexity level may provide the most efficient search device, but our final decision must await a future report.

Machine systems, including those of Dunn,[2] Wood,[3] and Dietrich,[4] provided helpful approaches to our problem. A particularly useful technique was Welt's,[5] segmentation of a problem into logical parts. Additional information relative to the use of "terms" as units of biological activity judgment, was obtained from the "deep" and "shallow" indexing approaches of Whaley.[6]

Acknowledgment.—The authors wish to acknowledge the contributions of Mr. Eliot Steinberg, Director of Research Administration, Mrs. F. Arendell, and Mrs. E. J. Pugh, Technical Information Section, to the development of the system and the present report.

BIBLIOGRAPHY

(1) Arendell, F. H., A Three-Symbol Code for Searching Chemical Structures, *J. Chem. Doc.*, 1, 47 (November, 1961).
(2) Dunn, E.E., and Lynn, G.E., Reporting and Indexing Biological Data by IBM Punched Card Methods, Presented at American Chemical Society Meeting, March, 1952.
(3) Wood, G. C., "Biological Subject-Indexing and Information Retrieval by Means of Punched Cards," *Special Libraries*, **47**, 26 (1956).
(4) Dietrich, E. V., "Machine Retrieval of Pharmacological Data," *Science*, **132**, 1556 (25 Nov. 1960).
(5) Welt, I. D., "The Detailed Indexing of Biological Effects of Chemical Substances," *Bulletin of the Medical Library Association*, **46**, (No. 3), 367 (1958).
(6) Whaley, F. R., "Operational Experiences with Linde's Indexing and Retrieval System," Conference on Information Retrieval, IBM, Poughkeepsie, N. Y., September, 1959.

LETTER TO THE EDITOR

Dear Sir:

In the article entitled "The Parke-Davis Code for Chemical Structures" by Geer, *et al.*, appearing in the April issue of the *Journal of Chemical Documentation* at page 110, reference is made in the fifth paragraph to "ozalid masters" and to "ozalid coated IBM cards."

We call to your attention that "Ozalid" is a registered trademark of General Aniline & Film Corporation for merchandise including light-sensitive copying papers, cloths and films, and machines for developing diazotype prints. Copies of our registrations Nos. 252,339, 358,594 and 393,828 are enclosed for your information.

Use of this mark written in lower case form as it appears in the article mentioned above fails to indicate its trademark status and gives the impression that the term is a generic one for the products for which it is used.

Continued use in this manner would destroy our rights in the mark. The mark should be used to refer only to merchandise supplied by our company and should be at least capitalized as has been done, for example, in another article on page 114—second column.

We assume misuse of our mark in the article on page 110 was inadvertent, and should appreciate your confirmation that you will avoid such misuse in the future.

In passing, we note also reference on page 110 to the "Ozalid Company." The correct designation is Ozalid Division of General Aniline & Film Corporation.

General Aniline & Film Corp.
Patent Department
435 Hudson St., New York 14, N. Y.

Walter G. Hensel