word matches. Implementation of this technique would necessitate extensive reprogramming of our system. Eventually, if we need to put CBAC on a cost-recovery basis, we will pursue one of these two possibilities.

Few subscribers rely totally on CBAC for all literature searching. Most use it primarily to gain coverage of peripheral and/or inaccessible journals. Virtually all users continue to subscribe to a few favorite journals for browsing and review *Chemical Abstracts* from time to time. One subscriber supplements CBAC coverage by perusal of *Current Contents*. It is doubtful that computerized current awareness searching would ever replace journal browsing, even if coverage of literature were complete and thorough for every journal.

FUTURE PLANS

In addition to investigation of more economical programming techniques for services which appeal to small numbers of subscribers, consideration of several other aspects of our system has high priority. These aspects include: expansion of our general search strategy to provide for Boolean logic, infixes, and suffixes; retrospective searching; revision of feedback techniques to include more extensive analysis of results at less frequent intervals, rather than simple analysis of the interest level of the output from every issue. We also hope to be able to provide more personal attention to sharpening established profiles in order to trim operating expenses.

CONCLUSIONS

Our four years' experience with CBAC has provided invaluable information in two respects. First, we learned a great deal about the literature-searching needs and habits of our research community by observing their reactions to CBAC's thorough coverage of a relatively small body of information. Secondly, we were exposed to the practical and economical problems of computerized free-text abstract searching by file inversion. These considerations have contributed to the over-all development of our current awareness system and will continue to help us define its future direction.

LITERATURE CITED

(1) Bond, Lynn, Carlos M. Bowman, and Dolores Hartman, "User Reaction to Three New Services Offered by Chemical Abstracts Service," Division of Chemical Literature, 152nd Meeting, ACS, New York, Sept. 1966.

(2) Brown, Marilyn T., "A Computerized Current Awareness System Using Chemical Abstracts Tape Services," 59th Annual Conference of the Special Libraries Association, Los Angeles, Calif., June 1968.

A Multi-Level Retrieval System

II. Medium-Sized Collections*

LEE N. STARKER, KATHERINE CRAWFORD OWEN, and BETTY COOPER BATSON
Warner-Lambert Research Institute, Morris Plains, N. J.

Received March 17, 1969

Retrieval systems based on the IBM-type variable field visual collation card can be converted to Termatrex systems as the collections grow in size and use. The conversion is accomplished by computer-generation of a corresponding term/document/IBM card deck which is used to drill a set of Termatrex cards with the J-400 Termatrex drill. The term/document cards can then be used in the preparation of a number of subsidiary search tools. Sample index work sheets and the input procedures are also discussed.

In the first paper of this series,[1] we described a visual collation or "peek-a-boo" approach to the control of relatively small collections of documents. This system was based on the common IBM card and was characterized by the fact that it could be handled readily by the ultimate user on a total do-it-yourself basis. It also could be operated with partial or total assistance from a central information group.

In our experience with the IBM card-oriented peek-a-boo technique, we found that several of the retrieval systems developed around it rapidly outgrew this particular device. The reasons varied from the large number of documents involved to the increasing number of searches required. Thus, a system for the indexing and retrieval of literature in the cosmetic field quickly passed the 6000 document mark (requiring 12 decks of peek-a-boo cards), while our library's use of the peek-a-boo cards, with a collection of 1500 papers (three decks) on Coly-Mycin, generated an increasing volume of searches. In the first case, the need to carry a relatively small number of searches through 12 decks of cards, and in the second instance, the requirement for a larger number of searches through three decks of cards, began to impose a burden on those users directly involved. The search procedures began to consume more time than was thought to be appropriate.

We did not feel that either of these situations warranted computerization at the time. Instead, we turned our considerations toward the Termatrex[2] approach. This

* Presented in part before the Third Middle Atlantic Regional Meeting, ACS, Philadelphia, Pa., February 1968, and in part before the 57th Annual Convention of the Special Libraries Association, Minneapolis, Minn., June 1966.

appeared to be a particularly desirable method, since the philosophy and techniques of input and retrieval were the same as those already in use. Most important, the application of the 10,000-document Termatrex card to our problems would give us the capability of running only one search per query through the files. Even if the larger of the two systems ran over 10,000 documents, it seemed likely that the use of two Termatrex decks, to give a capacity of 20,000 documents, would serve the Cosmetics group well for the foreseeable future.

We were also attracted to this technique since the user would not be dependent on card sorters or any other type of data-processing equipment for purposes of retrieval. He would still retain the ability to carry out his searches at his desk, and as he wished. He would, however, now have to rely on the Information Center to input the data for him.

CONVERSION TO TERMATREX CARDS

A J-400 Termatrex drill had been installed in the Information Center some time before for a completely different application, but time was available for additional uses. This drill (Figure 1) is a very sophisticated piece of equipment that can either be actuated by an adding machine type of keyboard or be tied to a card reader, such as the IBM 514 reproducer. The document addresses are then drilled under electronic control from the IBM punch cards as they are read by the reproducer.

Because of the volume of information that we wished to transfer from the peek-a-boo cards to the Termatrex cards, the ability of the J-400 to operate from punched cards became a very valuable attribute. We did not want to enter all of this information manually, nor did we wish to repunch all of the information from the multipunched peek-a-boo cards into the "term per document per card" format that was necessary to make use of the automatic features of the J-400.

Our problem was solved with the realization that, because of the logical pattern with which document numbers were entered on the peek-a-boo cards, a relatively simple computer program could be used to translate each peek-a-boo punch to its corresponding four-digit document number, and thus generate a deck of cards in the appropriate format for input to the Termatrex.

Figure 1. J-400 Termatrex drill

Two simple expressions were sufficient to define the conversion rules needed to translate the peek-a-boo punches to four-digit document numbers. Thus, if C = the column number, R = the row number, and D = the punch in column 80, then for any punch in the body of the card (cc 1–50, rows 0–9):

$$\text{Document number} = (10C + R) + 500\ (D - 1)$$

Applying this formula, a punch in cc 2, row 3, is translated to 0023 for deck 1 (cc 80/1), to 0523 for deck 2 (cc 80/12), and 2023 for deck 5 (cc 80/5). Similarly, a punch in cc 48, row 6, translates to 0486 for deck 1 (cc 80/1), 0986 for deck 2 (cc 80/2), and 3486 for deck 7 (cc 80/7).

Punches in the 11-row of cc 1–9 were translated by the expression

$$\text{Document number} = C + 500\ (D - 1)$$

Application to a card punched in row 11 of cc 3 gives document number 0003 for deck 1, 0503 for deck 2, 2003 for deck 5, etc.

When a deck of peek-a-boo cards was submitted to this program, the net result was the generation of a new deck of cards equal in number to the holes punched in the peek-a-boo deck. Each new card carried only one document number and the term name to which it was related (Figure 2), and was ready for preparation of the Termatrex deck.

The new term deck (Figure 3) was sorted into alphabetic order by term, after which a sequence number was punched in, so that future alphabetic sorts could be carried out on a five-digit numerical field rather than on the actual term field. Each group of term cards was then placed in the punch station of the 514 reproducer, and the corresponding Termatrex card was placed in the drill. As the document number of each card was read, the information was transmitted to the drill, which was then directed to the correct position for entering the document number in the Termatrex card. Using this equipment, it is possible to carry out up to 60 drilling operations per minute—a substantial increase over the manual methods.

After each group of term cards had been fed through the system, the next group was added, the Termatrex card was replaced with a new one, and the operation was repeated. At the conclusion of this sequence, we had a completely drilled set of Termatrex cards ready for searching.

It readily became evident that the IBM term cards which were used for input to the Termatrex could be used in several auxiliary ways. This was particularly true with the library's Coly-Mycin index. This file consists of a deep index (20 to 30 terms per paper) to a collection of approximately 2500 papers dealing with the drug Coly-Mycin. Because of the depth of the index, it proved valuable for searching out those documents that dealt with relatively subtle pieces of information, as well as for those questions that required the correlation of large numbers of terms.

Numbers of questions could be answered by referral to no more than two or perhaps three terms, and the

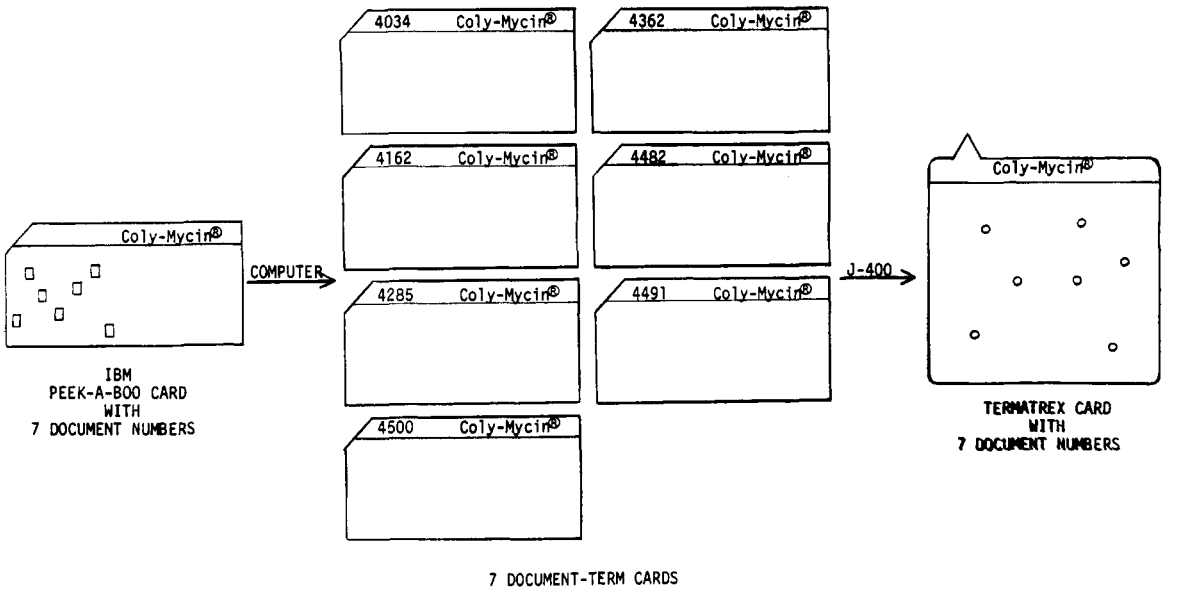

Figure 2. Peek-a-boo to term/document/card conversion

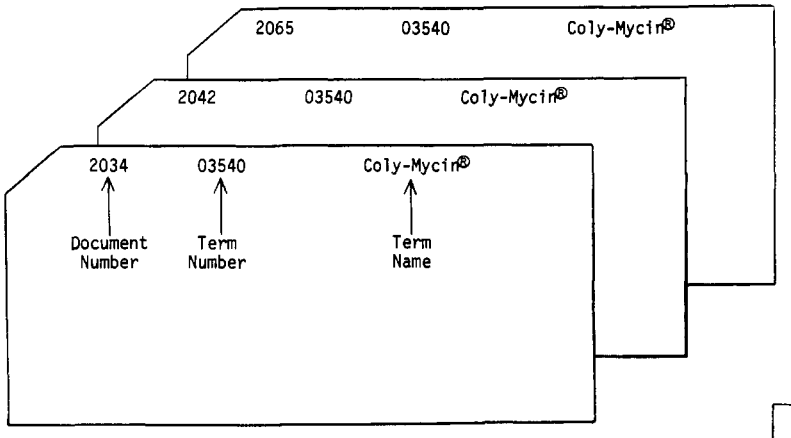

Figure 3. Term card layout

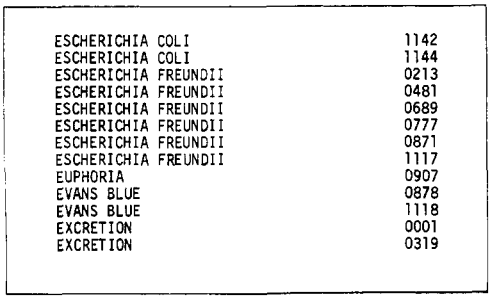

Figure 4. Term order printout

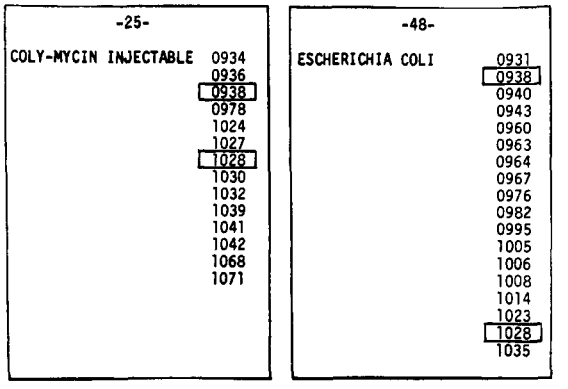

Figure 5. Double book dictionary

greater sophistication of the Termatrex deck was not always needed. If a simple search tool were available for desktop use, then these questions could be answered immediately by the individual scientist without need to refer to the library. Accordingly, the IBM card deck used to prepare the Termatrex file was sorted in term order and listed to produce an index of all terms with associated document numbers (Figure 4).

To further ease this kind of searching, we made two copies of the printout available to provide the convenience of a "double book" dictionary (Figure 5). These very simple tools quickly expanded the usefulness of the product turned out by the library's indexing effort, and resulted in a decline in the number of requests made of the central facility. These requests now, however, tend to be of a more complex nature, and can best—or can only—be searched via the Termatrex deck.

We next resorted the IBM term cards in document number order to yield a printout in which each document number was listed with all of the terms by which it had been indexed (Figure 6). With this listing, a search can be run, and the document numbers which appear as answers can then be further evaluated by consulting the document order printout. Because this listing does contain all terms associated with a document, it can often be of help in determining the relevance or lack of relevance of an answer, without having to locate the actual paper.

As a further fringe benefit, we have found that the depth of indexing of the Coly-Mycin file yields an unstructured abstract of the paper that has permitted us to do away with all formal abstracting. We now substitute the list of index terms in our *Abstract Bulletin*, with a significant saving of time and effort (Figure 7).

As a result of the increased speed with which searches of the journal literature can now be carried out, we are beginning to note a steady increase in the number of requests that we receive. This number is still not overly time-consuming, but we are finding that the maintenance problems involved in handling the card decks for the various printouts are becoming burdensome. We are, therefore, now involved in a systems study to develop a computerized approach to this problem. We anticipate that we will be able to make use of the same deck of term cards which was used to generate the Termatrex deck and the associated printouts, thereby avoiding the necessity to keyboard the data again. We would also plan to merge these term cards with a deck of cards carrying full bibliographic data on this file, and thus increase the versatility of the resultant product.

When this new system has been fully developed and made operational, it will be the subject of a further communication in this series. This should not mean the end of the Termatrex system for these applications. We still visualize many searches being run on a preliminary basis via this file, or being run only through the Termatrex deck where the fallout is relatively small, or where the logic may be handled more easily in this manner.

GERMAN	0904
REVIEW	0904
TOXICITY	0904
YEAR 1964	0904
BLOOD LEVELS	0907
BRUCELLA	0907
CEREBROSPINAL FLUID LEVELS	0907
CHEMISTRY	0907
COLIFORM	0907
COMPARISON	0907
COMPETITIVE	0907
CONTRAINDICATIONS	0907
DAYS TREATED	0907
DERMATITIS/TOX./	0907

Figure 6. Printout of document number with terms

UPDATING THE FILE

The indexing procedure which precedes the entry of new information into the files has been developed so that it is an integral part of the input to the retrieval system. This has been done by developing an indexing work sheet

COLISTIN CO 1942

Gillot, F. and Levilain, J.-C. (C.H.U. de Nantes): (Clinical tests on Albacyclin in pediatrics.) Gaz. Med. France 74:3581-4, June 25, 1967.

INDEX TERMS:
ORIGINAL DATA
CASE HISTORIES
INFANTS 1 mo.- 2 yr.
PATIENTS- 1
PEDIATRICS
DIARRHEA
OTORHINOLARYNGOLOGY
PHARYNX
NAUSEA-VOMITING
NOVOBIOCIN
TETRACYCLINE

68-310 (LE)

COLISTIN CO 1943

Hungerland H. (Bonn): (Treatment of pyelonephritis in children.) Acta Paediat. Belg. 21:145-60, 1967.

INDEX TERMS:
REVIEW
PEDIATRICS
PYELONEPHRITIS
INTRAMUSCULAR
EFFICACY
COLIFORM
ESCHERICHIA COLI
PSEUDOMONAS AERUGINOSA
CHLORAMPHENICOL
NITROFURANTOIN

68-311 (LE)

Figure 7. List of index terms in *Abstract Bulletin*

that can also be used as a source document for key-punching. Examples of these indexing forms are shown in Figure 8a, 8b, and 8c, and illustrate an increasing depth of vocabulary complexity that is directly related to the manner in which the retrieval file is used. These forms are used as source documents for the IBM cards that generate the Termatrex decks and the associated listings described earlier.

The "Coagulation Index" sheet shown in Figure 8a is a relatively shallow index that is used by one of our senior laboratory scientists. She scans and indexes all of the pertinent literature and carries out searches based on this body of data, both for herself and for associated research and sales groups. Since she is intimately involved with a relatively narrow segment of the literature, her prime need is for a retrieval system that will give her a series of generic "handles" to specific items of interest. This dictionary has approximately 180 terms, including a series of headings to indicate year of publication.

Because authors' names can also be important terms in this instance, we have added a four-letter author code. This code is made up of the first three letters of the author's last name, plus the initial of his first name—i.e., JOHNSON, LAWRENCE = JOHL. This information is input to the Termatrex system, using a unit-digit approach that devotes 26 A to Z cards for the first letter of the code, 26 A to Z cards for the second letter, etc. Thus, a total of 104 Termatrex cards is sufficient to file or retrieve (Figure 9) any possible four-letter code. False drops are possible because of code overlapping for multiple authors, but so far, they have not proved detrimental to the system.

COAGULATION INDEX WORKSHEET

Senior Author | Journal | Document No. (74-77)

Author Code (1-4) | Document No. (37-40) | 78 L | 79 | 80 T

✓	Term # (1-5)	Term Name	✓	(1-5)	Term Name	✓	(1-5)	Term Name
	001	Accelerator Factors		2250	Coag.-Microbiological Causes		4800	Prothrombin-Assay of
	050	Accelerator Factor V		2300	Coag.-Non-Specific Factors		4850	Prothrombin-Prothrombinogen
	100	Accelerator Factor VII		2350	Coag.-Pediatric		4875	PTT Assay
	150	Angina Pectoris		2400	Coag.-Screening Assay		4900	Reticuloendothelial System
	200	A/Coagulants		2450	Coag.-Vascular Effects		4950	Reviews and Books
	250	A/Coagulants-Circulating		2475	Dextrans		5000	Serotonin
	300	A/Coag.-Commercial Preparations		2480	EACA		5050	Species Specificity

Figure 8a

COSMETICS/TOILETRIES WORKSHEET

DOCUMENT NO._____

00020	ACNE	01120	ELECTROPHORESIS
00030	ADHESION/ADHESIVES	01140	EMULSIFYING AGENTS
00040	AEROSOLS-COSMETIC	01160	EMULSIONS - CHROMATIC
00060	AEROSOLS-GENERAL	01180	EMULSIONS - THEORY
00080	AEROSOLS-PHARMACEUTICAL	01200	EMULSIONS TRANSPARENT
00100	AGING-BIOLOGY	01220	ENCAPSULATION
00120	AIR FRESHNERS X ODOR	01240	ENZYMES
00140	ALLANTOIN	01260	EYES & EYE PRODUCTS
00160	ALLERGY	01280	FATS & OILS

Figure 8b

COLY-MYCIN® WORKSHEET

CO NUMBER_____

USE CHECK	TERM NO.	TERM	USE CHECK	TERM NO.	TERM	USE CHECK	TERM NO.	TERM
	SCHEDULE T - continued			050350	SHWARTZMAN PHENOMENON		BACTERIA - continued	
	020000	EUPHORIA		050950	SPEECH, SLURRED		007600	BORDETELLA BRONCHISEP.
	020300	FATALITY		054950	TOXICITY		007650	BORDETELLA PERTUSSIS
	020350	FATIGUE		055000	TOXICITY, ACUTE		008150	BRUCELLA
	020700	FEVER/TOX/		055050	TOXICITY, CHRONIC		010850	CHROMOBACTERIUM
	024050	GASTROINTESTINAL/TOX/		057550	URTICARIA		010900	CILLOPASTEURELLA
	024900	GRANULOCYTOPENIA		058000	VERTIGO/TOX/		011600	CLOSTRIDIUM BOTULIN.
	025650	HALLUCINATIONS/TOX/		058350	VISUAL/TOX/		011650	CLOSTRIDIUM BUTYRICUM
	025800	HEADACHE/TOX/		--	--		011700	CLOSTRIDIUM NOVYI
	026600	HEPATOTOXICITY		--	--		011750	CLOSTRIDIUM PERFRING.
	030250	IRRITATION/EYE/		SCHEDULE M - microbiology			011800	CLOSTRIDIUM SPOROG.
	030450	ITCHING		000900	ADDITIVE EFFECT		011850	CLOSTRIDIUM TETANI
	031400	LASSITUDE		001450	AGGLUTINATION		012250	COLIFORM

Figure 8c

Figure 8. Indexing worksheets

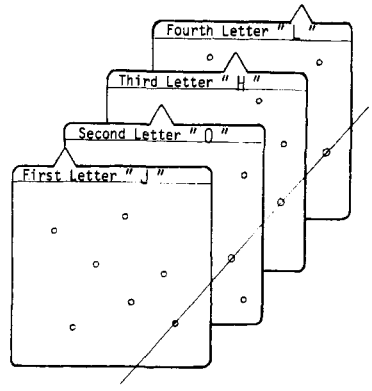

Figure 9. Retrieval of four-letter author code (JOHL)

One other device that has been adopted for this application is the generation of an IBM card for each paper indexed. This card carries the senior author's name, the journal citation, and the document number. These cards are sorted alphabetically by author, printed monthly, and cumulated quarterly. The listing is most useful when the indexer reviews secondary sources, since it permits a ready check on the possible previous inclusion of a given paper as an original publication.

A "Cosmetics and Toiletries" index sheet is shown in Figure 8b. Again, this is a completely self-contained operation where the laboratory scientist chooses, indexes, and subsequently retrieves his own papers. The index dictionary is, in this instance, somewhat larger because of a greater breadth of coverage, but not really deeper than the one just described. Searching may also be done, however, by other members of his group.

A portion of the work sheet used by the library staff for indexing their collection of papers on the antibiotic Coly-Mycin is shown in Figure 8c. This is by far the deepest and most complex index that we have attempted. It contains approximately 1200 terms, and the work sheet is printed on five double-sided pages.

Because of its length and complexity, however, the simple alphabetical or semi-classified arrangements shown in Figure 8a and 8b were inadequate, and this work sheet has been divided into a number of "schedules," each of which includes a number of related terms. Within each "schedule" the terms are arranged in alphabetical sequence. This allows for ready location of any term either for indexing purposes or for subsequent retrieval purposes.

The Coly-Mycin Index is also backed up by a complete thesaurus containing "see," "see also," "broad reference," "narrow reference," etc., designations. This system is used by our literature scientists in answering questions generated by a number of different research, clinical, and marketing areas.

In each of these applications the indexing term is always accompanied by a 4- to 6-digit number. This is called the "term number," and is assigned to each term so that the numerical sequence is the same as the alphabetical sequence. This serves two important functions.

First, the availability of the term number simplifies keypunching, since all that is necessary is to enter the term number and the document number into each punch card. Where the full term name is required for subsequent printout needs, this information can be gang-punched from a deck of master cards carrying both the term number and the term name.

The second use of the term number relates to the simplified approach to alphabetic sorting that was discussed earlier.

Index work sheets for all applications are now designed, using the general formats already shown. Except for the blood coagulation file, which requires entry of the bibliographic citation, the only requirement for the indexer is to assign and record a document number and then to check each term which relates to the paper.

The index sheets are transmitted to the Information Center in convenient batches, where the IBM cards are punched with the document number and term number. If the application requires printed term lists as well, the full term names are gang-punched into the card deck from a set of master cards.

The IBM cards are now ready for updating the Termatrex cards.

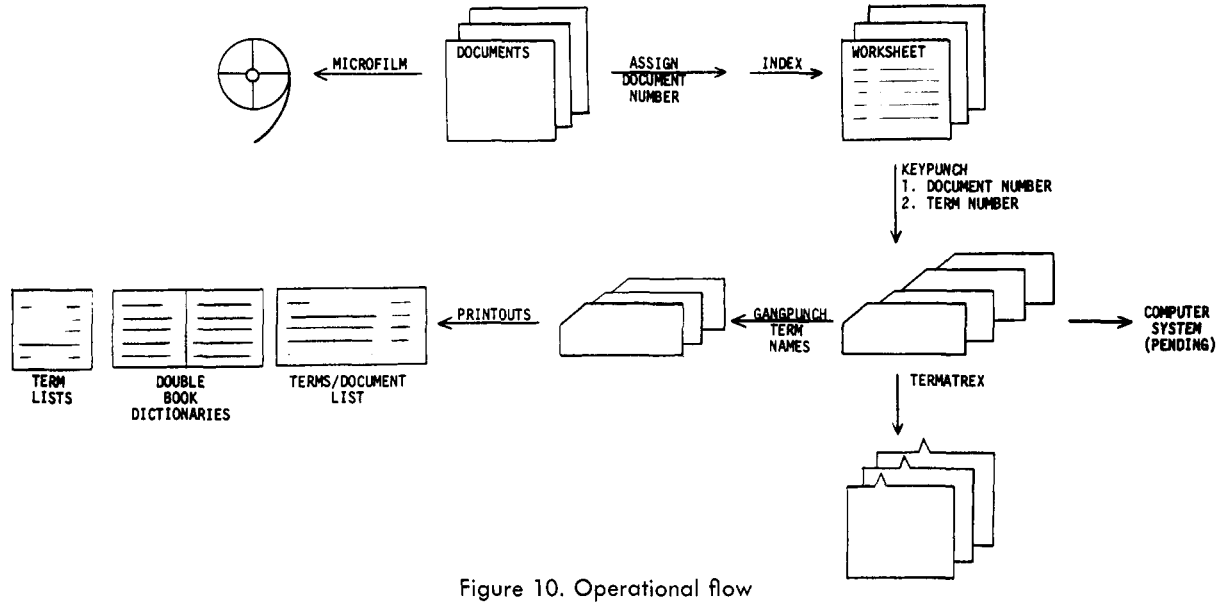

Figure 10. Operational flow

In a number of the satellite systems, we have also included microfilming of the documents after indexing. The film is used in either fiche or cartridge form.

The operational flow is outlined in Figure 10.

SUMMARY

We have described a flexible approach for the input and retrieval of information from medium-sized document collections. This system is based on the use of the Jonker J-400 Termatrex drill. IBM cards are used to drive the drill, as well as generate printed term and document lists. This system can be implemented directly from the visual collation cards described earlier, and makes for a logical extension of retrieval techniques for growing systems.

In addition, the system makes it possible to control individual literature collections with a minimum of assistance from the Information Center. Two such applications are described, as is the simplified index work sheet on which they are based.

Conversion of the Termatrex-based systems to computer manipulation for large document files is in process and will be described shortly.

LITERATURE CITED

(1) Starker, Lee N., and J. A. Cordero, "A Multi-Level Retrieval System. I. A Simple Optical Coincidence Card System," J. CHEM. DOC. 8, 81–5 (1968).
(2) Jonker Corp., Gaithersburg, Md.

The Reliability of Property Data, Or, Whose Guess Shall We Use?*

LEE J. KIEFFER
Joint Institute for Laboratory Astrophysics, National Bureau of Standards, Boulder, Colo.

Received April 14, 1969

Increased reliability of property data can be achieved by critical evaluation of the techniques used in making measurements. The necessity for and the implementation of such evaluations are explored.

The two titles I have chosen for this article, when juxtaposed, appear to be contradictory. The message I wish to convey is that the level of reliability claimed for most data compilations is based on the experimenter's estimate of his systematic and random errors, and that in almost all cases his estimate of the systematic errors is a guess. I hope to convince you that this is true, and then explore a possible way of overcoming this obstacle to reliability.

The Joint Institute for Laboratory Astrophysics Information Analysis Center is concerned with data on low energy atomic collisions—cross sections, rate coefficients, etc. What distinguishes an Information Analysis Center from an Information Center is that the former concentrates on critical evaluation of the information or data it deals with. Data are selected on the basis of quality judgments which have been clearly outlined and defended. In our own center's program we are trying to push this evaluation procedure one step further.

The word reliability will be used here in a strict sense. A short definition of reliable data would be data which are presented with error bars which were chosen so that the probability of the "true" value lying outside of these limits of error is extremely small.

The ultimate user of data is not interested in why the data may be unreliable. The penalty he pays for using an incorrect value is not alleviated by the fact that the error was only a clerical one. Preserving the integrity of the data identified and assuring the user that it was obtained using current "best" measurement techniques are absolutely necessary to achieve reliability. There is another step possible, and that is assuring the user that these current "best" measurement techniques are reliable—i.e., they can measure what they claim to measure to the accuracy claimed. The last criterion is the most difficult one to satisfy in this long chain.

This problem was noted a few years ago by W. J. Youden, at that time a consultant to the National Bureau of Standards, in a paper on "Systematic Errors in Physical Constants," published in *Physics Today*.[1] "When two laboratories make independent determinations, each may attach to its 'best' value a $\pm$ sign followed by an estimate X of the error. This estimate of the error is often based upon a series of observations made under carefully controlled conditions. Experimenters soon discovered that if laboratories A and B reported values C_A and C_B for the same constant, the difference Δ between C_A and C_B was almost always a large multiple of their estimated error. Obviously, these calculated errors had no more to do with the real errors than the neatness of the laboratory or the promptness with which the investigator answered his mail."

Determining the reliability of a measurement is not a simple problem. It is clear that progress has been made in making measurements more reliably, but such progress is not easy. The kind of effort which has been lavished on fundamental standards and on constants measurements is the kind of effort necessary to achieve reliability in the quantitative measurement of physical properties.

Every experiment which purports to measure quantitatively a property of a well-defined physical system is based on a theory. This theory usually consists of a number of mathematical expressions which explain how to connect the dial readings of the instruments to the value attached to the property. For completeness, include in this scheme whatever theory is necessary to connect the instrument readings with the fundamental standards

* Presented before the Chemical Documentation Sessions, 4th Middle Atlantic Regional Meeting, ACS, Washington, D. C., February 12–15, 1969.