mense volume of literature, with which he cannot cope. Some sort of "saturation" is to be expected soon and probably for some metals, the period 1968–1969 already marked this new tendency. In future years, we can probably expect only modest growth of the ANR for most metals or groups of metals.

4. Tables II and III show that the most spectacular growth rate was observed in the transition metal chemistry; in 1968, at least twice as many papers as in 1966 were published for most of the groups of the transition metals; unusual increases over a two-year period were observed also for thallium, arsenic, antimony, and bismuth. If a four-year period is considered, high growth rates of ANR are observed also for Mg, Hg, Si, Ge, Sn, and most of the transition metal groups.

We feel that such estimation of the growth tendencies in the literature can be useful in detecting the areas of most concentrated interest.

# Acronym Compilation by Computer (ACRODABA - ACROnym DAta BAse)*

PAULINE M. VAILLANCOURT**
School of Library and Information Science, State University of New York at Albany, Albany, N. Y. 12222

OLIVER H. BUCHANAN
Graduate School of Library and Information Science, Pratt Institute, Brooklyn, N. Y. 11205

Received March 1, 1972

**The complex and time-consuming task of preparing an acronym directory with fully documented entries for the library, information, and computer science fields was appropriately aided by methodology of these fields. The input record is constantly available for searching despite dispersal of the component 3 × 5 card records. The use of the data base as a teaching tool is interesting and meaningful for the students involved. The KWOC printout provides an index to significant words in the full text.**

Acronym compilation by computer gives continual integrity of records, permits dispersal of control cards for the gathering of additional data, and the resulting alphabetical and KWOC printouts optimize editorial surveillance of related terms.

Most directories of acronyms and initialisms give only the full text of the acronym. Often this information is not enough for the average reader. It was felt by the authors of the proposed "Directory of Acronyms and Initialisms in Library and Information Science" that additional information was required. To be able to look up and add to the main file and simultaneously separate the file into nine different categories to search for added information, it was decided to keypunch acronyms and full text in alphabetical order of acronym and to KWOC full text to terms on both IBM equipment (located at Pratt Institute) and UNIVAC equipment (located at S.U.N.Y. Albany).

Acronyms and initialisms in the literature, in speeches, and even in correspondence, present a serious problem to the neophyte, and, indeed, even to experienced individuals in the field.

As editor of the Science Associates International publication, "Scientific Information Notes" (now called "Information: News/Profiles/Sources"), the senior author received many news releases with acronyms sometimes explained in parenthesis, but surprisingly often unexplained with the assumption that all readers knew the terms for which the acronym stood.

A publisher who was interested in publishing a directory of acronyms for the library, information sciences, and computer fields approached the authors to consider compiling such a directory. The study of a large number of available acronym directories and dictionaries revealed that none gave adequate interpretive information, documentation, or references back to the original literature. For example, when you discover that CARES means "Central Advisory and Referral Service," what do you really know about that project if it happens to be unfamiliar to you?

What is needed is a directory that will tell the reader *where* the project is located (in this instance, New York Metropolitan Reference and Research Library Agency, which is also known as METRO) and hopefully what the project is about—rationale, procedure, output or results, methodology, and similar information. A complete and accurate bibliographic citation to a document would at least provide an answer to the source of this information, thus eliminating the need for knowing what reference tool might contain that information, and having to consult the tool that would provide the final answer to the reader's questions.

First, it was decided to include all acronyms in the strictly correct sense of the word. An acronym has been defined by Weik as "a *word* formed from the initial letter, letters, or syllables taken from a succession or group of words, and capable of being articulated .... An acronym may also be constructed from the initial letters plus the terminal letters of words ... [but] *must* be capable of being pronounced."[1] In addition, it was decided to include initialisms to broaden the scope of the book. These are usually included in most directories that purport to cover acronyms, although some compilers do not distinguish between initialisms and acronyms. An example of

*Presented at the 3rd Northeast Regional Meeting, ACS, Buffalo, N. Y., October 1971.

**To whom correspondence should be addressed.

a true acronym is POST-J for "POlymer Science and Technology Journal," while ACS for American Chemical Society and DCL for Division of Chemical Literature are examples of initialisms. For expediency, hereafter the term acronym will be understood to include both true acronyms and intialisms.

Since we were working with acronyms, we decided it would be appropriate to adopt one for our own project, calling it ACRODABA (ACROnym DAta BAse).

## INPUT TO THE SYSTEM

As shown in Figure 1, the authors searched the open literature to find acronyms which were then recorded on 3 × 5 cards. The source materials were primary journals, preprints, speeches, handouts at meetings, and proceedings; secondary literature, such as indexes and abstracts, directories, and dictionaries; and tertiary literature, such as reviews and handbooks.

The cards were kept in alphabetical order in duplicate decks by each author because of the geographical distance between the two schools. Cards had to be removed for the searching that was mandatory for the researching and verification operations. It therefore became advantageous to establish some mode of control over the total input to avoid inadvertent addition of duplicate cards for the same acronym.

For this purpose of control, acronyms and their full text were keypunched into IBM cards (by full text we mean the words for which the acronyms stand).

Each acronym record was tagged with an asterisk followed by a blank in card columns 1 and 2. The end-of-record was indicated by a dollar sign preceded by a blank. The acronym is punched into a free field beginning in card column 3 and is limited to 32 characters, ending in card column 34. The full text of the acronym is punched into a free field beginning with card column 35. If the full text of the acronym extends beyond 80 columns, it is carried on to a second card beginning in column 1, and may extend through column 78 followed by a blank and dollar sign. Each acronym record is limited to two input cards (160 characters) with a free-field maximum of

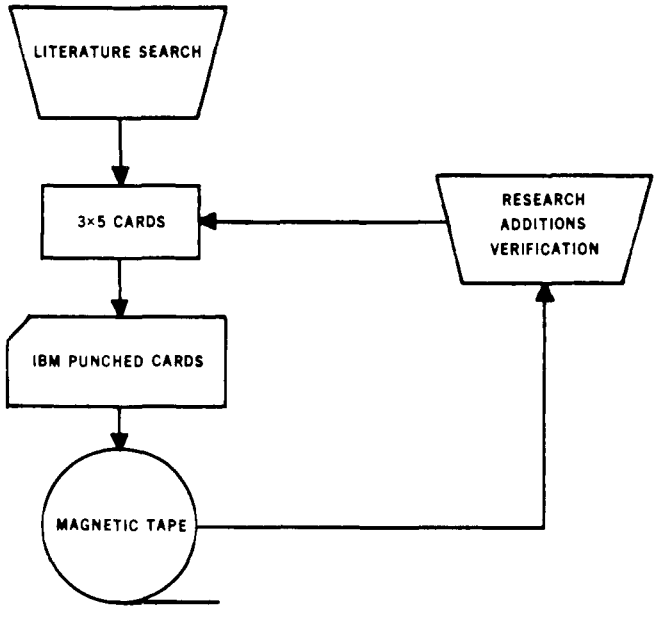


Figure 1. Flow chart of input data

| Card Number | Card Columns | Content | Comments |
|---|---|---|---|
| 1 | 1, 2 | tag | *β |
| | 3–34 | Acronym | Free field limited to 32 characters |
| | 35–80 | full text | Free field limited to 124 characters |
| 2 | 1–78 | " " | " " |
| | 79–80 | End-of-Record | β$ |

Figure 2. Punched card layout

32 characters for the acronym, and a free-field maximum of 124 characters for the field text as shown in Figure 2.

Since the longest acronym that had been found when the card was designed was less than 15 characters in length, the 32-character limit was considered a safe one. It was also felt that if all acronyms fell into the same pattern, the extra columns could be used for special coding purposes.

The data on the punched cards were transferred to magnetic tape. By means of an alphabetical sort program, the acronyms were rearranged, and an alphabetical listing was printed. The alphabetical printout allowed individual 3 × 5 cards to be removed from the working deck, while the printout served as a reference point for checking whether any given acronym was already in the data base. The 3 × 5 cards could then be sorted by category to permit similar questions to be looked up at the same time. This would eliminate backtracking during the searching operation. It also enables the searcher to become more familiar with the types and varieties of problems that arise within a given category. The system allows repeated experience to serve as a reinforcement which leads to the correct documentation more speedily and more accurately.

At one point, it was considered desirable to code the categories into each punched card. In any case, since this indexing, classifying, or categorizing is a mental activity, each entry would have to be examined and indexed. Consequently, there was little to be gained by machine coding, so it was decided to code manually.

The acronyms fell into nine separate categories, which determined the type of documentation needed to complete each entry. The categories are the following Table I):

**I. Association.** This includes any association, society, organization, committee, or commission—foreign and international. Mailing addresses are supplied for those having permanent headquarters. If the address rotates with the election of new officers, the reader is told to "contact the current president" or the current secretary, whichever is appropriate. It was not considered feasible to name the encumbent officer because this would become quickly outdated.

**II. Meetings.** This category is intended mainly for meetings sponsored jointly by groups or held infrequently, although any meeting known by an acronym is also included. These may be called conferences, symposia, colloquia, seminars, workshops, or meetings. The number and date of the last known meeting are listed, as well as the place held (if known), and sponsorship.

**III. Publications.** These are divided into books and journals. Complete bibliographic citations are given for both, but when the title of a journal does not contain the words "journal," "bulletin," or some other identifying term, the word "Journal" is supplied after the title. For books, in addition to title, the author or editor, place of publication, publisher, and date of publication are given.

Table I. Categories of Acronyms

| | |
|---|---|
| I. Associations | V. Terms |
| II. Meetings | VI. Library and Information Centers |
| III. Publications | VII. Project on Systems |
| A. Books | VIII. Commercial Firms |
| B. Journals | IX. Consortia |
| IV. Government Agencies | |

For journals, the date of volume one is followed by a dash only if the journal is still being published. If the journal is no longer published, the first and last years of publication are listed.

**IV. Government Agencies.** For agencies that are general and used by several countries or agencies that are unique to the United States, only the full text is given. Foreign agencies that are unique to a given country will have the designation for the particular country added. Those agencies that are not too well-known will have the next larger section or division given so that the reader may more readily identify it.

**V. Terms.** Only terms in library, information, and computer sciences will be included. Full text only will be supplied.

**VI. Library and Information Centers.** Full text and mailing address will be given.

**VII. Projects or Systems.** On-going or completed projects, studies, systems, computer programs, tapes, cooperative plans, and networks will be included. Full text, address (when known), and sponsor will be listed as well as a bibliographic citation, especially for those projects or systems whose complexity requires further explanation or whose purpose is not self-explanatory.

**VIII. Commercial Firms.** Publishers, consulting firms, hardware and/or software producers, and other such commercial organizations will be listed with mailing addresses.

**IX. Consortia.** Purpose, component members, and headquarters location will be included when known.

## PROCESSING

The hardware used for ACRODABA was a Philco 2000 (located at Pratt). The system was also designed for use on a UNIVAC 1108 (at SUNY Albany).

The software packages used were a Philco DOCSYS program designed by the Pratt Computer Education and Research Center, and a locally written alphabetical sort program. The original DOCSYS program was written for bibliographic citations, but was adapted to the ACRODABA project by substituting the acronym for the author-journal reference input while the full text of the acronym was used instead of the title of the journal or book. Thus the KWOC program gives an alphabetical listing of each word appearing in the full text of each acronym, the acronym, and the full text of the acronym. The original "trivial word" or "stop list" has been enriched to eliminate KWOC indexing on words considered insignificant in the ACRODABA project, plus certain insignificant non-English terms. Preliminary studies have been carried out on the effect of pre-coordination of selected acronym terms on the usefulness of the KWOC index.

## OUTPUT FROM THE SYSTEM

In addition to the alphabetical listing of acronyms with their full text, a second program is used to produce a Key-Word-Out-of-Context (KWOC) Index of all significant terms in the full text. This KWOC index will supply additional entry points to the main portion of the manuscript. It will be an index to key words in the full text.

The fully documented information on the 3 × 5 cards, when complete, will serve as the manuscript for the main body of the textual material.

## USE AS AN EDUCATIONAL TOOL

The building of a machine-readable data base of acronyms and the computer-manipulation of that data base is being used as a teaching tool in the Information Science curriculum at Pratt Institute. During the 1971 spring semester, 94 students were given the assignment of collecting acronyms and abbreviations pertaining to library, information, and computer sciences from the open literature. The acronyms were collected on 3 × 5 cards, together with their "full text" and literature source. The students were then taught how to key punch their acronyms into designated positions in IBM cards, and to submit their decks of cards for batch processing through the Computer Center Philco 2000 using pre-written alphabetical sort and KWOC indexing programs. Beginning students in Information Science are thus given the "laboratory experience" of the acquisition of information from the open literature, the preparation of input data to a live system, the use of a computer to prepare alphabetical and KWOC indexes of their data, and the results of the effect of variations and additions to a "stop list" (exclusion list) upon their KWOC indexes. This "hands on" experience appears to be a much more effective educational technique than simply listening to lectures or reading about KWIC and KWOC indexing.

## SUMMARY

That a need for such a bibliographic tool exists has been demonstrated by the keen interest of seven different publishing houses to produce the book. Holt Information Systems, the reference publishing division of Holt, Rinehart, & Winston, Inc. has been selected as the publisher and the book is expected to be released in 1973. The publisher will use 3 × 5 cards, KWOC printout, and magnetic tape as manuscript copy for the production of the book.

## ACKNOWLEDGMENT

The authors acknowledge the assistance of Ronald R. Fontana and his staff for writing the programs.

## LITERATURE CITED

(1) Weik, Martin H., "Standard Dictionary of Computers and Information Processing," p. 11, Hayden Book Co., Inc., New York, N.Y., 1969