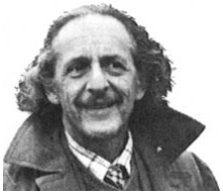


The Journal of
DOCUMENTATION

Volume 10 MARCH 1954 *Number 1*

THE PREPARATION OF SUBJECT-HEADING LISTS BY AUTOMATIC PUNCHED-CARD TECHNIQUES

by EUGENE GARFIELD

Grolier Society Library Fellow, School of Library Service, Columbia University, New York. Formerly of the Medical Indexing Project, Johns Hopkins University, Baltimore

Introduction

THE uses of punched-cards in documentation are already so varied as to confuse the potential user. Consequently, many organizations and individuals who could efficiently employ existing techniques are not aware of these possibilities. Even in the specialized problem of preparing subject-heading lists various approaches are in use. The purpose of this paper is to outline the uses of punched-cards in this special area.

The mechanization of bibliographical techniques has received considerable attention in recent years. There is a great deal of talk about robots, yet there are few if any 'mechanical brains' being used to solve bibliographical problems. In the realm of mathematical computation a great deal has been accomplished. What most people fail to realize or remember is that in applying 'robots' to various problems we are, essentially, merely speeding up activities we ordinarily handle without machines. Time is a most important feature in any operation—indeed were it not a question of time it is doubtful that we could call it an 'operation'. Consequently a giant computor does complicated mathematical problems by speeding up thousands of simple arithmetic operations. Without exploring further what computors actually do, it is necessary to state that, basically, in applying machines to problems in documentation we are again concerned with the time element. This is invariably the reason why large organizations are the first to require and use advanced machines. They are the first to face tasks of large magnitudes.

The Welch Medical Library Indexing Project at Johns Hopkins University is only one of several groups employing punched-cards to prepare subject-heading lists. The Library of Congress, Technical Information Division, prepares its subject-heading lists from such a file. The Subject-Heading

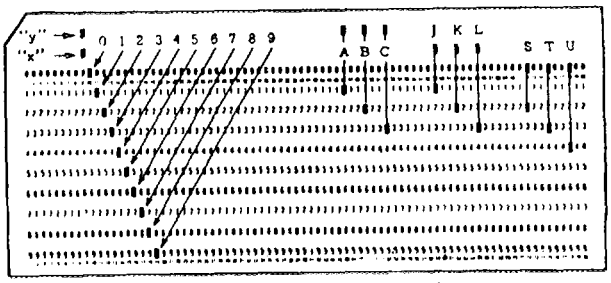
Authority List employed by the *Current List of Medical Literature*, published by the Armed Forces Medical Library, is also maintained on punched-cards.

The reader may already be actively engaged in preparing subject-heading lists, employing conventional clerical techniques. This usually consists of maintaining a file of 3 × 5-inch subject-heading cards or slips and then periodically typing a list. The latter task proves to be a stumbling-block sufficiently annoying so as to discourage the publication of current lists at frequent intervals. On the other hand, you may wish to begin the preparation of a subject-heading list without reference to an existing file. It is here that the sorting, filing, and editing abilities, as well as the printing features, of punched-card equipment can be used most efficaciously.

II

Before discussing the specific problems involved in these two cases it may prove useful briefly to explain what the punched-card is and how it is used in connexion with the various punched-card machines. The discussion here will make specific reference to the Hollerith (IBM) card, though similar principles may apply to the Remington Rand (Powers-Samas) card. Both of these systems are considered automatic in comparison with McBee or Zator edge-notched cards. The Hollerith punched-card is a paper card, $3\frac{1}{4} \times 7\frac{1}{2}$ inches, as is indicated in the accompanying figures. It is possible to write or print information on these cards as is done with the ordinary 3 × 5-inch catalogue or file card. It is when one adds holes to a card that it becomes a punched-card. Hence the Germans appropriately refer to them as *Lochkarte*, i.e. cards with holes in them. The French refer to punched-cards as *cartes perforeés*, i.e. perforated cards. Indeed, a punched-card is a card with holes that have been perforated. Hollerith cards require a special perforating machine called a key-punch. The key-punch is a glorified typewriter. It has the same keyboard found on standard typewriters except that it does not have a shift-key, since Hollerith machines are designed for upper-case letters only. However, there do exist a few specially prepared key-punches which do use every feature on the standard typewriter, including the shift, but the cards prepared on these machines cannot be used with standard Hollerith equipment. The reader will do best to think in terms of a machine using upper-case letters only. One operates the key-punch, then, exactly as one types on a typewriter. However, the depression of one of the keys does not activate a type-bar, but rather a set of punching-dies which perforate the card with a vertical pattern of holes. On Hollerith machines each hole is rectangular and about $\frac{1}{16} \times \frac{1}{8}$ inch in size. These patterns of holes are similar in purpose to the holes in a teletype tape or the patterns of holes in a Monotype tape. They may also be compared with the patterns on the old-style piano roll. Indeed, if one were to staple several punched-cards together lengthwise it might be possible to roll them up and feed the roll into such a piano. The resulting cacophony might be distressing, but not nearly as distressing as the difficulty presented in graphically explaining techniques which require actual observation for complete comprehension. It is not irrelevant to

mention that the reader who is faced with a documentation problem amenable to punched-card techniques should not be distressed with the details involved. Most of us can use typewriters without too much difficulty, but would find it extremely difficult to explain their operation. Once you have actually observed the preparation of a punched-card and its subsequent use in conjunction with sorters, collators, and printing units you will feel more at ease. As each key on the key-punch is depressed a pattern of holes appears on the card. The card feeds into the key-punch from right to left and the pattern of holes appears in a single vertical column. Since there are twelve rectangular punching-dies there are twelve corresponding punching or hole



Reproduced by courtesy of IBM

FIG. 1. Tabulating card showing 12 punching positions and combinations of punches to indicate letters.

positions. These are numbered from 1 to 12. However, the numbering does not begin at the uppermost position. The very first position is numbered 12, the next 11, and the third 0 or 10. The next nine punching positions are numbered from 1 to 9. The 12, 11, and 0 punching positions are the 'zone' positions. 12 is often designated as Y and 11 usually referred to as the X punch. These are not to be confused with the letters *x* and *y*, which have their own codes. When one depresses the A key, two punching-dies are electrically activated, namely the 12 or Y punch and the 1 punch. Similarly, depressing the B key activates the 12 or Y punch and the 2 punch. For each letter, then, there is a two-hole pattern or code. When one depresses the numerical keys only one punching-die is activated. Thus, pressing the 7 key activates the 7 punching-die which, as we have seen, is really the tenth of the twelve dies. The 9 key activates the last or twelfth punching-die. As one types words or phrases, vertical patterns of holes are made in the card which give it a laced appearance. As each key is depressed the card moves horizontally from right to left about $\frac{1}{8}$ inch. This distance constitutes a column, of which there are 80 on the standard Hollerith card. A smaller card is also used. Such is the case with the American Postal Money Order where a 51-column card is used. The size of the 80-column card was originally influenced by the size of the old American greenback, probably because of its contemplated use as a bank cheque. As each pattern of holes is perforated it is also possible to print the corresponding letter or number at the top of the

column, as is shown in Fig. 4. On older models this was done by combining a typewriter mechanism with the key-punch. Newer models use a more efficient printing unit.

One might reaonably ask why we prepare a punched-card. (Indeed, this question has been shown to be quite pertinent in situations where punched-cards were used in desperation with the hope that some magical solution might be forthcoming without clearly defining the problem at hand.) There is, nevertheless, very good reason to prepare a punched-card and even 'proof-read' it mechanically on the verifying machine. The verifier is a machine which will locate errors made by the operator in preparing the card.

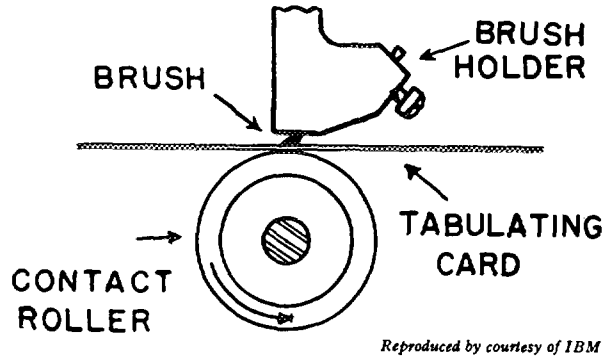


FIG. 2. Reading brush of punched-card machine.

The initial preparation of the punched-card and subsequent verification is essentially the last manual task performed. It is now possible to use the punched-card in conjunction with other Hollerith equipment.

People are often amazed that it is possible to alphabetize cards with a machine. The point is that we can only alphabetize punched-cards by machine. We are not yet alphabetizing library catalogue cards by machine because there is nothing on the card but the printed letters. Since we do not yet have a machine available which is capable of recognizing printed letters, the alphabetizing of catalogue cards in their present condition is impossible. Machines such as the sorter can recognize the punched-hole patterns we have perforated in the cards. At one end of the sorter is a mechanism with a little wire brush rubbing against a metallic roller. The punched-card is fed mechanically between the brush and roller. As soon as a hole appears in the card, contact is permitted between the brush and roller, completing an electrical circuit, which then energizes the armature of an electromagnet, which opens one of thirteen chutes for the conveyance of the card to one of thirteen pockets. As we have seen, if the letter A is coded 1, B coded, 2, &c., the alphabetic arrangement of the punched-cards can be achieved by sorting cards according to number. Indeed, numerical serializing is the key to all punched-card manipulations. Once we have sorted the cards it is then possible to place them in a printing unit which has a series of 80 brushes

enabling it to 'read' the entire card at once. However, instead of directing the cards to particular pockets, the punched-hole patterns activate a bank of type-bars or type-wheels, printing an entire subject-heading at a time at the speed of 150 lines per minute on the model 407 tabulator.

III

In the first case mentioned in the introduction the use of punched-cards is quite simple. It is only necessary to prepare a punched-card for each heading or cross-reference in the list corresponding to each 3 × 5-inch card. If a file of 10,000 cards is anticipated, with an average of 30 letters per heading, one can conservatively estimate that an average key-punch operator can prepare the file in 10,000 × 30/7,500 hours, i.e. about 40 hours. (Punch operators can achieve speeds up to 10,000 strokes per hour.) Once this initial work has been done, the time required to maintain the file is comparatively insignificant. It is definitely a part-time operation unless the file is rapidly expanding and/or changing. Maintenance-time requirements should compare favourably with conventional methods.

There now remains the problem of preparing the printed subject-heading list. A Hollerith tabulator can list (print) at the rate of 9,000 lines per hour (5,000 for older models). Consequently two to three hours of tabulator time is required whenever a new list is required. Multiple-copy carbon forms are commonly used for as many as a dozen copies. If a larger number of copies is required, continuous multilith mats can be used. For even larger quantities photo-offsetting techniques can be employed, by pasting the original printed copy on boards, followed by photography and the preparation of offset plates.

Tabulating machines only offer upper-case type. At the Library of Congress a card-activated typewriter is employed permitting the usual variety of upper- and lower-case characters. However, the latter equipment is not generally available. One hopes that in the near future the manufacturers of punched-card equipment will produce a more versatile tabulator, printing upper and lower case. The IBM 407 tabulator, while still employing one case, permits the use of eleven special characters, the comma, period, and asterisk being the only three of much use in subject-heading lists.

The reader must keep in mind that such an operation can only be economical if automatic punched-card equipment is already available (and this is commonly the case), or if the resources of a service bureau are utilized on an hourly basis. Both of these alternatives can usually be expected to permit a comparatively inexpensive method for preparing lists.

IV

Those who are contemplating new subject-heading lists must, of course, anticipate intellectual problems already tackled by those well established in the subject-heading business. Punched-card techniques can, however, facilitate methods already employed. It is first necessary to gather all the terms (subject-headings) of the subject field. The usual literature sources

PSYCHOSES INVOLUTIONAL		09530	6895
X	MELANCHOLIA INVOLUTIONAL	09530	6896
PSYCHOSES MANIC DEPRESSIVE		09530	6897
XX	DEPRESSION		6898
PSYCHOSES PRESENILE		09530	6899
X	ALZHEIMERS DISEASE	09530	6900
X	DEMENTIA PRESENILE	09530	6901
X	PICKS DISEASE OF BRAIN	09530	6902
X	SCLEROSIS PRESENILE	09530	6903
XX	PSYCHOSES		6904
PSYCHOSES SENILE		09530	6905
PSYCHOSES TOXIC		09530	69053
SA	PSYCHOSES ALCOHOLIC		69056
PSYCHOSOMATIC MEDICINE	S MEDICINE PSYCHOSOMATIC		6906
			6907
PSYCHOSURGERY		102	6908
SA	NEUROSURGERY		6909
X	GYRECTOMY	102	6910
X	LEUCOTOMY	102	6911
X	LOBOTOMY	102	6912
X	TOPECTOMY	102	6913
XX	NEUROSURGERY		6914
PSYCHOTHERAPY		101	6915
SA	PROJECTIVE TECHNIQUES		6916
PSYCHOTHERAPY GROUP		101	6917
PSYLLIUM		0141	6918
X	METAMUCIL	0141	6919
PTERINS		03399	6920
SA	FOLIC ACID ANTAGONISTS		69205
PTERYGIUM		09538	6921
XX	CONJUNCTIVA DISEASES		69215
PTERYGOID BONE S SPHENOID BONE			6922
PUBERTY		0432	6923
SA	ADOLESCENCE		6924
XX	ADOLESCENCE		6925
PUBERTY PRECOCIOUS		09527	6926
PUBIC BONE		02214	6927
SA	PUBIC SYMPHYSIS		6928
PUBIC SYMPHYSIS		02214	6929
X	INTERPUBIC FIBROCARTILAGE	02214	6930
X	SYMPHYSIS PUBIS	02214	6931
XX	PUBIC BONE		6932
PUBLIC HEALTH		1605	6933
SA	HEALTH INSURANCE		6934
SA	INDUSTRIAL HYGIENE		69345
SA	MEDICINE PREVENTIVE		6935
SA	MEDICINE SOCIAL		6936
SA	NATIONAL HEALTH PROGRAMS		69366
SA	OCCUPATIONAL DISEASES		69372
SA	PUBLIC HEALTH NURSING		69373
SA	SANITATION		69374
XX	MEDICINE PREVENTIVE		6938
PUBLIC HEALTH NURSING		1605	69382

FIG. 3. Specimen page of medical subject-heading list. Each heading is followed by a serial number for alphabetizing and a category number for categorization.

(existing subject-heading lists, dictionaries, indexes, glossaries, &c.) are investigated for new terms. A punched-card is prepared on the key-punch for each term. If one were merely interested in a list of terms, the key-punching work could stop at this point. This would be equivalent to the first step described in Part III except that the terms would not be in alphabetical order. The cards could be alphabetized on the sorter and a list of terms, i.e. subject-heading list, prepared by feeding the deck of alphabetized cards into the printing tabulator. However, subject-heading lists usually have as their ultimate purpose the goal of providing keys to larger files of information. Such is the case, for example, in preparing indexes and catalogues. Consequently one wishes to analyse the terms so that consistent decisions can be made in indexing and cataloguing. If you examine an alphabetical list of 10,000 terms, it proves rather difficult to make decisions in preparing consistent cross-references. For example, the term ADRENALINE may appear in the list as well as the term EPINEPHRINE. This is an example of two synonymous terms. Examination of such a lengthy list might not show the presence of both terms. It would then be possible for two indexers to be indexing information under both headings without knowing it. For example, the indexes to the *Journal of Documentation* use the two following subject-headings—*punched card systems* and *Hollerith systems*. There is no cross-reference between these terms. Unless the reader is aware that both headings are used it would be possible to overlook pertinent material. This is especially true in volume 4. Examination of the index shows that only one entry is common to both headings and yet there are a total of ten entries. At present, then, experience has been the guide in these matters. Through constant usage many redundancies are eliminated as they are encountered and necessary cross-references provided. However, experience does not always enable one to establish all necessary cross-references in a logical pattern. It is only through the approach of categorization that it is reasonably possible to cross-reference properly a subject-heading list and achieve a greater degree of access to all desired information.

V. *Categorization*

It has been found useful to establish approximately a dozen broad categories into which terms can fall. In the case of medical subject-headings, this has been quite rewarding. According to the proponents of Colon classification this is a general principle. The reader should not have too much difficulty in deciding on his categories. Just start to categorize terms and the broad categories will soon become apparent. Thus, ADRENALINE falls into the category of chemistry; TUBERCULOSIS in pathological conditions, and MONKEY in organisms. What one is striving for is a simple method of bringing together related terms. It is in the attempt at categorizing terms that the reader will locate redundancies in his basic list of terms while at the same time establishing a basis for sub-categorization. In the beginning more than a dozen categories will suggest themselves, but many are gradually eliminated.

Since it is not the purpose of this paper to discuss the merits of various classification procedures, it will be for the reader to decide what approach to use here. Once a category number has been assigned to a term it is then necessary to punch this number in the card for that term, and by sorting cards on the appropriate columns one obtains a category arrangement of

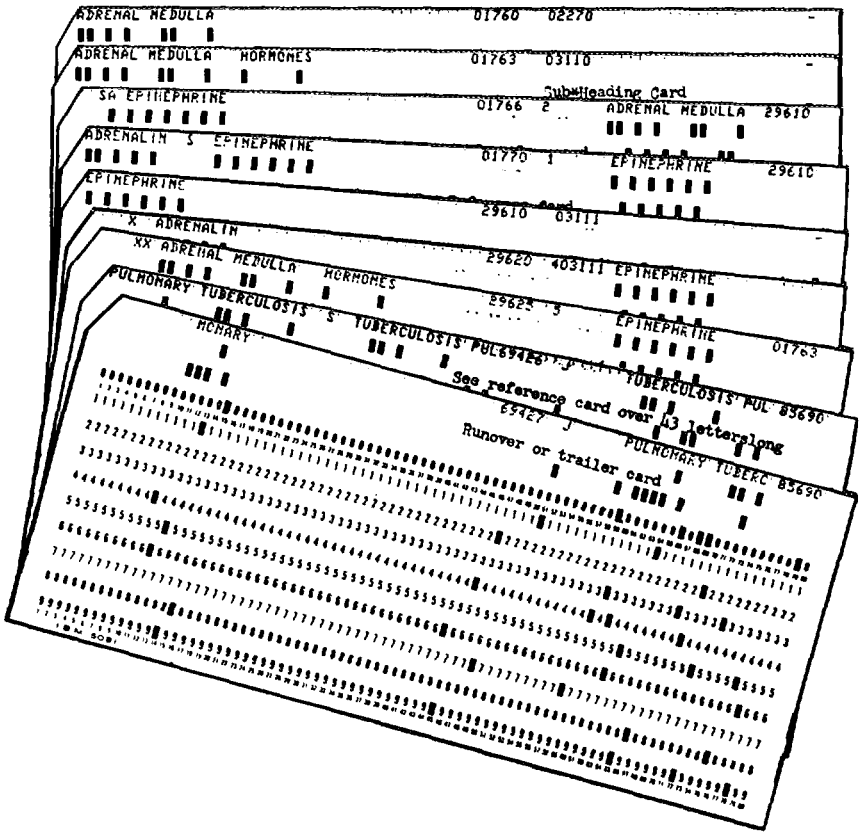


FIG. 4. Sample subject-heading cards.

terms. If more than one arrangement of terms is desired, a second or third category number can be used. It is possible to tabulate these categories, study the resulting lists of terms, make further sub-categorizations, and then make a decision for each term depending on the particular use to be made of the subject-heading list. Thus, in the case of ADRENALINE and EPINEPHRINE, both terms having been assigned the category number for chemistry, they naturally were found in a list of all terms in chemistry. Since the number of terms in this field was quite large, it was then necessary to further sub-categorize. These two items came together under the sub-grouping of hormones. It was only then that the obvious cross-reference from one term to the other

could be established. A *see* reference card, ADRENALINE SEE EPINEPHRINE, was prepared as well as an X reference card using the system of *see from* references employed in the Library of Congress List of Subject Headings.[1] The latter is extremely useful, but may not be necessary in some lists. It will also be found most useful to use category lists of headings during indexing as well as for the development of the original alphabetical list.

A serial number is assigned to each card, which permits one to alphabetize the file very rapidly, using a sorter. This is especially important during the preparatory stages. It also permits the retention of a single deck of cards and enables one to return more quickly to the alphabetized arrangement after the cards have been sorted by category or by type. It is often useful to remove all *see* references or all main headings and prepare lists of each separately. By assigning a type number to each card, e.g. a 1 punch for *see* references, it is possible to sort cards by type, bringing together all *see* references. These advantages will become quite apparent once the punched-card file has been in use a short time. This technique can be very useful in preparing X or XX cards mechanically. By use of the reproducing punch it is possible to prepare a matching X or XX card for every *see* or *see also* card. The serial numbers for alphabetizing by number should be added after the deck of new punched-cards has been alphabetized by letter. In this first alphabetization one must sort the cards on as many as thirty or forty columns. Although quicker than manual techniques alone, it is possible to achieve much faster alphabetic sorting by use of the 101 Electronic Statistical Machine[2] or by combined machine-manual techniques. The latter is achieved by sorting the cards on the first ten letters by machine and finishing the job manually. It is also possible to cut down on alphabetizing time by use of the collator. It will also become apparent that if X and XX references are used it will not be possible to alphabetize mechanically the cards without the use of serial numbers.

In conclusion, of the many possible uses of punched-cards, the preparation of subject-heading lists can be shown to be a most suitable and profitable application. The use of this medium will encourage frequent publication of current subject-heading lists, the importance of which cannot be over-estimated in library practice, documentation, and bibliographical control in general.

[1] *Subject headings used in the Dictionary Catalogs of the Library of Congress*, Fifth Edition, prepared by Subject Catalog Division, Library of Congress, Wash., D.C., 1948.

[2] Cf. Garfield, Eugene. 'The mechanical analysis of information by use of the 101 Electronic Statistical Machine'. *American Documentation*, Vol. 5, no. 2.

Appendix on p. 10

Appendix

Directions for punching cards for Subject-Heading Authority List of the Current List of Medical Literature

1. *Main headings.* The alphabetic subject-heading is punched in columns 1–43, since there are only 43 alphabetic type-bars on the standard tabulator. The five-digit serial number is punched in columns 44–48. The category number is punched in columns 52–57, and finally an 11 (X) punch is made in column 80. Always punch in zeros in categories 1 through 9. Thus category 2 is always punched 02·0 and not 2·0.

   All sub-headings are considered as main-heading cards. There should be three spaces between the heading and sub-heading. Thus LIVER is one main heading card and LIVER–PHYSIOLOGY is another main-heading card. These terms would not have the same category numbers.
2. *See references.* First 43 columns alphabetic, 44–48 for the serial number, a 2 in column 51, as much of the alphabetic heading referred to in columns 58–74 as possible, followed by the serial number of the heading referred to in columns 75–79.
3. *Run-over cards. See* references usually are the type of card necessitating a run-over or trailer card, though occasionally 43 columns is not enough for a main heading. Start the run-over card in column 10 and use as many as 34 columns. In columns 44–48 assign a serial number one digit higher than the master card. In column 51 punch the same information as the master card. In columns 58–74 punch the first 27 letters of the master card and in the case of *see* references punch the serial number of the heading referred to in columns 75–79.
4. *See also cards.* Start the *see also* card in column 3, punching SA in columns 3 and 4, followed by a single space followed by the heading referred to. In 44–48 punch the serial number, in column 51 a 3 punch, in 58–74 the heading under which the *see also* appears, and in 75–79 the serial number of the heading referred to.
5. *See from or X cards.* Punch X in column 4, skip two spaces, and punch heading referred from. Punch serial number in 44–48, punch 4 in column 51, and punch category number of heading referred to in columns 52–57. Heading referred to in columns 58–74.
6. *See also from or XX cards.* Punch XX in columns 4 and 5, skip one space, and punch heading referred from, punch serial number in columns 44–48, punch 5 in column 51 and the heading referred to in columns 58–74.
7. *Multiple see references.* The second card of the group should not have the heading referred from punched. In this way the heading referred from will only print other *see* references; e.g. CEREBRUM S BRAIN 1671
   S CEREBRAL CORTEX 1672

	Start in column	*Serial number*	*Column 51*	*Category number*	*Heading referred from or to*	*Col. 80*
Main . .	1	44–48	1	52–57	—	11
See . .	1	,,	2	—	58–74; 75–79 (serial no.)	—
Run-over .	10	,,	same	—	,, ,,	—
See also . .	3	,,	3	—	,, ,,	—
X . . .	4	,,	4	52–57	,, ,,	—
XX . .	4	,,	5	—	,, ,,	—