»PIRS« (Personal Information Retrieval System)

A Multi-purpose Computer Program for Storage and Retrieval of Reference Files *)

From the Research Computation Center of the University of Texas, Medical Branch, Galveston
(Director: Prof. Dr. John E. Overall) and the University of Missouri, Medical School,
Radiology Computer Research (Director: Prof. Dr. Peter L. Reichertz) Columbia, Missouri

P. L. REICHERTZ

»PIRS« is a personal information retrieval system for individual reference files. It has been designed to store and retrieve references by identification number, keywords and any string of characters in the name, title or bibliography section of the reference. The keywords have to be defined by the user and hence may be chosen according to his particular needs. The selected references may be printed with or without full text keywords.
»PIRS« is available in FORTRAN II-D for an IBM 1620 and in FORTRAN IV-360 for 360 models.

»PIRS« — EIN VIELZWECK-COMPUTER-PROGRAMM FÜR DIE LITERATURDOKUMENTATION

»PIRS« ist ein Informationssystem zur Bearbeitung von individuellen Literatursammlungen. Es vermag Literaturzitate nach Identifikationsnummer, Schlüsselworten und Buchstabengruppierungen in Autorennamen, Titel und Bibliographie auszuwählen und auszudrucken. Die Schlüsselworte werden vom Benutzer selbst definiert und können somit den speziellen Anforderungen angepaßt werden. Das Ausdrucken der Zitate kann mit oder ohne Schlüsselworte (in Volltext) erfolgen.
»PIRS« existiert in einer Version für einen kleinen, aber weitverbreiteten Computertyp (IBM 1620) und in einer zweiten leistungsfähigeren und schnelleren Version für die 360-Modelle.
Programm und Anwendungsbeschreibung stehen auf Anforderung zur Verfügung.

The spectrum of computer application in medicine is manifold. It comprises administrative data processing, menu planning, monitoring, storage and retrieval of literature and patient history, generation of reports (1, 6), design and planning of experiments and clinical trials as well as development of hypotheses, simulation of hospitals, departments and diseases. It encompasses diagnostic aids, computer-aided diagnosis and computer-aided instruction (8).

Usually the application of computers requires special training. This restricts their applicability. Scientists outside computer oriented teams are mainly interested in statistical analyses, and storage and retrieval of information or medical records. Modern communication devices and physician-computer-interfacing (PCI) by means of cathode-ray tubes or similar terminals will expand the applicability for clinical purposes.

The program libraries of most computer centers contain general statistical programs that may be adapted to individual problems. Processing of literature data is more difficult, due to the diverging interests and projects of the prospective user. Intensive and comprehensive collection of all medical information requires large centers with great staff and storage capability. Continuous updating is necessary.

```
          COMPUTER-AIDED DIAGNOSIS   (IN GENERAL)

          _____

   32 DESCRIPTION OF THE DIAGNOSTIC PROCESS
   21 GENERAL DESCRIPTION OF MATHEMATICAL MODELS
   22 DESCRIPTION OF MATHEMATICAL MODELS; FORMULAS, THEORY
   28 ARTICLE DESCRIBES DIAGNOSTICAL RESULTS
   91 DIAGNOSTICAL PRINTOUT IS CLINICALLY APPLICABLE
```

```
*******************************************************************************

          DISEASE GROUPS; DISCIPLINES; FIELDS OF RESEARCH, RESEARCH METHODS

          _____

   80 GENERAL DISEASES (GENERAL DIAGNOSTIC)
   65 CARDIOLOGICAL DISEASES IN GENERAL
   66 CONGENITAL HEART DISEASES
  210 ACQUIRED VALVULAR HEART DISEASES
  149 DISEASES OF THE RESPIRATION TRACT
   68 HEMATOLOGICAL DISEASES
   79 MALIGNANT TUMORS
   86 DISEASES OF THE GASTRO-INTESTINAL TRACT
  120 LIVER DISEASES

   63 RADIOLOGY      (FILM READING, REPORT GENERATION, INFORMATION STORAGE)
   23 RADIOLOGY, DIAGNOSTIC
   95 RADIATION THERAPY
  107 NEURORADIOLOGY
   88 IMAGE RECOGNITION AND ANALYSIS
  105 IMAGE ENHANCEMENT  (COMPUTER APPLICATION)
  106 IMAGE ENHANCEMENT  (LASER AND FILTER TECHNIQUES)
  100 PRIMARY BONE TUMOR
  101 LUNG NODULES

  199 ENDOCRINOLOGY
   67 THYROID GLAND

   83 NUCLEAR MEDICINE   (EXCEPT THYROID GLAND)
   30 PATHOLOGY
   36 MICROBIOLOGY AND PREVENTIVE MEDICINE
  201 GYNECOLOGY
  172 GERIATRICS
  110 PSYCHIATRY
  216 NEUROLOGY
  139 PEDIATRICS
  205 DENTISTRY

  109 VARITYPING OF CHROMOSOMES

*******************************************************************************
```

Figure 1: Part of a keyword list. The keywords are ordered according to their meaning and hierarchical order. Version (a) (IBM 1620) accepts 1—3 digit keyword numbers; version (b) (IBM 360) is programmed for 1—4 digit keyword numbers

Information systems have been developed for all types of computers. Powerful sorting and merging programs are provided by the manufacturers. But all these systems have to be as general as possible, which sometimes is a compromise between the requirements of the different user groups.

The individual researcher usually compiles a reference file during his work on a project. He would like to have the information system adapted to his own terminology and requirements. His keywords are not necessarily similar or equal to those that would be used in a general information system. The chronological order of keywords or key phrases may serve him as a short summary of an article.

These reflections initiated the development of »PIRS« which has not been designed in order to collect and distribute general information but is intended to be a practical aid for the scientist in his work with his own reference file. The programs select references of particular interest and may compile reference lists for publications.

The advantage of »PIRS« is its adaptability to individual reference files. It has been designed in two versions (a) for a very common but small computer type (IBM 1620) that does not require a large computer center and (b) for IBM 360 machines providing speed and large storage capacity. It may be used without special programming experiences (7).

Version (a) has been written in FORTRAN II-D, version (b) is coded in a 360-FORTRAN IV. Program listings and special instructions are available on request.

Description of Input Data

The format of the data input was kept variable (compare 2, 3, 4, 5). It thus takes account of the different lengths of the individual references. The input consists of two information categories:

1.) *Keyword lists.* The keywords or key phrases have to be formulated according to the objectives of the reference file. Each keyword or key phrase is identified by a unique number. Version (a) accepts 1 through 3, version (b) 1 through 4 digits. The keyword or key phrase may contain up to 75 characters. Consequently, words, phrases or small sentences may be used. The upper limits of keywords for version (a) is 300. Version (b) currently accepts up to 5000. This number may be increased according to the available core.

For the practical use of the keyword list it is suggested that the keywords be listed in their hierarchical order or according to their meaning. Such a listing may be used while screening the literature and thus the pertaining keyword numbers can easily be dictated. Figure 1 gives an example of a keyword list used for the problem of computer-aided diagnosis. When a new keyword or -phrase has to be added, the punched card may be inserted in its proper place and a new listing can be made either by an IBM 407 or with the computer itself.

Keyword lists may be translated into any desired language. In using version (b) it is possible to refer to the wanted keyword list by changing the pertaining job control language (JCL) card. For Figure 4 a German keyword list has been used. The exchange of a single JCL-card in the deck causes the English translation to be printed.

Naturally the data deck for the job itself must not contain the comments shown in Figure 1. It is to be suggested that a copy of the keyword list be made and sorted according to the keyword numbers. In such a fashion duplications of keyword numbers are easily avoided.

A labelling of keyword numbers may also be done by using numerical ranges for hierarchical groups or particular research fields.

The program tests the number of keywords and prints error comments when the provided maximum is reached.

2.) *References.* Each reference is divided into five levels of information: identification number, name of author(s), title, bibliography and keyword numbers. All cards belonging to the same reference must have the same identification number in columns 1—4, right justified. Name of author(s), title and bibliography are punched in one sequence, separated by a »$« sign.

In order to make the system flexible, it is not necessary that this information sequence contain all three levels. Thus, for instance, the title section may be missing. More levels may be added in version (b).

If more than one punch card is necessary for the information levels 1 through 4, up to 5 cards may be used. A continuation must be indicated by punching a »1« in column 80. [System (b) accepts any non-blank character].

The keyword numbers have to begin with a new card after the information units 1—4. Since this card must have the identification number also in column

```
207KRUESKEMPER•H•L••REICHERTZ•P••MORGNER•K•D•$ELEKTROGRAPHISCHE SCHNELLMETHODE1
207ZUM NACHWEIS VON CORTICOTROPIN$ACTA ENDOCRINOLOGICA• IN DRUCK
207199158195
208REICHERTZ•P•L••NOELL•G••HEMMATI•A•$ZUR BEHANDLUNG VON STOERUNGEN DES HERZMU1
208SKELSTOFFWECHSELS MIT KALIUM-MAGNESIUMASPARTAT$DIE MED•WELT•IN DRUCK
208202
209TEMPLETON•A•W••SIDES•S•D•$RADIATE-A RADIOLOGY (AND HOSPITAL) COMPUTER-ORIEN1
209TED COMMUNICATING SYSTEM$NATIONAL CONF• ON COMPUTER APPLICATIONS IN RADIOLO1
209GY•COLUMBIA/MO••1967
209 96 63
210LAMSON•B•G•$STORAGE AND RETRIEVAL OF MEDICAL DIAGNOSTIC STATEMENTS IN FULL 1
210ENGLISH TEXT$NATIONAL CONF•ON COMPUTER APPLICATIONS IN RADIOLOGY•COLUMBIA/M1
2100••1967
210 96 97 30 108
```

Figure 2: Randomly chosen data of an own reference file. All cards belonging to the same reference must be labeled with the same identification number. Name of author(s), title and bibliography are separated by a »$«-sign. A continuation of the record is indicated in column 80. The keyword numbers start with a new card. Here also continuation cards are possible

1 through 4 and, as described above, a continuation has to be indicated in column 80, up to 25 1—3 digit keyword numbers may be punched into each. If in version (b) four digit keyword numbers are used, this has to be indicated in a header card and consequently only 18 numbers may be accepted per card. (The documentation of the programs gives detailed descriptions).

Currently the maximum numbers of cards per reference without keyword number cards is 5. Consequently 375 characters may be used. Version (a) accepts up to 225, version (b) up to 5000 keyword numbers per reference.

These limitations are checked and, in case of violations, error comments are printed. In addition to this, other error conditions are tested.

Figure 2 shows a randomly chosen example of a literature file. In this fashion the deck has to be prepared for the run.

Version (a) accepts up to 2000 different references, when only one card is used per reference and per keyword numbers. If 5 cards are used, 560 different references may be processed. Version (b) is currently programmed for up to 5000 references. According to available disk space and core storage this number may be increased.
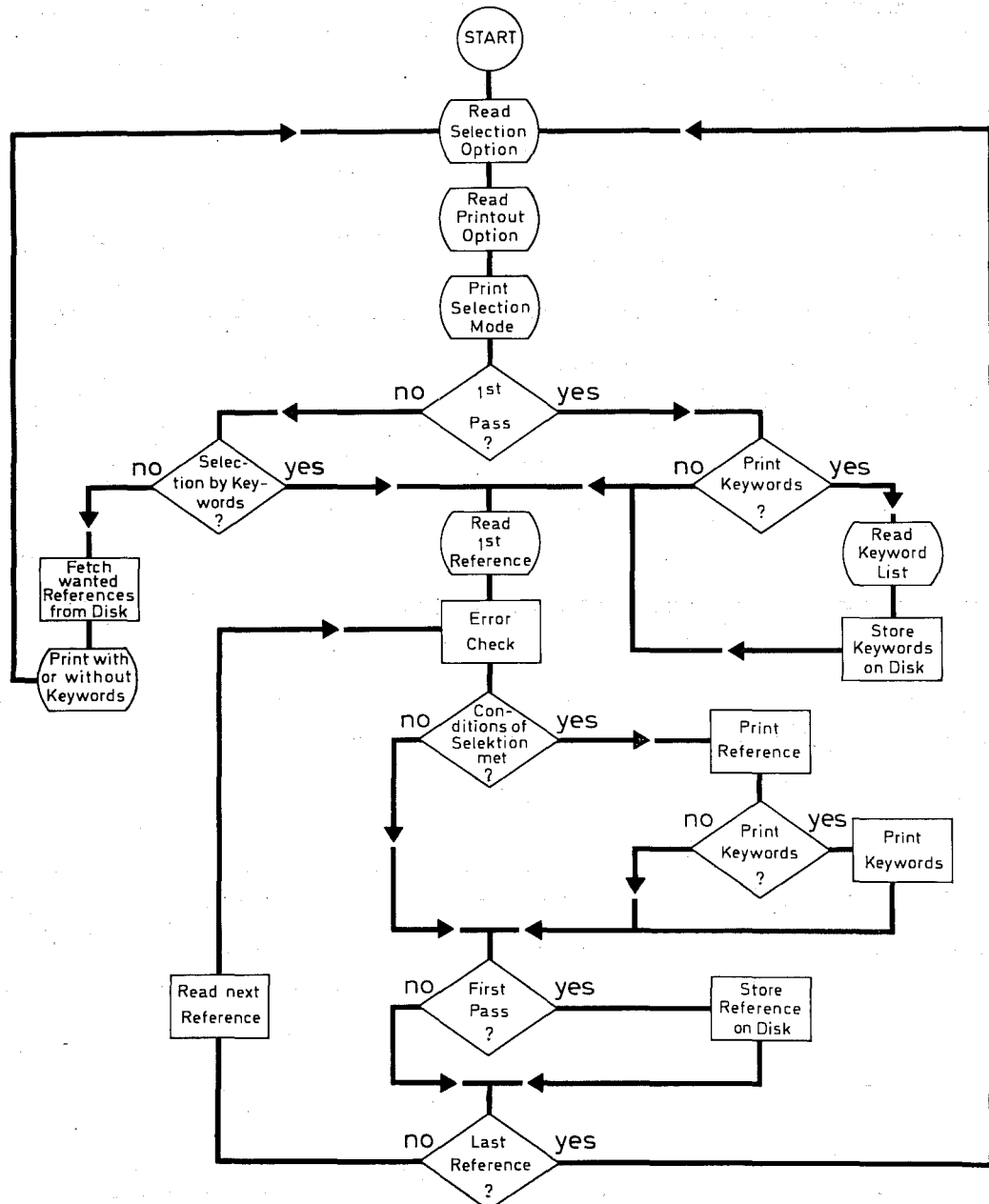

Figure 3: Simplified and schematized flow chart of »PIRS«, version (a). During the first search the data are read in, for subsequent searches the information will be fetched from disk. The selection option by any string of characters follows the same logic as the selection by keyword numbers. Version (b) stores the data permanently on disk. Updating is possible

Program Logic

Figure 3 is a schematized flow chart for version (a). Always a number of statements is combined in a descriptive block. For the first pass the data are read in from punch cards; all following searches fetch the data from disk storage.

The different error checks are not shown in detail. The flow chart does not describe the provided selection by a string of characters. The pertaining program flow basically follows that provided for the selection by keyword numbers.

In this way the flow chart could be made as simple as possible and will be understood even by those who do not have programming experience. As already indicated, version (a) needs always to read in all data for the first search. All secondary searches are performed in a more rapid way from disk storage.

```
*                         LIST OF REFERENCES
(THIS IS A COMPUTER-PRINTOUT. DISREGARD  FOLLOWING SELECTION-COMMENTS WHEN PRIN-
TING)


TO BE SELECTED
 108   96


OPTION      2      REFERENCES SELECTED FROM KEYWORDS
PRINTOUT WITH KEYWORDS
      KEY-WORDS (SCHLUESSELWORTE)
      VOLLTEXTVERARBEITUNG
      BEFUNDDOKUMENTATION


   1  KOREIN,J.,BENDER,A.,ROTHENBERG,D.,TICK,L.
      COMPUTER PROCESSING OF MEDICAL DATA BY VARIABLE-FIELD-LENGTH FORMAT
      J.AMER. MED. ASS. 196,957-963,1966


      KEY-WORDS (SCHLUESSELWORTE)           REFERENCE NR.        24
      BEFUNDDOKUMENTATION
      VOLLTEXTVERARBEITUNG


   2  LAMSON,B.G.
      STORAGE AND RETRIEVAL OF MEDICAL DIAGNOSTIC STATEMENTS IN FULL ENGLISH TEXT
      NATIONAL CONF.ON COMPUTER APPLICATIONS IN RADIOLOGY,COLUMBIA/MO.,1967


      KEY-WORDS (SCHLUESSELWORTE)           REFERENCE NR.        210
      BEFUNDDOKUMENTATION
      AUSWERTUNG VON BEFUNDEN NACH SYNTAXREGELN
      PATHOLOGIE
      VOLLTEXTVERARBEITUNG


   3  BARNHARD,H.J.
      AUTOMATIC CODING AND MANIPULATING OF RADIOLOGY DIAGNOSTIC REPORTS
      NATIONAL CONF.ON COMPUTER APPLICATIONS IN RADIOLOGY,COLUMBIA/MO.,1967


      KEY-WORDS (SCHLUESSELWORTE)           REFERENCE NR.        211
      BEFUNDDOKUMENTATION
      RADIOLOGIE, DOKUMENTATION UND BEFUNDUNG
      AUSWERTUNG VON BEFUNDEN NACH SYNTAXREGELN
      BEFUNDKONDENSATION
      ORTHOGRAPHIEKORREKTION
      VOLLTEXTVERARBEITUNG


   4  KRICHEFF,I.I.,KOREIN,J.,CHASE,N.E.
      COMPUTER PROCESSING OF NEURORADIOLOGIC REPORTS BY VARIABLE-FIELD-LENGTH
      FORMAT
      RADIOLOGY 86,1100-1106,1966


      KEY-WOR4S (SCHLUESSELWORTE)           REFERENCE NR.        238
      NEURORADIOLOGIE
      BEFUNDDOKUMENTATION
      VOLLTEXTVERARBEITUNG
```

Figure 4: Example of a selection by keywords. Printout with keyword text. Here a German keyword list has been used. (Volltextverarbeitung = processing of full text; Befunddokumentation = documentation of findings). By changing the pertaining JCL-card, the English text may be used. (The actual list has been shortened for reproduction)

Version (b) is expanded and more flexible. Here the data are stored permanently on disk. By means of control cards the program flow can be modified for three conditions: (1) loading of new data. (Control word »LOAD«). This option will be used for the first application of the program and the initialization of the data sets. (2) The control word »UPDATE« updates the information stored on disk. This version accepts either keywords, references or both. (3) »RETRIEVE« performs a search without updating.

The control cards »LOAD« and »UPDATE« may contain information on which media the data are to be found (punch cards, disk or tape) and the pertaining data set number. It may also be specified that a control printout of the loaded data is desired. Version 1 and 2 branch automatically to version 3 after loading or updating, so that further selection tasks are possible.

Selection Modes

The system may follow three modes for selection:

1.) A selection is possible using the keyword numbers or a combination of keyword numbers coded for an individual reference. This option is used when a literature search is wanted to retrieve articles with certain information. Version (a) provides for a combination of up to 225, version (b) up to 5000 keyword numbers for a single selection pass. Only those references are printed that contain all given keywords. Thus, if more than one keyword is specified for the selection, a logical »AND« is performed. (See Figure 4).

The flow chart (Figure 3) indicates that the selected references are printed in the order in which they are stored. An alphabetic sorting would be

```
                        LIST OF REFERENCES
THIS IS A COMPUTER PRINTOUT. DISREGARD FOLLOWING SELECTION COMMENTS WHEN PRIN-
TING

TO BE SELECTED
 211   22   23   24 238 210 262 122 490

OPTION       1      REFERENCES SELECTED BY IDENTIFICATION NO.
PRINTOUT WITHOUT KEYWORDS

   1 BARNHARD,H.J.
     AUTOMATIC CODING AND MANIPULATING OF RADIOLOGY DIAGNOSTIC REPORTS
     NATIONAL CONF.ON COMPUTER APPLICATIONS IN RADIOLOGY,COLUMBIA/MO.,1967

   2 KOREIN,J.,GOODGOLD,A.,RANDT,C.
     COMPUTER PROCESSING OF NARRATIVE DATA
     NEUROLOGY 16,848-857,66

   3 KOREIN,J.,GOODGOLD,A.,RANDT,C.
     COMPUTER PROCESSING OF MEDICAL DATA BY VARIABLE-FIELD-LENGTH FORMAT
     J.AMER. MED. ASS., 196,950-956,1966

   4 KOREIN,J.,BENDER,A.,ROTHENBERG,D.,TICK,L.
     COMPUTER PROCESSING OF MEDICAL DATA BY VARIABLE-FIELD-LENGTH FORMAT
     J.AMER. MED. ASS. 196,957-963,1966

   5 KRICHEFF,I.I.,KOREIN,J.,CHASE,N.E.
     COMPUTER PROCESSING OF NEURORADIOLOGIC REPORTS BY VARIABLE-FIELD-LENGTH
     FORMAT
     RADIOLOGY 86,1100-1106,1966

   6 LAMSON,B.G.
     STORAGE AND RETRIEVAL OF MEDICAL DIAGNOSTIC STATEMENTS IN FULL ENGLISH TEXT
     NATIONAL CONF.ON COMPUTER APPLICATIONS IN RADIOLOGY,COLUMBIA/MO.,1967

   7 REICHERTZ,PETER L.
     A GENERAL PURPOSE REFERENCE PROGRAM FOR A SMALL COMPUTER
     TEX. MED. REPORTS, IN PRESS

   8 REICHERTZ,P.L.
     ELEKTRONISCHE DATENVERARBEITUNG IN DER MEDIZIN
     DER LANDARZT,43,1201-1207,1967

   9 REICHERTZ,P.L.,TEMPLETON,A.W.,DELURGIO,L.J.
     A FORTRAN COMPATIBLE MULTI-TERMINAL SYSTEM FOR PHYSICIAN-COMPUTER-INTERFAC-
     ING (FCMTS)
     SIXTH ANNUAL SYMPOSIUM ON BIOMATHEMATICS AND COMPUTER SCIENCE IN THE LIFE
     SCIENCES,HOUSTON/TEX.,MARCH 1968
```

Figure 5: Selection by identification number without printout of keywords. Figure 5 is the list of references of this article.

possible, but has not been programmed in order not to decrease storage capacity and speed. This option can easily be added.

2.) The second selection option is the retrieval by identification number. This option is used when the identification numbers of the wanted references are known (for instance, after searches in mode 1). In this fashion reference lists may be generated to be submitted for publication. Also a personal bibliography can be made in this way.

The flow chart for the 1620-version (Figure 3) shows that the first pass in selection mode 2 does not print the references in the pre-given order, but in the order as they appear in the data deck. The specified order will be followed for all subsequent passes. In case this order only is wanted, a perfunctory selection

may be performed during the loading of the data before the wanted selection. Naturally version (b) does not have restrictions. Here the selection is always in the strict order of the pre-given sequence of identification numbers.

3.) The third selection mode provides the option to select a reference by any string of characters (a: up to 225; b: up to 5000). The program flow follows selection mode 1.

This option allows the selection of references according to name of author(s), words in the title section or journal names in the bibliography. Using stems such as »DIAGNOS«, all titles will be selected that contain the words »Diagnose«, »diagnosis«, »diagnoses«, »diagnostisch«, »diagnostic«, »diagnostical«, »diagnostique« and so on.

```
*                          LIST OF REFERENCES
(THIS IS A COMPUTER-PRINTOUT. DISREGARD  FOLLOWING SELECTION-COMMENTS WHEN PRIN-
TING)


TO BE SELECTED
PIPBERGER


OPTION      3      REFERENCES SELECTED BY CHARACTERS
PRINTOUT WITHOUT KEYWORDS


   1  BERSON,A.S.,STALLMANN,F.W.,BRODERS,J.H.,PIPBERGER,H.V.
      TELEPHONE TRANSMISSION OF ELECTROCARDIOGRAMS AND ON-LINE COMPUTER DIAGNOSIS
      AMER.J.MEDICAL ELECTRONICS 35-37,65


   2  COSMA,J.,STALLMANN,F.W.,PIPBERGER,H.V.
      DIGITAL COMPUTATION OF THE SPATIAL VENTRICULAR GRADIENT IN NORMAL AND PREMA
      TURE VENTRICULAR CONTRACTIONS AND INTERMITTENT BUNDLE-BRANCH BLOCK
      CIRCULATION 24,910,61

   3  PIPBERGER,H.V.,FREIS,E.D.
      AUTOMATISCHE ANALYSE KARDIOLOGISCHER ANALOGDATEN MITTELS ELEKTRONISCHER REC
      HENMASCHINEN
      MED.DOK.4,58-61,60

   4  PIPBERGER,H.V.,FREIS,E.D.,TABACK,L.,MASON,H.L.
      PREPARATION OF ELECTROCARDIOGRAPHIC DATA FOR ANALYSIS BY DIGITAL ELECTRONIC
      COMPUTER
      CIRCULATION 21,413-418,1960


   5  PIPBERGER,H.V.,ARM,R.J.,STALLMANN,F.W.
      AUTOMATIC SCREENING OF NORMAL AND ABNORMAL ELECTROCARDIOGRAMS BY MEANS OF A
       DIGITAL ELECTRONIC COMPUTER
      PROC.SOC.EXPER.BIOL.MED.106,130-132,61


   6  PIPBERGER,H.V.,STALLMANN,F.W.
      USE OF COMPUTERS IN ECG INTERPRETATION
      AMER.HEART J.64,285-286,62


   7  PIPBERGER,H.V.,STALLMANN,F.W.,BERSON,A.S.
      AUTOMATIC ANALYSIS OF THE P-QRS-T-COMPLEX OF THE ELECTROCARDIOGRAM BY DIGIT
      AL COMPUTER
      ANN.INTERN.MED.57,776-787,62
```

Figure 6: Example for the selection by any string of characters. For this selection the name »Pipberger« was used. The printed list has been shortened for reproduction. Words, word stems or numbers in the title or bibliography section may be used as well

Output Description

Output examples are given in Figures 4—6. The selection mode and the specification of the search is printed. Name, title and bibliography of the individual references are separated and started in new lines. The printed references are numbered.

For the printout two options are provided (Figure 3). The first prints the keywords or -phrases that have been used for the selection in full text and specifies, also in full text, the keywords associated with each selected reference. Also the identification number is printed (Figure 4). (In using the selection option by identification numbers those numbers will always be printed in the selection comment; see Figure 5).

The references may also be printed without keywords. This option may be used for the compilation of a reference list. Figure 5 is the reference list of this article.

Version (b) provides options for multiple printouts and punched cards for separate listings.

Core Requirements

Version (a) requires a 32 K-character core memory. In its current form version (b) requires 100 K-bytes.

References

See Figure 5.

Author's Address: Peter L. Reichertz, M.D., Associate Professor and Director, Radiology Computer Research, University of Missouri, Medical School, 406 Turner, 410 Lewis Hall, Columbia, Missouri, 65201, USA.