V. PESCARU
M. IONESCU-BUJOR
V. DUMITRESCU

Information Storage, Retrieval and Editing System — S.I.R.P.I. *) — the first version

„Bring near more and more the future days to present days".

Glenn T. Seaborg

Men become more and more conscious as the years go by that the "information explosion" they have generated is a consequence of the natural striving to find the truth, to know the reality, to improve their work and their products. The reactions to this ascertained fact have been and are very much varied, beginning with the most pessimistic of them, who consider that the future generations will be "smothered" by the avalanche of information whom they cannot resist and ending with the most optimistic ones who completely overlook the novelty character of the contemporary information, considering that the pace the information system increases has remained unchanged in the course of history.

As a defence reflex against all the theories trying to speculate upon this critical situation engendering a state of mind leading to progress hindrance, as well as against those theories trying to ignore or to minimize the indisputable acuteness of the information problem, men started to scientifically approach this phenomenon: to analyse the "explosion" causes, to exactly measure its dimensions, to establish its effects, to find out its laws and, first of all, to take the measures required for reducing its negative effects. A new science of high resonance in the whole mankind life is forming beneath our very eyes: THE INFORMATION SYSTEM — the science of modern procedures and methods for information collecting, recording, transmitting,

*) S.I.R.P.I. = Sistem de Inmagazinare, Regăsire şi Prezentare a Informaţiei (Information Storage, Retrievai and Editing System)

storage, retrieval, processing and editing, and of human and material efforts made by men in this connection.

Information exists and continues to be generated in any field of human activity. Although men are making serious efforts to record all issued information, it seems, however, that if this is not impossible, it is at least useless. Selection and holding of this information including a large amount of news of interest for an as large as possible field of man activities is nearer the mankind interests and considerable results have been obtained to this end. From all the fields of activities it seems that the science and material production fields are asserting themselves as the main originators of information that is worth recording.

The information problem is no more a fashion problem only, as it is preocupying a larger and larger number of scientists and specialists in various fields of activity, statesmen and politicians, who are conscious of the special importance of the problem, of the economic and social implications it may have.

The achievement of the Information Storage, Retrieval and Editing System — S.I.R.P.I. — is also included in the context of scientific and practical concerns for putting more and more information at the service of the material and intellectual progress of the society.

S.I.R.P.I. is a modern instrument of ample and effective information for all the fields of activity where a great amount of information is accumulating in view of subsequent consultations. The system has a high degree of automation by using to the full the advantages of a third generation electronic computer: high speed, increased realiability, processing accuracy, magnetic supports (disks and tapes) for information holding, remote processing (teleprocessing), etc. It is not dependent on the application field or on language, and allows, for each application, the selection and addition of those functions of interest for the respective field.

With a view to retrieving and editing this information which is relevant for a certain moment, the respective data should fulfil two conditions:

— to be stored by magnetic grouped media, in units (e.g. documents) corresponding to some real units from the application field (e.g. source documents);

— these real units should be indexed by means of keywords (descriptors) pertaining to an already established and organized vocabulary.

The system allows data organization as well as vocabulary organization in four levels (e.g. the documents may be grouped in subsections, the subsections in sections and the sections in magazines). The retrieval speed and the information editing clarity directly depend also on the collection organizig degree.

The collection consultation with a view to finding out all the documents dealing with the same subject is the information retrieval operation. The subject should be expressed by means of keywords from the vocabulary and of the logical operators "and, "or" and "no", the documents being then automatically selected.

For example, in order to select the documents that have been printed in 1970 in connection with the WHEAT CULTURE ON NON-CLAYEY

FIELDS in conditions of a TEMPERATE OR TEMPERATE-CONTINENTAL CLIMATE, fill in a retrieval inquiry as follows:

1970. AND. WHEAT CULTURE
.AND. (TEMPERATE CLIMATE. OR. TEMPERATE CONTINENTAL CLIMATE)
.AND. (.NO.CLAYEY SOIL)

The inquiries may be filled in by the applicant in natural language, the transcription in the descriptors language being made by specialists, or directly in the descriptors language. In the conditions of a system with teleprocessing, the inquiry demands may be entered from consoles located in the most varied places and at no matter how remote distances with respect to the computer where the S.I.R.P.I. has been implemented. The responses to the retrieval inquiries are printed by the computer printer and then displayed to the applicants.

The system allows the periodical editing, or on demand, of all the entered (or existing) information of the collection The information editing form depends on the collection organization degree.

The appicability fields of S.I.R.P.I. are numerous. The most appropriate of all of them is the field of scientific, economic and technical information and documentation.

By some creative efforts applications of S.I.R.P.I. may be conceived in almost all the activity fields, such as, for example:

— in management work, for following up the activity of the units and of the subordinate staff, of task (duty) distribution and execution, etc.;
— in the activity of investigation, study and designing, inventions and standards, for following up who and what is he investigating, studying or designing, who and what inventions has he patented, who might investigate, study or design in our country or abroad, being a valuable support in drawing up and correlating the investigation, study and designing plans, in eliminating some significant material and human efforts now spent with research and designing to discover and create what is already existing, but is not yet known;
— in the home and foreign trade activities, for effective information on the market: who, what, how and how much he produces, sells or purchases;
— in the education and culture activity for following up the school syllabus, the activities of schools, teaching staffs, students, libraries, museums, theatres, cinemas, artists, writers, painters, sculptors, etc.;
— in justice, for following up the activities of the courts, judges and advocates, for selecting the law clauses with a view to substantiating decisions and sentences, etc.;
— in medicine and public health, for following up the disease spreading on the country territory, the distribution of medical staff and instruments, hospital activities, as well as those of maternities, and polyclinics, and for helping to determine the diagnosis.

The conception and designing of the present system has had as starting point the experimentation of the Information Retrieval and Management

System — I.R.M.S.* — of the firm I.B.M. with a view to its implementing to a documentation unit. While investigating the I.R.M.S. system, we have reached to the conclusion that it may be developed and completed with new programs for achieving additional functions.

With a view to emphasizing the original facilities of S.I.R.P.I., we are giving below a brief description of the I.R.M.S. system. Detailed information may be found in [8], [9].

The I.R.M.S. is an improved version according to the standards imposed by the I.B.M. firm, of the French SAGESSE ** system. We cannot say that I.R.M.S. is only a set of programs for the management and retrieval of information. It is a whole system including:

a. the storage, retrieval and editing concept of information;
b. the retrieval language: possibility of organizing the descriptor vocabulary, rules for filling in the inquiries for document selection, conversational possibilities;
c. the primary document design, temporary files, permanent files and those for the output documents;
d. a set of 7 programs: DCT1,2,3,4, INV1,2 and SRCH, written in ASSEMBLER language for computers I.B.M. 360 from the model 30 upwards, that may be carried out in a Disk Operating System — DOS;
e. management-operation procedures of the system;
f. the physical means for system operation: materials, equipments, staff.

a. I.R.M.S. Philosophy

I.R.M.S. is a retrieval system based on descriptors. The determination of the descriptors which are associated to a document is not automatically carried out, but remains a duty of indexing personnel, i.e. of the indexers. All the descriptors associated to a document have the same weight and are considered of the same significance for the respective document. Likewise, we cannot determine any kind of semantic connections or other kinds of connections between two or several descriptors at the indexing moment. The document resulting following the indexing — the analysis or indexing sheet or-card — includes:

— the recording number of the respective document — DAN (Document Accession Number);
— the descriptors associated to the document;
— title, author and, if possible, other bibliographical information;
— in case of need, the summary, or even the document, if this is of small size. The I.R.M.S. is generally not proposing to integrally record the documents, but only information about them.

---

*) Information Retrieval and Management System
**) SAGESSE = Système Automatique de GEstion et de SÉlection de Documentation

This information on the indexing form is then piloted on two directions: by means of the vocabulary, the descriptors are replaced by their numerical codes — DNR (Descriptor NumbeR) and coupled to the relevant DANs, forming a temporary file of direct searching, the other information, also associated with the relevant DANs, generate the so-called bibliographical file. For the retrieval operations, I.R.M.S. uses the procedure of the inverted searching files. The direct searching files, which are periodically generated, are reversed by associating to each descriptor (DNR) all the documents (DANs) that have been indexed with the respective descriptor or with its synonyms. There results an inverted file for searching, which is periodical, too; but this file is each time merging with a cumulative searching inverted file: the periodical one, by means of which the retrieval inquiries may be solved and which is referring to the respective period only — subscriptions — and the cumulative one, up to date, by means of which the accidental inquiries generally appealing to the entire document fund of the collection are solved.

The retrieval has as starting point a retrieval (selection) inquiry for the whole store or only for what has been lately stored. The retrieval inquiries are usually filled in in natural language, and afterwards translated by a qualified personnel in descriptor language. In the main, an inquiry represents a list of descriptors by which the documents must have been indexed or not, or, finally, have been indexed with at least one from a sublist of the list. These three types are, in fact, the expressions of the logical operators "and", "no" and "or", by which the descriptors are interconnected in an inquiry. Searching is carried out in several steps: the inquiry is decoded, the descriptors are sought in the vocabulary wherefrom their DNR is taken, then by means of these DNRs, the (periodical or cumulative) inverted file is consulted for extracting the DANs of the documents that have been indexed with the respective descriptors. The logical relations between the descriptors that have formed the retrieval inquiry are solved within these lists of DANs by the usual operations with sets: intersection for "and", inclusive for "or" and complementary for "no" operations. At last, when we have the list of the DANs of the documents that might be the answer to a retrieval inquiry, the number of documents may be the only imprinted, if it exceeds a limit imposed by the applicant, or all the documents may be printed from the response list; it is obvious that for this printout the bibliographical file may be used. The printed cards on which the answers of the retrieval inquiries are printed are sent to applicants. The retrieval inquiries are not solved in block, but one by one. So as to increase the rapidity of approach to the vocabulary and searching inverted files we use the method of access tables and the vocabulary is held in two forms: one sorted in alphabetical order and the other one in DNR increasing order, i.e. the so-called synonyms form.

In levity and for memory economy, as well as for being able to also use punched or magnetic tapes as storage media, and for retrieval flexibility, the access to the bibliographical file is sequential. The retrieval language and the retrieval moment have a conversational character, which facilitates the implementation of I.R.M.S. in a comuting system with teleprocessing.

b. I.R.M.S. Retrieval language

I.R.M.S. may use the whole set of characters of the code EBCDIC *), which are divided in four categories:

— letters: the capital and small letters of the Latin alphabet
— figures — Arabic figures
— special imprintable characters — $ % + — */ . , ; etc.
— non-printable control characters.

The language descriptors are formed from these characters through joining in chains. In the I.R.M.S. conception, a descriptor is a sequence of 20 characters, which has, of course, a certain significance. For example:

OPERATIONAL RESEARCH
1971 — 1975
RESEARCH
* I R M S *

From the semantic point of view, a descriptor may consist in a single or several words, abbreviations of words, initials, figures, etc. The words may be written in one or several languages. The set of descriptors forms the descriptor vocabulary of the system. With a view to allowing the achievement of a descriptor universality, as well as for facilitating the delimiting of their semantic spheres, I.R.M.S. provides the possibility of organizing the vocabulary as a thesaurus. The descriptors are connected within a thesaurus by various semantic fields, to other descriptors, generating in this way true hierarchies. The semantic fields and the logical and semantic relations are codified with a single character in two correspondence tables of 60 components each. This means that a thesaurus cannot use more than 60 semantic fields or 60 semantic relations. If this restriction concerning the number of semantic fields could become inconvenient in a certain field, the restriction of the number of relations to 60 is not practically constituting a restriction. Some examples of logical and semantic relations: synonymy, homonymy, antonymy, proximity, generality, restriction, etc. Except the synonymy relation, I.R.M.S. is using for the retrieval operation proper none of the mentioned or possible relations. This is the reason why the code of this relation is fixed: 0 — main synonym, 1 — secondary synonym. The remaining codes may be established by the user in terms of the application field and of the set of available characters in the respective computing system. These relations are intended for facilitating the operation of accurately defining the semantic sphere of the subject, which is the object of a retrieval trial.

The retrieval inquiries are filled in by allowing for some rules specifying their syntax and orthography:

— each retrieval inquiry should have an identification line, at least a descriptor list (line) (15 at the most) and a command specifying

*) EBCDIC = Extended Binary-Coded-Decimal Interchange Code

which logical combination among the descriptor lists existing in the inquiry should be solved at retrieval;

- a list should include: an indentifier, one of the logical operators "and", "or" and a row of maximum 10 descriptors (their order in a row is immaterial); the entire inquiry should not include more than 150 descriptors; there is a possibility of marking those descriptors for which we wish to print all the connected descriptors in the thesaurus;
- a command should include: an identifier and a logical expression consisting of the identifiers of the lists, each appearing only once, the logical operators: "and", "or" and "no" and maximum two levels of parentheses; it may also include a number establishing the number of response-documents that are printable: if the maximum limit evaluated by the applicant is exceeded, only the number of response documents is printed, otherwise the documents are also printed; where there is no such limit, only the number of response-documents is printed;
- the descriptors should appear as in the vocabulary, not through DNR, possibly eliminating the insignificant blanks (spaces) up to 20 characters;
- in the lists, the descriptors are separated through commas and with a full stop after the last one.

The introduction of a retrieval inquiry in the computing system to be solved is carried out either by punching it on cards, or by typing it on a typewriter coupled to the system. The console operator may achieve a dialogue with the system allowing: the specification of the device through which the inquiry is introduced, the immediate correction of the syntax or punctuation errors, the partial reformulation of the inquiry, the more accurate indication of the area of document codes — DAN — where the retrieval should be carried out and the indication of work end. The I.R.M.S. program achieving the retrieval of information — SRCH — disposes of a system of messages informing the operator of either an error, or of the necessity of a decision, or the acceptance of one of his decisions. The language to be used is not too advanced; it reduces itself to the essentials.

c. I.R.M.S. Documents and files design

The I.R.M.S. primary documents are: the vocabulary forms, analysis sheets and retrieval inquiries. All these documents are then punched on cards to be processed by the I.R.M.S. programs. The retrieval inquiries are the only exceptions, as they may be input from the computer console typewriter too, or from a console connected to computer by a teleprocessing system. The I.R.M.S. is not providing a format, this being left with those that are operating this system. The I.R.M.S. provides however patterns for all the three categories of primary documents.

The card pattern for vocabulary is as follows:

Columns	Symbol	Significance
1—20	D1	Descriptor 1
21—25	DNR	Descriptor 1 code
26	SF	Semantic field
27	RC	Relative code
28—47	D2	Descriptor 2
48—79	—	Free
80	OC	Operation code

For the numerical code of a descriptor — DNR — five decimal positions have been reserved, therefore the maximal number of descriptors for vocabulary is 100,000; the secondary synonyms have the same DNR as the main synonyms and therefore the effective number of descriptors may be much higher than 100,000. The operation code identifies the operation which is carried out with the respective record: 1 — insertion in the vocabulary, 2 — cancellation from vocabulary.

The card pattern for the analysis sheet has the following structure:

Columns	Symbol	Significance
1—6	DAN	Document access number
7	—	Free
8—10	CC	Card code
11—70	TX	Text
71—79	—	Free
80	OC	Operation code

Six decimal positions have been reserved for the document digital code — DAN — therefore the maximum number of documents that may enter a collection is of 1,000,000. This means that I.R.M.S. is a system that may be applied for the case of small and middle colections. The card code, for which three decimal positions have been reserved, therefore of maximum 1,000, is designed to separate the information on the analysis sheet into three groups: descriptors — up to 100, bibliographical information — up to 200, and document or its abstracts — up to 1,000; this dimensioning makes impossible the integral storage of some larger documents than 800 rows x 60 letters a row = 48,000 letters (approx. 20 pages). The operation code has the same values and significances as those for the vocabulary.

The card pattern for the retrieval inquiry is as follows:

Columns	Symbol	Significance
1–73	TX	Text
74–75	C	Code
76–80	SQ	Sequence

The text includes, in principle, a part of identification, separated by the sign = from the rest of the information. The structure of the identification part has the following aspects:

— for the identification line: ID — input from console
OO — input on punched cards
— for the descriptor lines : **, OL [,P] *, where:
$01 \leqslant xx \leqslant 15$ — represents the list proper identifier,
OL — represents the list asociated logical operator which may be OR (or), AD (and),
P — represents the printing indication of the thesaurus branches the descriptors on the list line where this indicator appears pertain to
— for command: QU [,N] — input from console,
99[,N] — input on punched cards;
$N < 10{,}000$ — represents the max. number of printable documents.

The code is the same for all the lines of the retrieval inquiries, 05, and the sequence is related to the number of an inquiry in a pack of inquiries and is the same for all the lines of an inquiry.

The files used by I.R.M.S. are of five categories:

— permanent basic files: vocabulary in alphabetic order and in DNR order, cumulative inverted search file and cumulative bibliographical file;
— periodical basic files: periodical inverted search file, periodical bibliographical file;
— auxiliary files: statistic file, erasing (corrective) file, tables (for the vocabulary and inverted files);
— intermediate files: direct searching file, update files;
— work files.

We will only present the structure of basic files.

The vocabulary logical recording has a structure similar to the following one, with the following differences: the DNR is held as a packed decimal form of 3 characters, the free space is suppressed, as well as the operation code. Thus, the logical record length is of 45 characters; the physical record represents a block of 25 logical records. Having this structure, the vocabulary

*) The brackets [ ] show the optional character of the row of letters inside; they are not appearing in the retrieval inquiry.

is sorted in ascending order: first according to D1 and DNR, generating the alphabetic form, and the second time reversely, generating the synonym form. Access to the alphabetical form of the vocabulary is obtained by means of two tables: the first holds for each character used, as first character in descriptors, the pointer in a second table, and this one, for each combination of the first two characters, the pointer in vocabulary; the identification of the other characters of a descriptor, which is searched in a vocabulary, is carried out from a much more restricted subset.

The cumulative inverted searching file, as well as the periodical one, have a logical record of the following design: the 3-character packed decimal DNR, followed by a list of 10 DANs of the documents that have been indexed with the respective descriptor, held as a binary form of 3 characters. If the document list is greater than 10, a new record of the same type should be formed. The file is sorted in ascending order according to DNR and DAN. The physical record includes 25 logical records. The access to the files is obtained by means of two tables: the first holds for each last DNR on one cylinder of the inverted file, the pointer into a second, and this one, for each ast DNR from a block the pointer in the inverted search file.

The cumulative bibliographic file, as well as the periodical one, have the logical record of the following form: binary DAN of 3 characters, CC — decimal, packed of 2 characters and the text — of 60 characters. The logical records of 65 are blocked by 10. The access is sequential.

The output documents provided by the I.R.M.S. are of two kinds: those required for maintenance of the system, i.e. the vocabulary lists and the error lists, and those including the answers to the retrieval inquiries.

For maintenance, the vocabulary is listed in three forms: alphabetical, of synonyms, organized (of thesaurus). Each of them is arranged in the printer page on 3 columns, each page being numbered and bearing the printing date. The first two forms are differing only by the sorting order: alphabetical and according to DNRs, respectively. The third form prints all the existing linkages within the thesaurus and is similar to the format of a dictionary.

The lists of errors are in connection either with the absence from the vocabulary of some descriptors by means of which documents have been ndexed or retrieval inquiries filled in, or with the absence of some documents irom the bibliographical file, etc.

d. I.R.M.S. Programs

All the I.R.M.S. programs are written in ASSEMBLER-I.B.M. system for disk operating systems — DOS. They are using for the files they operate magnetic disks only. Excepting the searching program — SRCH —, the I.R.M.S. programs are not allowing any kind of conversation with the operator. They have no control cards, the options concerning the input, processing or output being entered through JOB CONTROL.

// UPSI h — where h is a sequence of 8 binary digits having a certain significance for each program. The I.R.M.S. programs are not checking the numerical information entering the system, considering it as correct. Each program consists of a single phase and does not occupy more than 24 K bytes

from memory. For input-output operations, the LIOCS (*) macro instructions referring to the disk files and the PIOCS (**) referring to the card reader or printer are used.

The DCT1 program allows the creation and updating of the vocabulary file. Input is carried out from cards or from disk. Selection of the operation as well as of input, is carried out, as already specified, by the control // UPSI 000000XX, with the following significances for the last two binary positions:

00 — updating with input from disk
01 — updating with input from cards
10 — creation with input from disk
11 — creation with input from cards.

Irrespective of the operation which is carried out, the DCT1 program generates each time the two access tables to the vocabulary. The DCT3 and DCT4 programs are pointing the vocabulary in its three posibile forms: DCT3 — thesaurus form, DCT4 — alphabetical and synonym form (the option is defined by UPSI card).

The DCT2 program is processing or only checking the file with document indexing. The check refers only to the presence or absence from vocabulary of the descriptors used for indexing. Processing consists in generating the intermediate direct searching file and the periodical bibliographic file.

The INV1 program achieves the reversal of the sorted direct searching file, generating the periodical inverted file and its access tables, or two inverted files are merging and generating at the same time the new access tables to the merged file, as well as a statistical file that may be consulted for following up which are the more used descriptors and which are not at all or less.

The INV2 program allows to carry out alterations in the inverted files, as well as in the bibliographical ones. These alterations are of the deletions kind; they may be input either from the console, or from cards. It also generates a file with the deletions carried out.

The SRCH program solves the retrieval inquiries.

e. I.R.M.S. Operating procedures

The above specified 7 programs are not enough for the I.R.M.S. file creation, maintenance and operation. In order to be able to carry out some programs, the input or reference information requires to be edited under a certain form, as achieved by other programs which being of the multipurpose type could include them in the programming set of the operating system. It is a matter of the sorting and merging program — SORT — and of the transfer from cards to disk program — CDDK.

---

*) LIOCS — Logical Input-Output Control System
**) PIOCS — Physical Input-Output Control System

Program calling is carried out according the norms stated by the disk operating system — DOS — i.e. by using the JOB CONTROL program language for defining the files and establishing the physical devices on which magnetic media have been installed where the files are installed or with whom we work, as well as for determining the working options and for program calling for execution.

When operating an I.R.M.S. system, the following main phases are distinguished:

— the vocabulary creation and updating;

— the creation of the basic, inverted searching and bibliographic files;

— maintenance of the basic files;

— solving of the retrieval inquiries.

Flow chart for the vocabulary creation, updating and listing is given in Fig. 1.

Flow chart from Fig. 2 may be used with a view to obtaining the basic files: the bibliographic and inverted search ones:

The inquiry retrieval may be carried out periodically or not, according to the flow chart from Fig. 3.

WHAT IS NEW WITH S.I.R.P.I. AS COMPARED TO I.R.M.S.?

First of all, the higher organization, on several levels, of the vocabulary and of the collection of documents and even of the information within a document, with a view of self-editing of volumes (reviews) exhibiting the respective collections or the collection periodical entries in an intelligible form, which is easier to be understood and to be checked. Secondly, this self-editing itself, through programs, of the volumes with the vocabulary and the periodical entries into the system at the computer printer. These volumes are required in many applications to be diffused in many places; the repeated typing operations or the lithographic operations, or even the printing may be replaced in this way by a more simpler operation: xerography multiplication. Thirdly, the information entering the system is checked in the numerical fields by thus avoiding the number of abnormal terminations of the I.R.M.S. programs due to this kind of errors. Fourthly, the sequential magnetic devices (magnetic tapes) have been also included in the system flow, as these are not only making possible the operation of such a system in a computing system of only two disk units, but they are also considerably increasing the storage capacity and reducing the expenses for the information supports.

The first version of S.I.R.P.I. has modified none of the 7 I.R.M.S. programs; it has aimed at completing the system with new functions, and therefore with new programs for operating the I.R.M.S. existing basic files, without requiring a great many new information. The 7 new programs, written so as to achieve: the indexing control and updating (correction), for editing the descriptor vocabulary, the information bulletins with author and descriptor index positions and the cumulative author and descriptor indices, have been drawn up in a different way as compared to the I.R.M.S. programs, with the purpose of obtaining a greater flexibility in their operation.

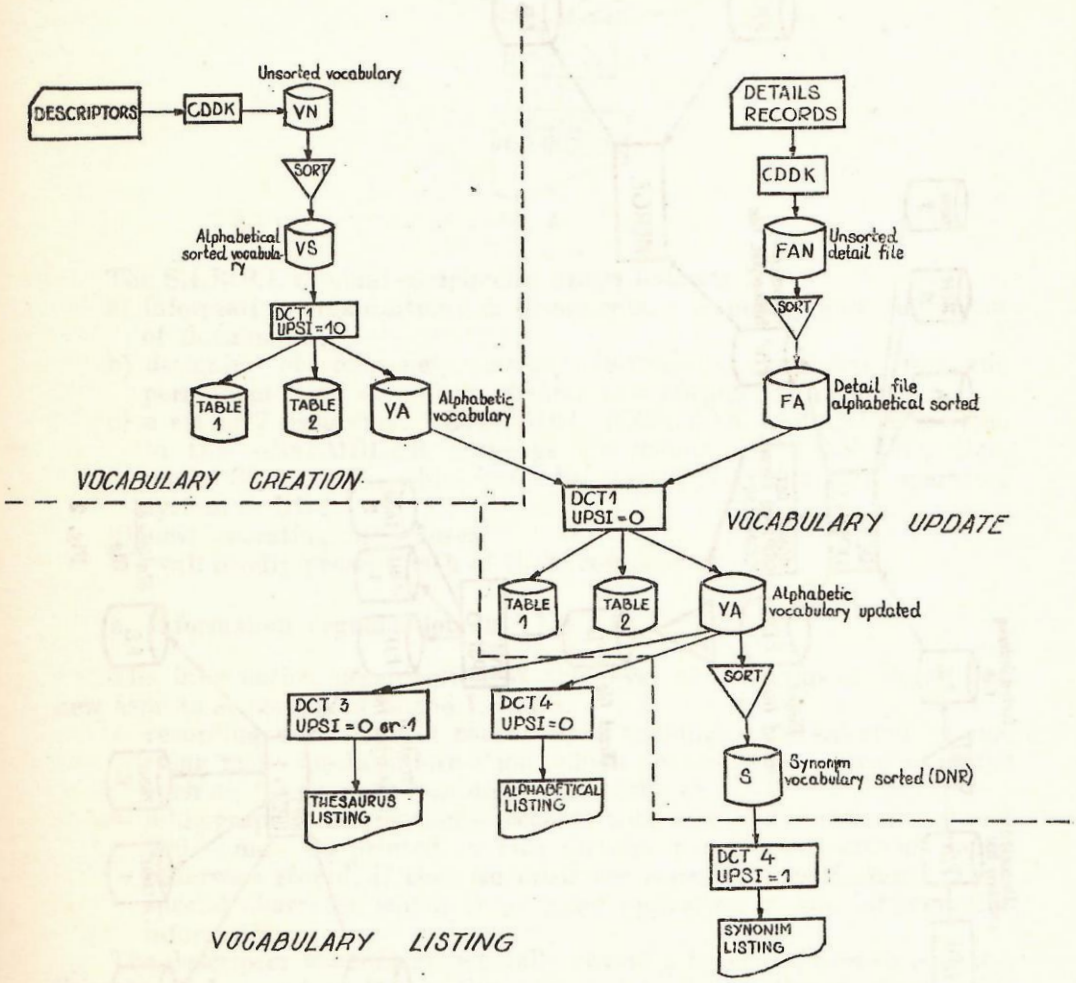

DESCRIPTORS
CDDK
Unsorted vocabulary
VN
SORT
Alphabetical sorted vocabulary
VS
DCT1 UPSI=10
TABLE 1
TABLE 2
VA
Alphabetic vocabulary
VOCABULARY CREATION
DETAILS RECORDS
CDDK
FAN
Unsorted detail file
SORT
FA
Detail file alphabetical sorted
DCT1 UPSI=0
VOCABULARY UPDATE
TABLE 1
TABLE 2
YA
Alphabetic vocabulary updated
SORT
DCT 3 UPSI=0 or 1
DCT4 UPSI=0
S
Synonim vocabulary sorted (DNR)
THESAURUS LISTING
ALPHABETICAL LISTING
DCT 4 UPSI=1
SYNONIM LISTING
VOCABULARY LISTING

Fig. 1

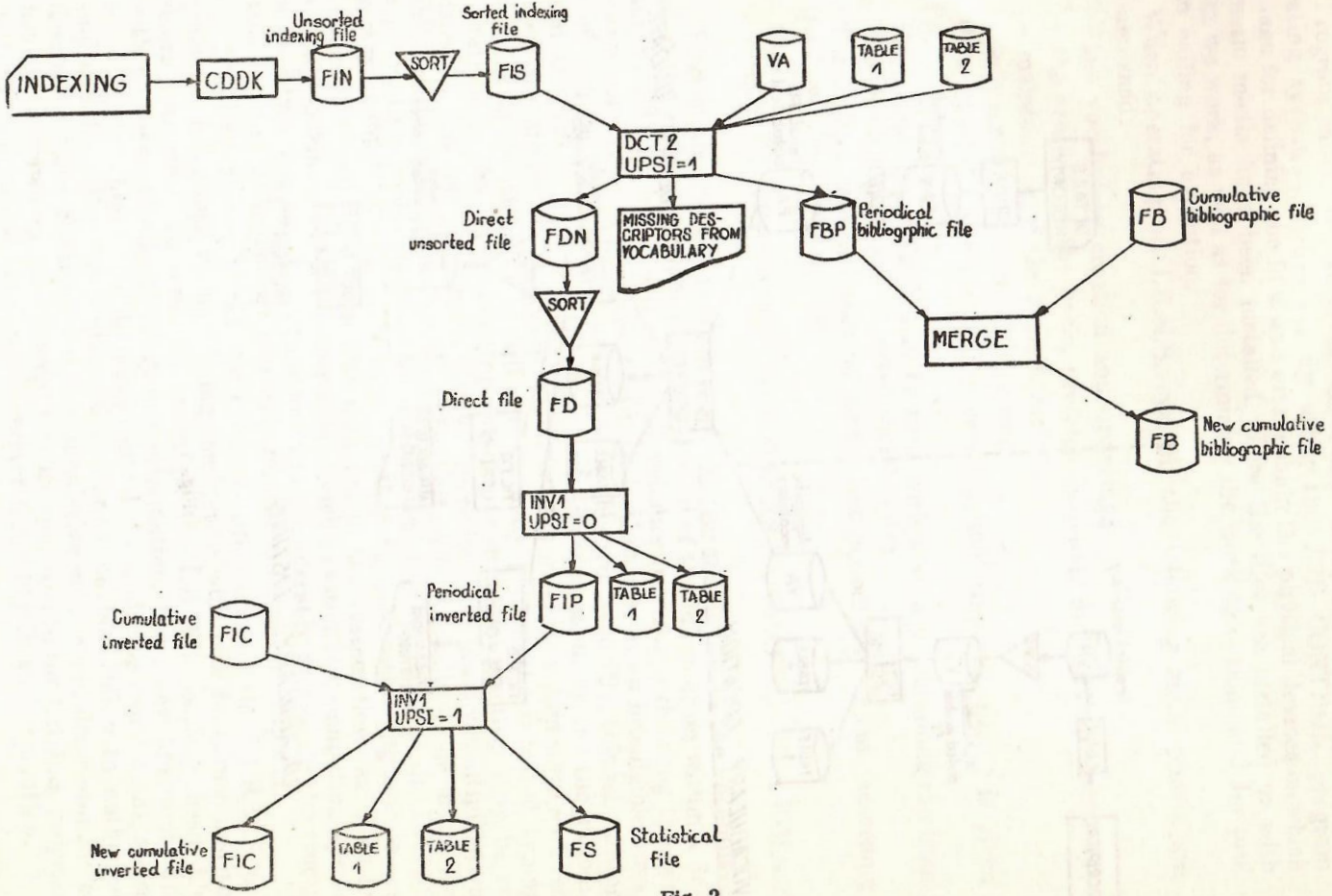

INDEXING
CDDK
Unsorted indexing file
FIN
SORT
Sorted indexing file
FIS
VA
TABLE 1
TABLE 2
DCT 2
UPSI=1
Direct unsorted file
FDN
MISSING DESCRIPTORS FROM VOCABULARY
FBP
Periodical bibliographic file
FB
Cumulative bibliographic file
SORT
MERGE
Direct file
FD
FB
New cumulative bibliographic file
INV1
UPSI=0
Periodical inverted file
FIP
TABLE 1
TABLE 2
Cumulative inverted file
FIC
INV1
UPSI=1
New cumulative inverted file
FIC
TABLE 1
TABLE 2
FS
Statistical file

Fig. 2

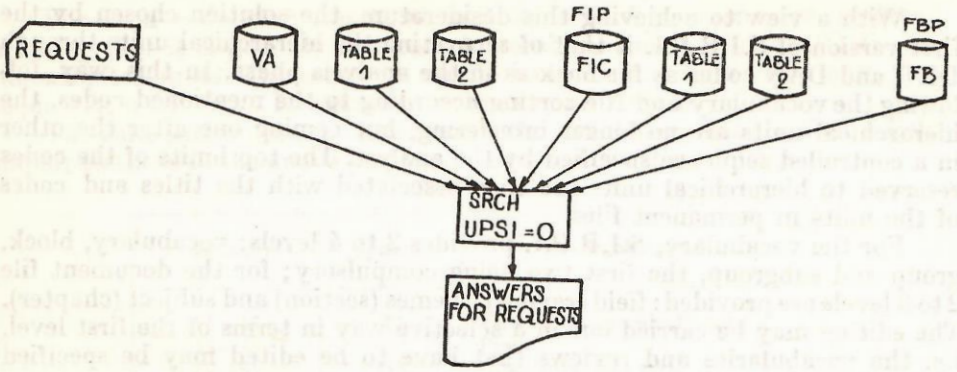


Fig. 3

The S.I.R.P.I. original components are as follows:

a) information organization: in documents, vocabulary and collections of documents;
b) designing of some new primary documents, temporary files and permanent files, as well as of some new output documents;
c) a set of 7 programs: DCT5, BIB1, REF1,2,3,4, and INDX, written in the ASSEMBLER language for computers I.B.M./360, from model 30 upwards, which may be carried out in a disk operating system — DOS;
d) new operating procedures.

We will briefly present each of these components.

a. Information organization in S.I.R.P.I.

The information organization at the level of a document shows two new aspests as compared to the I.R.M.S. :

— recording with code of card 100, if existing, is interpreted as carrying more special information, which deserves to generate an index, such as, for example, the document authors;
— bibliographic information — records with card code between 100 and 200 — may be printed in two distinct paragraphs, without being otherwise stored, if the two areas are separated, for instance, by a special character which is no more appearing in the bibliographic information.

The descriptor vocabulary generally covers a branch of human activity. Criteria might exist to divide this branch into several fields, these fields may be, in turn, divided in several subdivisions, and so on, generating a hierarchical structure, each unit from this hierarchy having its own individuality. This is also happening with the document collection The printing of the vocabulary of descriptors and of the collection or of the periodical inputs in the collection in such a form, where the descriptors and the documents, respectively, are grouped in hierarchical units bearing significant titles, is a necessary operation both for those that are entrusted with the system maintenance, and for those that are turning it to account.

With a view to achieving this desideratum, the solution chosen by the first version of S.I.R.P.I. is that of separating the hierarchical units through DNR and DAN codes as far back as in the analysis phase. In this way, following the vocabulary and file sorting according to the mentioned codes, the hierarchical units are no longer interfering, but coming one after the other in a controlled sequence specified by the analyst. The top limits of the codes reserved to hierarchical units are then associated with the titles and codes of the units in permanent files.

For the vocabulary, S.I.R.P.I. provides 2 to 4 levels: vocabulary, block, group and subgroup, the first two being compulsory; for the document file 2 to 3 levels are provided: field (review), themes (section) and subject (chapter). The editing may be carried out in a selective way in terms of the first level, i.e. the vocabularies and reviews that have to be edited may be specified when the system includes several vocabularies or several reviews.

The inside covers of the respective volumes may be also edited through programs, as well as their contents, an introduction, and the hierarchical structure specifying for each unit the page where it is to be found within the volume. Moreover, two indices are provided to be edited with reviews: the descriptor index and, if need be, the special index generated by the information included in the records — with the cord code 100 from each document (for exemple, the author index). The possibility of yearly editing of some cumulative indices in separate volumes has been provided.

b. Document and S.I.R.P.I. file design

The primary documents appearing in plus as compared to those of the I.R.M.S. are the forms for titles. These are not standardized. There are, however, patterns for punching this information on cards. For the vocabulary, two distinct title files are provided within this first S.I.R.P.I. version: one for printing the hierarchical structure and the other for their insertion among descriptors; their separation was required by the fact that the vocabulary is edited on two columns on each page, while the hierarchical structure is edited on a single column; hence the difference in length of the two file records. Besides, the titles that are inserted among descriptors are ampler, also including various comments of the kind of some references to other hierarchical units.

The card pattern for the vocabulary hierarchical structure is as follows:

Columns	Symbol	Significance
1—2	V	Vocabulary code
3—4	B	Block code
5—6	G	Group code
7—8	SG	Subgroup code
9	SQ	Sequence
10	—	Free
11—70	TX	Title of hierarchical unit
71—80	—	Free

For each code of hierarchical unit 2 decimal positions are reserved, meaning that it is possible that each hierarchical unit has max. 100 directly subordinated units. The sequence should be a decimal figure or blank. Consequently, a title may be extended on maximum 10 lines.

The card pattern for the titles inside the vocabulary has the first 6 fields similar to those in the above pattern, but further on there follows:

Columns	Symbol	Significance
11—45	TX	Title of hierarchical unit
45—46	—	Free
47—51	UDNR	Descriptor last code
52—80	—	Free

The card pattern for the file hierarchical structure is as follows:

Columns	Code	Significance
1—2	D	Field code
3—4	T	Themes code
5—6	S	Subject code
7	SQ	Sequence
8—10	—	Free
11—55	TX	Title of hierarchical unit
56—57	—	Free
58—63	UDAN	Last document code
64—80	—	Free

Concerning the files, S.I.R.P.I. furnishes the following novelties:

— permanent basic files: the three title files, cumulative descriptor index file, comulative special index file;
— periodical basic files: periodical descriptor index file, periodical special index file;
— auxiliary files: files for control cards;
— intermediate files: updating file for bibliographical file, non-sorted index files;
— work file.

We will briefly present the structure of basic files.

The three title files have the structure of logical record similar to those shown in the card patterns, with the following remarks:

— the free spaces are suppressed;
— the last descriptor code is kept in packed decimal form of 3 characters;
— following the last document code two characters are left free for receiving, during editing, the page — in packed decimal form — where the respective hierarchic unit begins.

The files of descriptor index and special (author) index both cumulative and periodical, have a logical similar record: field (review) code, DAN and descriptor — 20 characters — or special information (author) — 22 characters.

All these basic files have large physical records — within 3,000 and 3,600 characters — for ensurig an as large density of information as possible on the magnetic media used and for requiring as few input-output operations as possible, which are slower. The only exception is that of the file with hierarchical structure of the collection where, as it has been declared to be a work file, the logical record have been no more blocked. The title files are stored on magnetic disks, while the index files may be either on disks, or on magnetic tapes.

The final documents proper to S.I.R.P.I. are: the list of control and errors for indexing and the volumes with descriptor vocabulary, with the collection fields (reviews) and the cumulative indices.

The control list and for indexing errors includes: list of all the records from the indexing file, messages for errors in numerical fields — DAN and card code —, records with the same numerical fields; when updating the indexing, add still the messages anouncing the trial to insert a record with the same numerical fields with an already existing record in the file, as well as the trial to erase a non-existing record in the file.

The vocabulary is printed on two columns; the pages bear on the outer corners the block title including the respective descriptors. The blocks are on separate pages; the blocks title always appears alone on the page, on an odd page; the blocks are numbered with Roman figures. Any subgroup should belong to a group and any group to a block; the groups may not have always subgroups, but the blocks should have at least a group.

The collection reviews, especially if there are more, may have the inner covers also programmed edited. The documents are automatically numbered in the code of each review; each document should pertain to a subject (chapter), or theme (section); any subject should belong to a theme and any theme to a field (review); any field should have at least a theme, but not all theme should have a subject.

c. S.I.R.P.I. Programs

All the S.I.R.P.I. programs are written in ASSEMBLER — I.B.M. system, for disk operating systems — DOS. They use for the files they operate both the disks and the magnetic tapes. Almost all the programs allow certain conversations with the operator, as a rule, in order to announce the exceptional situations occurring in the system and controlled by programs, and, sometimes, so that the operator takes the decision imposed by the new situation. All the programs have control cards, establishing: the processing duties, selection of input or output devices, as well as certain processing parameters. All the programs are controlling the syntax of their proper control cards and work begins only if there is no error on them. Also, all the programs of execution and announce the errors and the place where these are occurring.

All the S.I.R.P.I. programs are made up of a single phase, which is not filling more than 44 K. The I.R.M.S. conventions have been held for input and output, i.e. LIOCS for tape and magnetic disks and PIOCS for reader and card-punch and for printer. All the S.I.R.P.I. programs are editing at the execution end a small summary of the manner how the duties they had have been fulfilled.

The DCT5 program has the duty to edit the program in a form allowing the xerographying by the printer and then their binding in volumes. The inside cover is edited, as well as the contents, introduction, vocabulary hierarhical structure, vocabulary proper and an appendix that belongs in fact to the vocabulary also, but which includes more special descriptors, such as: geographical denominations, collection reviews, calendar years used, main processed reviews, etc. Within each hierarchical unit, the descriptors appear in alphabetical order, not in code order. Sorting is carried out in memory from an area where no more than 2,000 descriptors may be included, and this is the reason why no hierarchical unit from the last level should include more than 2,000 descriptors.

The BIB1 program may carry out multiple functions. It is a service program. Its duty consists, in the main, in the indexing control and updating. Delimitation of its duties is carried out by means of its control cards. BIB 1 may carry out in the indexing file or in that of indexing updating controls concerning: numerical fields, record duplication (according to the numerical fields) insertion and erasing of some already existing records or, on the contrary, which are not existing in the file to be updated.

The BIB1 program may list these files, copy the correct records on disk, erase from the file to be updated all that exists within two records; it has the imput either on tape, or on disk.

The REF1 program is designed to achieve the selective editing of the reviews with the collection documents. But, moreover, REF 1 generates the author file required for editing the author index. The bibliographical file, which is the main program input, may be located either on tape, or on magnetic disk; the same thing holds for the author file. The REF1 program also fills up the pages in the title file for the file hierarchy and transmits its control cards, when modified and filled in with the head page of the author index, to an auxiliary file for control cards.

The REF2 program achieves the editing of author index for each review that was selected, as well as the summary from each review.

The input files used by the REF2 programs are as follows:

— author file (alphabetically sorted) on tape;

— title file on disk;

— auxiliary file for control cards, on disk.

The program accepts the input of control cards from reader and transmits the control cards that are modified with the head pages of the descriptor index.

The REF3 program creates the descriptor file for those having been used for indexing the documents in the reviews selected for editing.

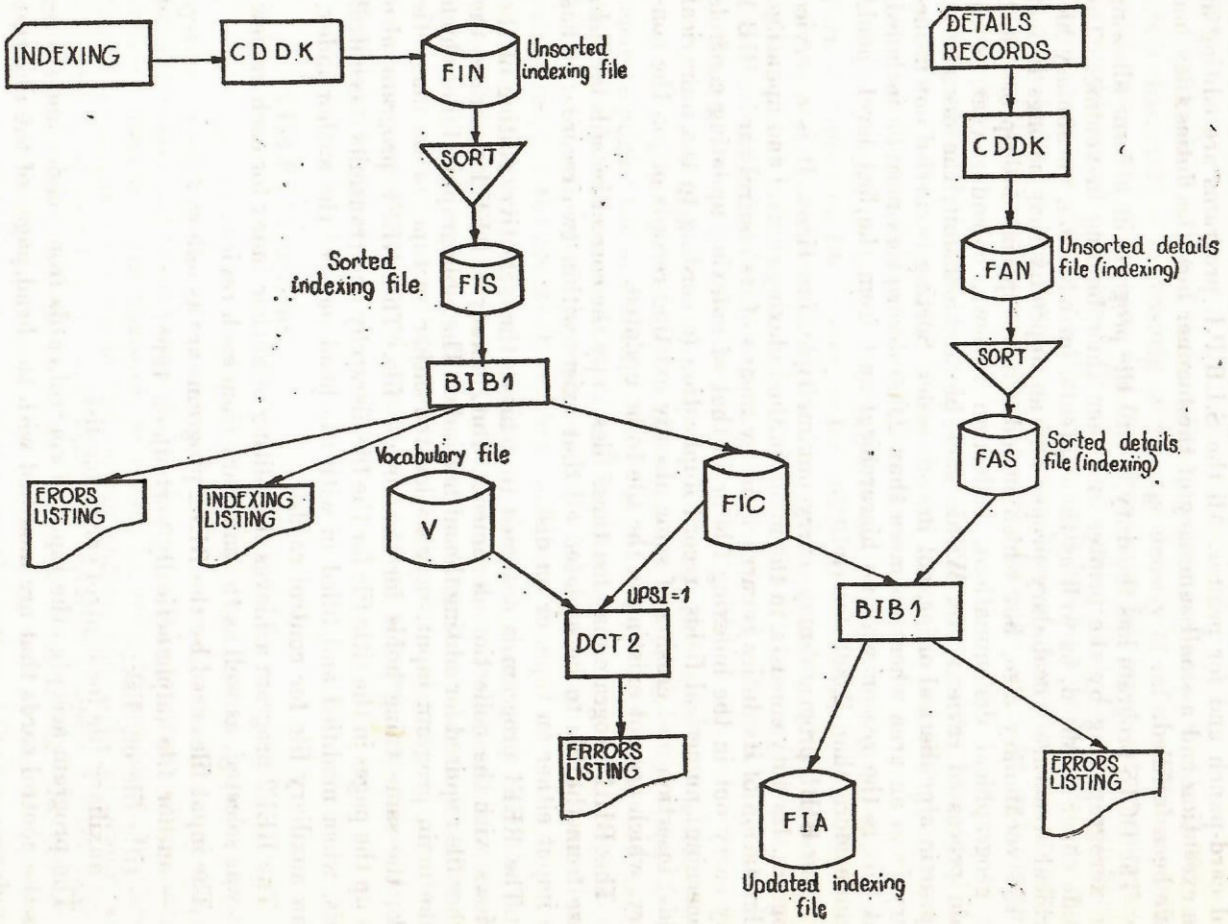

INDEXING
CDDK
FIN
Unsorted indexing file
SORT
Sorted indexing file
FIS
BIB1
EROR'S LISTING
INDEXING LISTING
Vocabulary file
V
FIC
UPSI=1
DCT2
ERRORS LISTING
DETAILS RECORDS
CDDK
FAN
Unsorted details file (indexing)
SORT
FAS
Sorted details file (indexing)
BIB1
FIA
Updated indexing file
ERRORS LISTING

Fig. 4

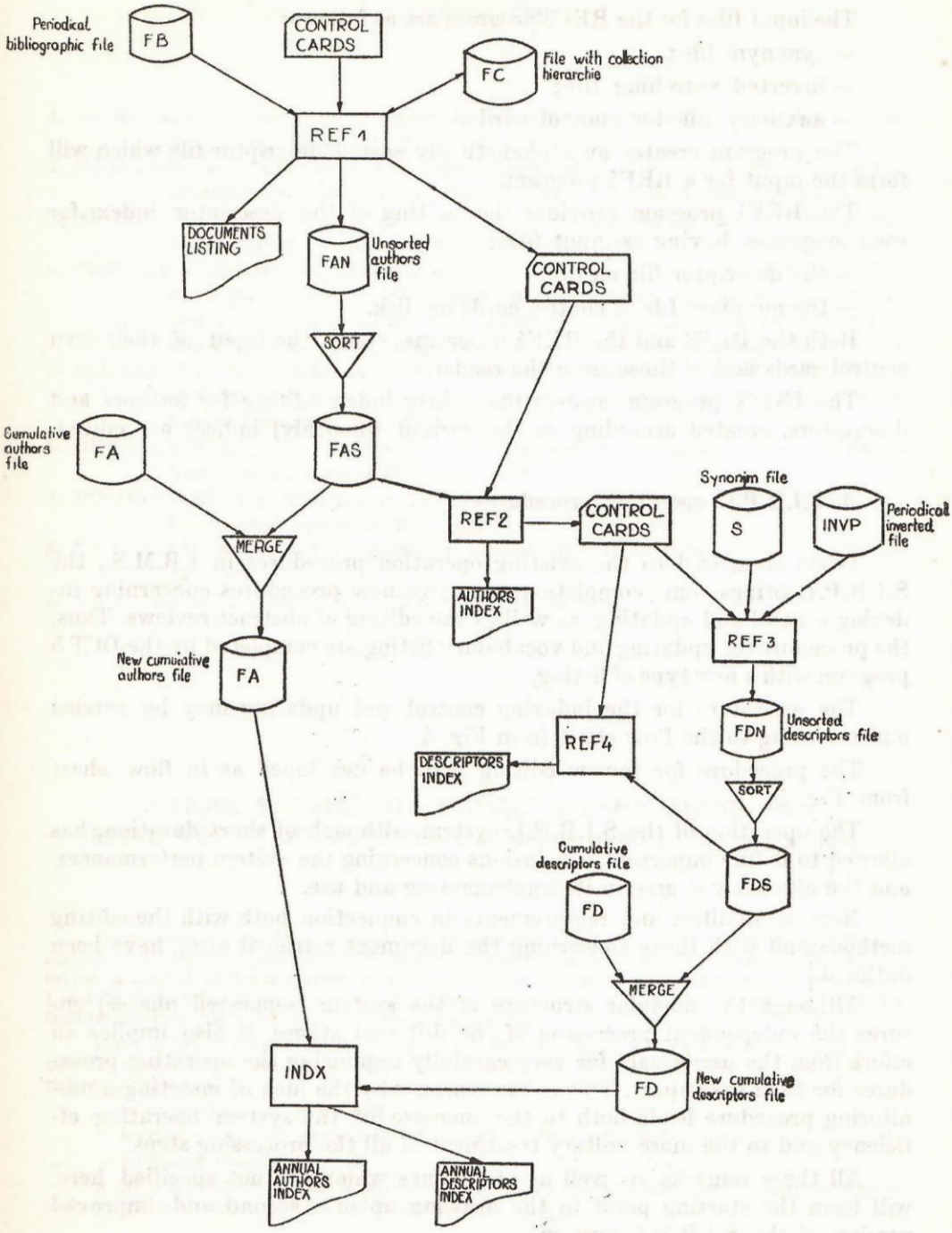

Periodical bibliographic file
FB
CONTROL CARDS
FC
File with collection hierarchie
REF 1
DOCUMENTS LISTING
FAN
Unsorted authors file
CONTROL CARDS
SORT
Cumulative authors file
FA
FAS
Synonim file
S
INVP
Periodical inverted file
REF2
CONTROL CARDS
MERGE
AUTHORS INDEX
REF 3
New cumulative authors file
FA
FDN
Unsorted descriptors file
REF4
DESCRIPTORS INDEX
SORT
Cumulative descriptors file
FD
FDS
MERGE
INDX
FD
New cumulative descriptors file
ANNUAL AUTHORS INDEX
ANNUAL DESCRIPTORS INDEX

Fig. 5

The input files for the REF3 program are as follows:

— synonym file;

— inverted searching file;

— auxiliary file for control cards.

The program creates an alphabetically sorted descriptor file which will form the input for a REF4 program.

The REF4 program provides the editing of the descriptor index for each magazine, having as input files:

— the descriptor file on tape;

— the auxiliary file of control cards on disk.

Both the REF3 and the REF4 programs accept the input of their own control cards and of those from the reader.

The INDX program ensures the yearly index editing for authors and descriptors, created according as the current (monthly) indices are edited.

d. S.I.R.P.I. operating procedures

When compared to the existing operation procedures in I.R.M.S., the S.I.R.P.I. brings some completions and two new procedures concerning indexing control and updating, as well as the editing of abstract reviews. Thus, the procedure for updating and vocabulary listing are completed by the DCT 5 program with a new type of listing.

The procedure for the indexing control and updating may be carried out according to the flow chart from Fig. 4.

The procedure for review editing may be developed as in flow chart from Fig. 5.

The operation of the S.I.R.P.I. system, although of short duration, has allowed to define important conclusions concerning the system performances, and the efficiency degree in its implementing and use.

New possibilities and requirements in connection both with the editing methods and with those concerning the document retrieval also, have been outlined.

Although the modular structure of the system (separated phases) ensures the independent processing of the different stages, it also implies an effort from the user's part for very carefully organizing the operating procedures for each case apart. This is the reason why the idea of inserting a monitoring procedure leads both to the increase of the system operating efficiency and to the more unitary treatment of all the processing steps.

All these remarks, as well as other ones which are not specified here, will form the starting point in the drawing up of a second and improved version of the S.I.R.P.I. system.

REFERENCES

1. NEGOITA, C. V.: *Modelarea matematică în documentare* (Mathematical simulation in documentation). Documentary synthesis, IDT, 1971.
2. MIHAILOV, A. I., CEORNÎI, A. I., GHILEAREVSKI, R. S.: *Informatica documentară* (Documentary Information). Translated from Russian, Editura Stiințifică, 1970.
3. NEGOITA, C. V.: *Sisteme de înmagazinare și regăsire a informațiilor* (Information Storage and Retrieval Systems). Editura Academiei, 1970.
4. TRIFU, I., PESCARU, V., LEO-MIZA, S., IONESCU-BUJOR, M., DUMITRESCU, Vl.: *S.I.R.P.I. — Sistem de Înmagazinare, Regăsire și Prezentare a Informației* (Information Storage, Retrieval and Editing System — S.I.R.P.I.). (Internal use publication CIDAS and Computing Centre of the Economic Cybernetics Department).
5. ALEXANDRU, T. I.: *Rolul informării documentare în dezvoltarea cercetării științifice* (Role of documentary information in the scientifical investigation development). Documentary synthesis, IDT, 1971.
6. DE SOLLA PRICE, D.J.: *Știință mică, știință mare* (Small science, great science). Translation from English, Editura Stiințifică, 1971.
7. MOLDOVEANU, V.: *Metode pentru indexarea documentelor* (Methods for document indexing). Documentary synthesis, IDT, 1971.
8. * * *: *I.B.M. System/360 — Information Retrieval and Management System, Application Description.*
9. * * * : *I.B.M. System*/360 — I.R.M.S. *Operation Manual.*

S.I.R.P.I. SYSTEME AUTOMATIQUE D'EMMAGASINER, DE RETROUVER ET DE PRÉSENTER L'INFORMATION (première version)

On présente un système automatique d'emmagasiner, de retrouver et de présenter l'information — S.I.R.P.I., réalisé au Centre de calcul appartenant à la Faculté de cybernétique économique. S.I.R.P.I. utilise le système pour gérer et retrouver l'information- I.R.M.S. — de l'I.B.M. Les auteurs présentent les principaux aspects techniques de l'I.R.M.S., ainsi que les nouvelles facilités prévues par S.I.R.P.I. — la première version.

Dans sa forme actuelle, utilisant les mêmes données, S.I.R.P.I. peut réaliser tant la sélection des documents relevants par rapport à un sujet, que l'édition complète de trois types de volumes, comprenant:

— le vocabulaire des mots-clefs;

— les documents emmagasinés périodiquement, un index des mots-clefs et un index spécial;

— les index cumulatifs.

S.I.R.P.I. peut être appliqué pour tout système de calcul I.B.M. 360, à partir du modèle 30F, qui dispose d'un système d'opérer sur disques -D.O.S.

S.I.R.P.I. — АВТОМАТИЧЕСКАЯ СИСТЕМА СКЛАДИРОВАНИЯ, НАХОЖДЕНИЯ И ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИИ

В настоящей статье представляется автоматическая система складирования, нахождения и представления информации — S.I.R.P.I., осуществленная в Расчетном центре кафедры экономической кибернетики. S.I.R.P.I. имеет в своей основе систему манипулирования и нахождения информации I.R.M.S. фирмы I.B.M. В статье излагаются важнейшие технические аспекты системы I.R.M.S., а также новые облегчения работы, предусматриваемые системой S.I.R.P.I. в первом ее варианте.

В своей настоящей форме, используя ту же основу данных, система S.I.R.P.I. может осуществлять как селекцию выявляющих какой-либо субъект документов, так и полное издание трех типов томов, охватывающих:

— вокабуляр ключевых слов.
— периодически складируемые документы вместе с индексом ключевых слов и специальным индексом,
— кумулятивные индексы.

Система S.I.R.P.I. может быть включена в любую систему расчета I.B.M. 360, начиная от модели 30 F, располагающую системой оперирования на дисках — DOS.