Közlemenyek 30/1984 61-71

SOME EXPERIENCES IN THE DEVELOPMENT OF DBMS [1]

P. DIPOTET

Institute of Mathematics, Cybernetics
and Computing Sciences
Havana, Cuba

Both, United Nations REPORT-71 and the first Intergovernmental Conference on Strategies and Policies (SPIN) organized by UNESCO and IBI in Spain in 1978, underlined the importance of developing a strategy for informatics in order to make the best use of the domestics resources of the country.

In Cuba, a national organization, "The Institute for Management Information Systems and Computing Techniques", has been created to coordinate cuban efforts in research, development and applications of computer science (INSAC).

Some important points in Cuban Informatics strategy are:

- the high priority given to the application of informatics in the development and control of the national economic plan;
- the development of the national statistical information system;
- the development of some special mini and microprocessor systems and software packages applied to some services and mainly to health services;
- the R/D works in software and hardware and the education and training of computer specialists supporting the former three application lines.

Next we present some of the most important experiences and approaches to the development of DBMS in our country.

---

[1] Received on April, 1982

APPLICATIONS OF DBMS TO PLAN PROCESSING

In the Electronic Processing Division of JUCEPLAN (Cuban Central Planning Board) a data model together with its retrieval language has been developed in order to be used by informatics experts working in the United Systems for Plan Processing (SUPP).

SUPP objectives are*:

- To reduce the global data processing time for daily tasks.
- To normalize the supports and media for developing, implementing and processing the functional subsystems.
- To unify the Computer Center's data base, with the possibility of being up-dated and consulted through diaplays.
- To achieve data integrity and protection in all data processing phases and stages.
- To improve the working efficiency of the whole computer center's staff.

Good ideas on relational data models have been used in order to derive the data model to be used by computer specialists in the Computer Center to implement the new processing demands for the information contained in the data base.

In SUPP a sharp differentiation between retrieval and up-dating language is done, in order to reduce the requirements stated to each one of them.

The former fact and the possibility of non-instantaneous retrieval allow us to simplify the computer implementation of the data model and their corresponding languages.

The set of relations present in SUPP may be divided in the following way:

- The classifiers, formed by the list of codes of the objects to be identified by the system, their descriptions and other informations associated to them.

---

* Notes are taken from: BARRERA J. et. al. "Use of the Relational Model within the Data Base of Plan Processing System". Economia Planificada III/2/1980, La Habana.

The classifiers are relations with simple primary keys (only one domain)

- The planning indicators, formed by the ecomonical data needed for the elaboration of the plan. They are identified by the code sets defined in the classifiers.
- The attributes of the planning indicators. They are formed by other informations needed by the exhaustive definition of the indicators. They are identified by the codes defined in the classifiers.
- A combination of the former relations.

The first three subsets are the updatable relations.

These subsets allow any information to be stored or retrieved in SUPP, to be in the third normal form defined by Codd.

OPERATIONS IN THE RETRIEVAL LANGUAGE

The retrieval language is an algebraic one. The proposed operations are a selection of the ones described by Codd. There is also the operation "transposition". This operation set allow us to derive any meaningful relation using only the relations stored in SUPP.

The operations with only one relation are: projection, restriction and transposition. We will define the last one. The others are similar to Codd's ones.

The transposition operation allows us to change the representation of the relation. There are two variants:

- to disgregate in various M-tuples a group of indicators defined in the same N-tuple;
- to fuse in one N-tuple various indicators defined in one group of M-tuples.

The operations with two relations are: union, intersection and difference which are very similar to Codd's ones.

PERSPECTIVES

A group of specialists of this organization (JUCEPLAN), is developing (under the scientific direction of Academic Lavrov) a DBMS using some ideas of Codd's relational model. There are some interesting results in this work which is expected to be finished in the second half of 1983.

Another group is working in the improvement of SUPP present version. They are dealing with some semantic problems related to the stored information. They are also trying to implement some concurrent processing facilities and the bank of method concepts.

A complete information about these works is possible to read in: "Using the Relational Model within the Data Base of Plan Processing System", Rev. Economia Planificada III/2.

APPLICATION OF DBMS TO CUBAN STATISTICAL OFFICE

The aim of the work (1) is to set Cuban requirements for statistical data bank considering GDR experiencies and our conditions and planned development of statistical service in Cuba.

The first thing to be considered is that the data bank should be constructed under the frame of the Automated System for State Statistics of Cuba, a project being designed now. The second aspect is the five year strategy plan for the development of statistical service that was recently approved in the Cuban State Committee for Statistics.

The work brings some results in the field of data bank, information modelling, general architecture, and a methodology for the formal description of the informative model.

The proposed topics being considered in this work are:

1. Definition of Cuban requirements for the statistical data bank from the point of view of the information needed (indicators, micro- and macro-data, registers etc.) and form the point of view of user requirements

for output results, evaluation and analysis of data etc.. At this point the statistical branch to be considered in the data bank (industry, buildings etc.) has to be decided.
To carry out this work, the experience of GDR data bank is considered, specially in the field of data bank statistical service.

2. The actual state and future trends in the development of data banks. After having the general requirements for Cuban data bank, a review of the application of data banks, - mainly in CMEA countries -, is done in order to gather experiences and to point out the main trends in this field.

3. A definition of a databank management system as a high integrated system for storing, retrieval, evaluation and analysis of statistical data should be given.

   See Hernández, O.; Lastra, O.; "Definition of Requirements for the Cuban Statistical Data Bank System" - Cuban Statistical Office. March. 1981.
   In this sense the main idea was pointed out in paper "A brief contribution to the study of statistical information, systems architecture" presented in ISIS'78 seminar containing basic theoretical principles of what has been recently developed as interface between data bases and bank of methods.
   An important result of this is the definition of the informative model as the formal description of concepts, semantic of data, user requirements for evaluation and analysis of data etc.. This theoretical part would not be the main part of the work but it is considered important to give support to the rest of the work.

4. Development of a methodology for the description of the informative model.

5. Application of the methodology to describe the informative model for the Cuban statistical data bank, according to the requirements and goals defined in the first point. After this being done we will have the complete view of the data bank to be implemented in Cuba from the information point of view in terms that could be comprehensible for statistical users and could serve as discussion tool and as the basis for the construction of the data bank.

6. Description of the software requirements and functions of a Data-Base System for Cuban Statistic (DBSCS).

7. Advantages of SPAZ as data base management system of the DBSCS. The capability of SPAZ to store data, according to the types and requirements of the information, is quite explicit, as well as other factors, like integrity, protection etc..

The system provides such advantages for the evaluation and analysis of data stored in the data bank, as:

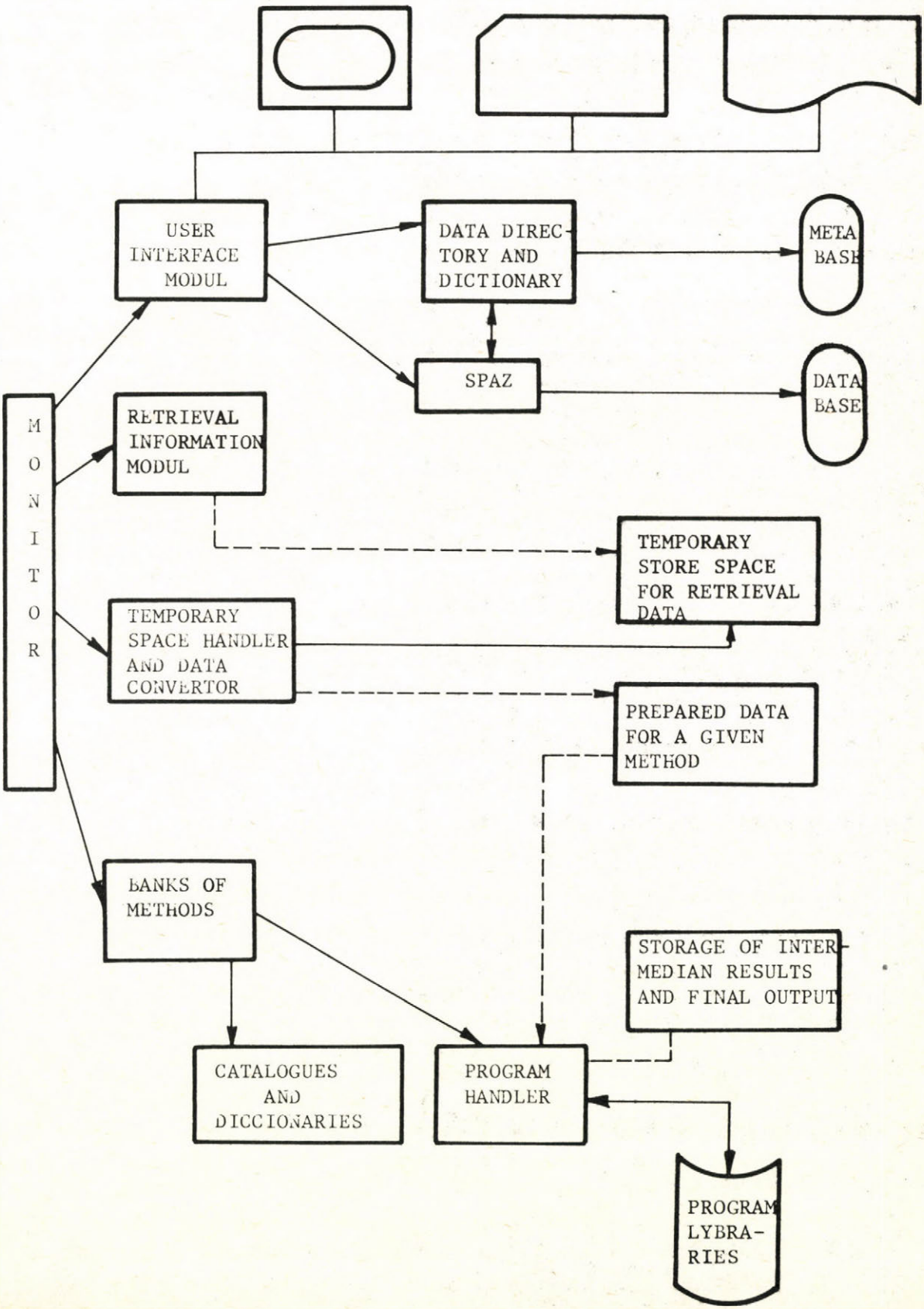
- mathematical computation and matricial algebra;
- descriptive statistics;
- index number;
- preparation and printing of complex output tables;
- graphics;
- regression and correlation analysis;
- time series analysis and forecasting;
- econometrics;

and others that can be included.

In fact, further versions of MGCE will approach a bank of methods specially in the fields of regression, time series analysis forecasting, and econometrics.

This approach will also guarantee the work to give practical results, since we have version 1.3 of MGCE working in user interface moduls and the retrieval programs would be themes for other works.

Definition of DBSCS, it general architecture and components.



DBMS IN THE CUBAN ACADEMY OF SCIENCES

In the Institute of Mathematics and Cybernetics of the Cuban Academy of Sciences, the works in DBMS are oriented in two main directions: DBMS for minicomputer and the development and/or implementation of well known systems as, for example, the Hungarian SDLA[1].

- *DBMS for minicomputers*

This work is done in cooperation with the Computer Center of the USSR Academy of Sciences. The system being developed is called SINOD.

The SINOD has twoo main blocks, an adaptive dialogue system (ADS) and a data base (BD), both are written in FORTRAN for the CM-4.

The ADS has three main modules: SYNTAX, MONITOR and INTERP. The firstone transforms the input text into a normalized format. The MONITOR analyzes the condition appearing on each line of the program written in the transformation language of the ADS and when the condition is valied then it passes the control to the INTERP.

The INTERP performs the actions corresponding to the conditions analyzed by the MONITOR. The arithmetic operators used in the transformation language are: +, -, *, /, with the same meaning they have in the common programming language.

The transformation language has also the following transfer operators:

PUT; for transferring a read data to one memory address called INPUT,

---

[1] Knuth, E., Preliminary description of SDLA. Tanulmányok, 105/1980. MTA SZTAKI, Budapest

STOR; to transfer one arithmetic operator from the memory address INPUT to the memory address OP,

and the following operators executing some actions:

EXEC; doing effective the operation indicated by the arithmetic operator,

EXIT; ending the program execution.

SINOD is being applied to the development of a data base system for planning the sugar cane harvest in a socio-economic region. We expect the system to be finished in the middle of 1983.

- *Another development of DBMS*

With the support and cooperation of the SZTAKI it has being installed in the Computer Center of the Cuban Academy of Sciences, the Hungarian version of the ADBMS, a special CODASYL type data base management system originally developed at the University of Michigan and improved at SZTAKI.

This system consists of 175 FORTRAN and 29 assembly subroutines. FORTRAN and COBOL can be used as host language of ADBMS commands. The commands can be invokated from a used program by CALL. The schema possibilities are limited compared with those of CODASYL approach.

There is no possibility to define a subschema and also to access concurrently the same data base at the same time by different users, but the ADBMS contains the main features of CODASYL and also some extensions to it. The ADBMS makes possible it to get practice in data base management systems and also it can be applied for problems having not too large mass of data (e.g. 15000-20000 records).

Formerly the strategy of locating records in ADBMS was the following: records are put into the pages of data base essentially in the order of arrival, sequentially. Modifications made in the Computing and Automation Institute in Hungary completely altered this strategy. The solution was

a hash algorithm by which the physical address of record is computed. So looking for a record does not mean several pages replacements between the storage device and the main memory, because using the computed physical address, the systems can find the page required only with one page replacement. In such a way the number of page replacements during the search of a record was decreased, and the system became more efficient.

Previously the load of a mass of records under certain circumstances required a total time of 45 minutes on an EC-1020 machine. After the modifications, under the same circunstances, the total time of the same load was 9 minutes.

DBMS IN THE INSTITUTE FOR THE DEVELOPMENT OF MANAGEMENT INFORMATION SYSTEMS AND COMPUTER TECHNIQUES (INSAC)

The main works in DBMS are related to the implementation of the SOMIS which is a DBMS with several important limitations. Direct access in SOMIS is provided by chained lists which may be structured in a hierarchical way. Users can take advantages of this hierarchies in order to avoid redundant records.

SOMIS uses direct access files which separate the data in two classes. Master files contain one of the classes and are ordered in logical sequences according to the keys. The second group of files, linked files, contain the other class of data. These files are organized as chained lists with variable record lengths. Each list is associated to a record in the master file. It is possible to access the Base using COBOL, PL/1 and ASSEMBLER.

SOMIS main difficulties are the low speed and the high frequency of maintenance. These maintenances are needed because of thw two related files and the fact that changes in one file are reflected in the other one.

The low speed is a consequence of the linked lists. The first record is accessed in a direct way (via its key) and the

others are accesses sequentially. Obviously the search by multiple keys on big files may be very slow.

For this reason SOMIS is not recommended for on-line application with big files and critical response times.
There are other important results in DBMS which are going to be published in near future.

ÖSSZEFOGLALÁS

ADATBÁZISKEZELŐ RENDSZEREK A KUBAI NÉPGAZDASÁGBAN

Perfecto Dipotet

A szerző ismerteti a Kubában használt információs rendszerek közül a legjelentősebbeket. Felhasználói szemszögből elemzi őket, de rövid funkcionális leirást ad szerkezetükről is.

ОПЫТ ПРИМЕНЕНИЯ СУБД В ЭКОНОМИКЕ КУБЫ

Перфекто Дипотет

Дается обзор информационных систем разработанных и применяемых на Кубе. Системы анализируются с точки зрения пользователя, но автор пытается дать и краткое описание функциональных характеристик.