SOFTWARE ENVIRONMENT OF AN OPEN MULTIFUNCTIONAL IBM PC COMPATIBLE INFORMATION SYSTEM FOR HIGH RESOLUTION SPECTROSCOPY

O.K. Voitsekhovskaya and S.V. Kuznetsov

Institute of Atmospheric Optics,
Siberian Branch of the Russian Academy of Sciences, Tomsk
Received December 28, 1993

An IBM PC compatible information retrieval system (IRS) for spectral characteristics of the atmospheric and trace gases in wide frequency and temperature ranges is described. A model and logical structure of an open software environment for the multifunctional IRS which generates data banks problem–oriented (spectral line parameters, absorption coefficients, molecular constants, etc.) have been developed. The structure of the IRS is arranged according to indexed sequential method as it is recommended for network data-base management system. Special attention we pay to the IRS architecture and man-machine interface. We also propose the ways of arranging data banks of IRS and some possible operations with databases on the experimental and recommended values of spectral line parameters and with the archives of molecular and spectroscopic constants. It is emphasized that relational and associative relationships dominant in this IRS.

Nowadays it is not possible to develop scientific study without an application of modern information technologies raising a level of scientific process. Continuous improvement of personal computers extends essentially technological possibilities of the information retrieval system (IRS) and requires a proper adaptation of the systems created before. New approaches to the data structure arrangement (see, for example, Ref. 1) are required to be introduced to the IRS arrangement for the problems of high resolution spectroscopy too.

The present paper describes a model and logical structure of an open software environment for the multifunctional IRS. This system generates simultaneously the problem–oriented data banks (on spectral lines parameters, absorption coefficients, molecular constants, etc.). It is the further development of IRS realized before on the ELBRUS computer.[2] Special attention is directed to describing of the model and logical structure of IRS.

This IRS includes the separate subsystems realizing their processing algorithms and man–machine interfaces and is integrated by the common menu for the operating mode selection. Data retrievals from the database are generated and remained in transfer file for the computational tasks. To save a main memory at the time of run of the computational task, a kernel of IRS is swapped out from the memory, and the task is loaded. Finishing an operation, all calculational tasks initiate a load of the kernel of IRS software environment or subsystem of IRS with which the given task is connected. A direct writing of the calculational results to the database is not allowed since the insert operation requires a rather long time and slows down the calculation process. Therefore, the calculated data are entered into the database from the transfer file which is analyzed first and then a decision to include the obtained data to the base is arrived (completely or partially writing or rejecting).

An access to IRS starts from the main menu where it is proposed to start one of the DBMS subsystems. Each subsystem of DBMS has its man–machine interface in the form of menu or map–list with the certain arguments of the database where the key arguments must be set to restrict the considered class of data in the database and to specify the operating modes for processing of all used fields (dimensions, types of errors, medium conditions, etc.). An initial run and preparation of IRS for operating with the users is began from determination of the main reference data for the man–machine interface and calculational tasks (molecules classes, molecules codes, isotopes codes, physical magnitudes and its dimensions, etc.). These operations are carried out by the system manager. All other data banks are users ones and can be changed at their discretion. It should be remembered that the notions (specifications, etc.) used for the calculational tasks can be added but a change or removal of these notions can lead to misinterpretation of the input data in the tasks developed before.

1. MANIPULATIONS ON THE DATABASES OF THE EXPERIMENTAL OR RECOMMENDED PARAMETERS OF SPECTRAL LINES

First, it should be noted that all fields in the data banks of experimental and recommended parameters of spectral lines (PSLs) have fixed dimensions of the physical magnitudes. These dimensions are determined during the first run of DBMS and can not be changed further. So, the interface must provides for a work on correction and input of PSLs in those measurement units which are used by the authors in the papers or more convenient for the user.

Block of input and correction of the experimental PSLs allows user to collect an information in the mode of manual input from any information sources. Before the start of input one needs to specify such characteristics as measurement units for the PSLs fields, types of errors for the separate fields, temperature and pressure of the medium, which remain constant throughout the process, as well as to add or attach a number of reference, to determine the isotopic modification of gas under study and buffer gas composition. As a result, the user operates under the stated conditions only, whereas the other PSLs from the archive which do not correspond to these

conditions, are inaccessible. Provision is made for specification of a solely vibrational transition to speed up the PSLs input within a single rovibrational band. In this data bank a determination of several variants of PSL under the ordered conditions and fixed rotational transition is allowed.

Block of input and correction of the recommended PSLs for the manual operating mode is arranged in the same manner, but the presence of standard medium conditions is presumed and, therefore, the medium characteristics and errors in determination of the parameters are not used. The data retrieval from the bank of the experimental PSLs can be done in accordance with the following search parameters: frequency range for the line centers, intensity range, number of reference, class and code of the molecule, code of the isotopic modification. temperature range, and pressure range. All these conditions are independent excluding the class and code of molecule and code of isotopic modification (the molecule code can be specified only for the class being zero, the code of isotopic modification can be specified only for nonzero molecule code).

Further, if it will be necessary, the filter potentials for the PSLs banks can be extended, but in this case the retrieval will be carried out somewhat slower, since in the certain cases whole archive must be browsed for sampling.

The samples for the recommended PSLs are generated analogously but the standard ambient conditions are used in this case (P = 1 atm, T = 296 K) and, therefore, it is not needed to set the temperature and pressure ranges. For fast construction of the samples mentioned above the keys are used (Fig. 1), which make it possible to reduce the time taken to find the first record with the specified values of key parameters in the file and to terminate file viewing when the set of key parameters comes out of the limits of specified values.

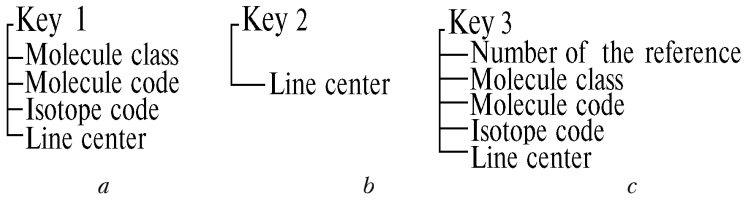

FIG. 1. Key files for the fast search in the PSL archives: speeded search on compound classification (a), on spectral interval (b), and on bibliography (c).

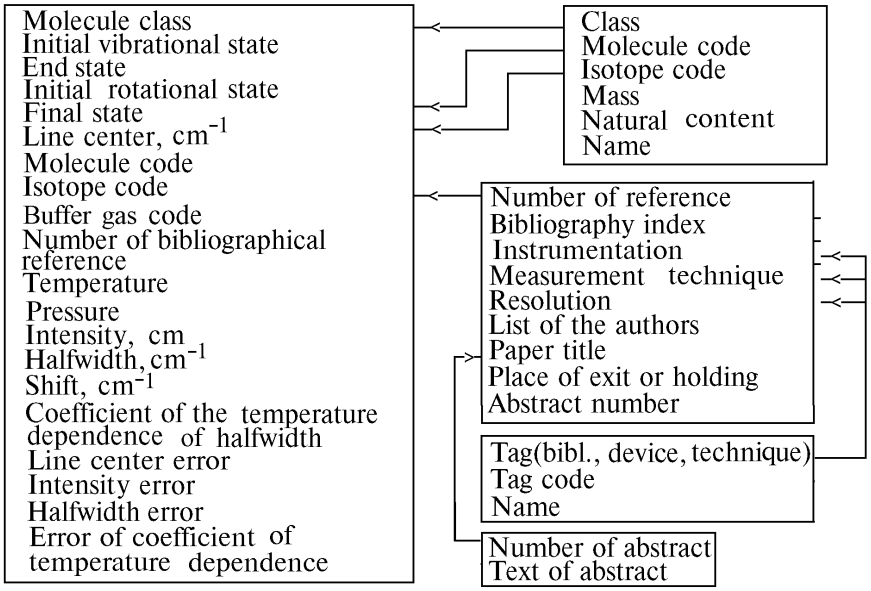

FIG. 2. Organization structure of manage for data base of experimental PSL.

At the given diagrams the field with the lower priority of filtration is correspondingly lower. Parameter sequence in the sample is determined by the key used for reading of records and the parameter sequence in the key. For example, if in the key *1* a class is only specified then the records are read only whose class coincides with the specified one. The molecule codes line up in increasing order. The same takes place for the isotope codes, in this case the records are filtered by such an order that the line centers fall in the specified frequency range, and the record sequence in the sample is determined by the increasing values of the centers for the fixed values of the class and code of molecule and isotope code. When increasing the isotope code the centers line up in the increasing order again, and so on. Figure 2 shows a generation of the archives for the experimental PSLs. An arrangement of archives for the recommended and experimental PSLs is identical. But the last archive does not require the reference, fields determining the medium conditions (temperature, pressure, code of buffer gas), and errors of measurements of line centers, intensities, halfwidths, and shifts. The sole additional field is filled with a tag of the way to obtain PSLs for the given tuple (calculation or experiment).

For the archives of the recommended PSLs the provision is made for the modes for entering the calculated PSLs to the data banks overwriting the records or separate record fields, statistical processing of the experimental PSLs, and their insert to the archive with the complete overwriting of record or certain fields of records. In this time PSLs inserted into the archive before on the basis of the experimental PSLs (or their certain fields) can not be overwrited by the calculated values. For this purpose a tag characterizing the way to obtain each information field is included to the tuple in the archive of recommended PSLs. Introduction of such a tag helps to carry out the overwriting operations.

2. MANIPULATION ON BIBLIOGRAPHICAL DATA

Bibliographical reference system serves for collection of information on used references when forming the database. It can include any publication or a short user's notices supported by the unique number for the reference being entered. The bibliographical reference is a tuple described by the block *8* in the diagram (Fig. 3).

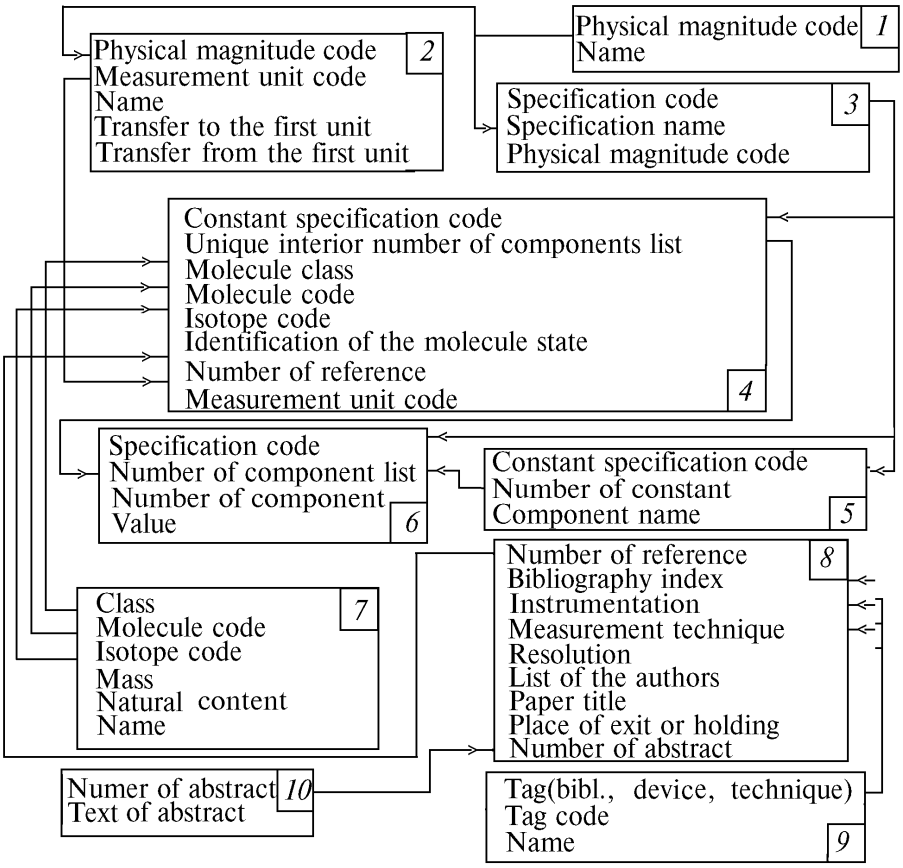

FIG. 3. Database structure for molecular and spectroscopic parameters.

During a real work every user has several hundreds of the references to be accumulate, so that a loss of disk storage is not essential. This subsystem gives a numbers of the bibliographical reference to the PSLs tuple and, moreover, can be used to construct a list of used references with the certain classification tags to browse the reference list when it is needed and add new references.

3. ARCHIVES OF THE UNIVERSAL, MOLECULAR, AND SPECTROSCOPIC CONSTANTS

Block of generation and support of the banks of physical, molecular, spectroscopic constants, and intermolecular interaction constants is needed to organize a flow of initial data to calculational tasks to obtain PSLs and others. Four levels of the data can be distinguished here: the universal physical constants, molecular constants, spectroscopic constants, and intermolecular interaction constants.

The universal constants are those physical constants which are obtained in the theoretical relations and not connected with the conditions of experiment. The same constant can be presented in different measurement units, therefore, the tuple corresponding to the universal constant must contain the unique interior number of the constant, name, value, and code of the measurement unit, which is determined by the lists of physical magnitudes and measurement units. Since every universal constant is an unique value, only one tuple for every constant must be stored. Note, that every constant is presented in any reference, therefore, the number of the reference, where its origin is noted, can be added to the tuple.

The molecular constants are those constants which are connected only with the molecule (mass, potential functions, inertia momentums, etc.) and electronic state of the molecule. These constants are determined by means of several tuples, the relationships between these tuples are illustrated by diagram (see Fig. 3).

Table I describes a function of every used tuple. Every rectangle in diagram describes one tuple and the coming arrow characterizes the fact that the given parameter is passed from the other tuple and is generated one.

TABLE I.

Tuple No.	Assignment
1	Physical magnitudes
2	Relation between the measurement units
3	Constant (constant list) type specification
4	Description of variant of constant list
5	Titles of components of constants list
6	Archive record of component values
7	Description of the molecule and isotopes classes
8	Description of the reference
9	Description of the bibliographical index, instrumentation, and measurement techniques
10	Description of added abstracts

For the molecular constants in the tuple *5* (Fig. 3) an identification of states always has a zero value, for the spectroscopic constant this parameter is determined by the vibrational state of the molecule, and for the constants of intermolecular interaction (or for the transition from the one state to the other for the same molecule) additional four fields are added. These fields determine the state of the second molecule or the second state of the molecule under study (class, codes of the molecule and isotope, and an identification of the vibrational transition). The archive of components values for molecular, spectroscopic constants, and the constants of intermolecular interaction are the common. Therefore, the tuple *6* is the main archive tuple. Distinction between the constants types in this archive is determined by the tag which is inserted to the key and allows us to use only one type of constants in every operation regime. This tag is not shown in diagram in the tuple.

4. ORGANIZATION OF THE DATA COMMUNICATION FROM THE IRS

Blocks of data forming for the calculational tasks must be developed and added with an expansion of area of used calculational tasks. It should be borne in mind that the direct access of the calculational tasks to the archive files is undesirable in the many cases due to the fact that direct access to the archive files is not nearly so rapid as that which could be organized by an algorithm of calculational task if the using the main memory and disk space at an optimum. For this purpose IRS presents the possibilities to sample the data from the archive in the most convenient form for the calculational tasks. In this case the samples can be generated as either the text or binary fails.

Designing new task one should take into account the availability of needed information in the database, and with deficit of this information it is necessary to determine the new constants (adding rather than deleting the old ones) and then only to realize an algorithm of the initial data input to the task and algorithm of output of the results to the file of standard structure which will be added to the database further.

5. CALCULATIONAL DATA INPUT TO IRS

Calculational data sink blocks in IRS are organized as IRS evolves, and they connect the data banks with calculational results. When creating the banks of the recommended values preference is given to the experimental values of parameters, therefore, the tuple for the recommended PSL must contain inside the tag determining the way to obtain the parameter. Now four parameters are used which can be experimentally measured: the line center, intensity, halfwidth, and shift. Indexing of the parameters is carried out in accordance with Table II.

A general value of the parameter is obtained as a result of summing of tags presented in Table II. For example, 11 means that the center, intensity, and shift were obtained as a result of experiment processing, whereas the halfwidth was obtained by calculation. Therefore, analyzing the tuple which has such a tag we can replace the halfwidth by new calculated value and remain other parameters unchanged.

TABLE II.

Parameter of line	Origin	
	Experiment	Calculation
Center	1	0
Intensity	2	0
Halfwidth	4	0
Shift	8	0

Similar tags can be included to the tuples of molecular and spectroscopic constants. Rules of tags forming can be the other in dependence on a number of used components in the constant.

CONCLUSION

Described principles of organizing of user interface and architecture of IRS can be realized on any algorithmic language. Most fast this process can be carried out on the basis of DBMS PARADOX and CLARION oriented to databases creation. The other database programming system known us are unpromissing due to slower record processing in the archives and inconvenient programmer interface.

The authors of the present paper elaborated the IRS on the high resolution spectroscopy for the IBM PC AT 286/287 on the base of DBMS CLARION. The architecture and interface of IRS oriented to fundamental and applied tasks of molecular spectroscopy and gaseous media optics.

REFERENCES

1. D. Tsichitzis and F. Lochovsky, *Data Models* (New Jersey, 1982), 381 pp.
2. O.K. Voitsekhovskaya, A.B. Rosina, and N.N. Trifonova, *Information System on the High Resolution Spectroscopy* (Nauka, Novosibirsk, 1988), 150 pp.