Geophysical Research Abstracts, Vol. 9, 01906, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-01906

Distributed information system on atmospheric spectroscopy

A.D. Bykov (1), **A.Z. Fazliev** (1), N.N. Filippov (2), A.V. Kozodoev (1), A.I. Privezentsev (1), L.N. Sinitsa (1), M.V. Tonkov (2) and M.Yu.Tretyakov (3)

(1) Institute of Atmospheric Optics SB RAS, Russia, (2) Saint-Petersburg University, Petersburg, Russia, (3) Institute of Applied Physics RAS, N.Novgorod, Russia

ABSTRACT

The urgency of creating the information-computational systems (ICS) on molecular spectroscopy follows from the circumstance that for some molecules the number of calculated energy levels counts hundreds of thousands, and the number of spectral lines sometimes reaches hundreds of millions. Publication of such data volumes in regular journals is inappropriate. Comparison of different calculated spectral characteristics or their comparison with experimental data beyond computer processing is hopeless. We consider information systems to be an adequate form for holding such data volumes and a toolkit for handling them. Correct digital data processing requires appropriate sets of metadata arranged in the form of ontology of molecular spectroscopy. First step on the way is task ontology of the molecular spectroscopy.

Our information system provides the data on spectral line parameters, water molecule energy levels, and absorption coefficients. Within this distributed IS one can solve two types of problems: manipulation with data and calculation of spectral functions. Among the latest experimental data in the IS there are data obtained at the Institute of Applied Physics RAS [1,2]. To calculate the absorption coefficients for the molecules of carbonic acid gas, we take into consideration spectral line interference[3].

1 INTRODUCTION

In Russia, the information resources on molecular spectroscopy have been progressively developed since the eighties at the Institute of Atmospheric Optics[4]. A progress in the development of information systems (IS) occurred with the advent of personal computers in the early nineties, after construction of the AIRSENTRY system[5].

The Internet-technologies promoted development and implementation of a shared information-computational system (ICS) on molecular spectroscopy. An Internet-accessible information resource[6] was created on basis of the known spectroscopic databanks HITRAN and GEISA and the original data[7]. The data of these databanks determined the list of domain-oriented applications available for users, the main ones among which were meant for the calculation of spectral functions. Further expansion of spectroscopy data structures was made by S. Mikhailenko *et al.*[8] in the information system S&MO devoted to ozone molecule. This IS was new to involve the data on the fundamental characteristics of an isolated ozone molecule, namely, energy levels, potential surface and wave functions, etc. With these systems, we managed to proceed from the concept of the molecular spectroscopy databank to that of the information system. The applications relevant to the problems of molecular spectroscopy, previously existing separately from the data, were integrated to an Internet-accessible system.

From the viewpoint of e-Science[9], a model information system has three layers: data and applications layer, information layer and knowledge layer. In our previous version of the ICS[6], we have realized only the first layer.

To introduce the information level to the IS, we took the concept of resources instead of data[10]. In the ICS "Atmospheric Spectroscopy"[11] the information layer was constructed based on the abstracts of the information resources downloaded from the outside and the resources having originated as task solutions[12]. Also, implementation of the information layer required the development of the ontology on molecular spectroscopy and the task ontology. The first version of the ontology on molecular spectroscopy, much simplified, was published in 2003[13].

The stage of developing of the knowledge layer started with building of a distributed information system on molecular spectroscopy, within which we supposed to realize a machine-aided exchange of annotations and thus to form a knowledge base.

This paper considers the data structure used in the distributed ICS, examples of downloading information resources, the procedure of calculating the absorption coefficient, and comparisons between experimental and calculated results.

2 HIERARCHY OF TASKS

The general approach to classifying allows us to divide all the tasks of molecular spectroscopy in two groups: direct and inverse problems.

The classes of direct fundamental tasks are:

1. **Determination of the physical characteristics of an isolated molecule (T1).** The result of solution is the calculated energy levels of a molecule, the wave functions corresponding to the stationary states of the molecule, and the motion integrals, whose quantum numbers identify the energy levels.

2. **Determination of the spectral line parameters of an isolated molecule (T2).** The result of solution is frequencies of transitions and the Einstein coefficients. The input data for the task is energy levels, wave functions, and motion integrals.

3. **Determination of the parameters of spectral line contour (T3).** The results of solution are half-widths, shifts, intensities, the parameters characterizing line interference, and statistical weights. The input data are transition frequencies, wave functions, Einstein coefficients, etc.

4. **Calculation of spectral functions (T4).** Absorption coefficients, transmission function, etc. are calculated for chosen thermodynamic and electromagnetic conditions. The input spectral data is the spectral line parameters of the interacting molecule.

5. **Measurements of spectral functions (E1).** The result is the metadata about the experimental environment and the values of measured spectral functions.

It is important that these tasks form a hierarchy. In the simplest case, for example, to solve task T3, you will need to have solved task T2. In other words, input data for task T3 must include the output data of task T2.

The classes of inverse fundamental problems are:

1. **Determination of the spectral line parameters of the molecule (ET1).** The input data is the measured spectral functions and measurement conditions. The output data is the spectral line parameters of interacting molecules.

(a) **Subtask of determining the spectral line centers (ET1.1)**. The result of its solution is transition frequencies (two types: vacuum transition frequencies and transition frequencies for a certain thermodynamic and electromagnetic environment).

(b) **Subtask of determining the spectral line intensities (ET1.2).** The result is the intensity of each transition frequency for a given thermodynamic and electromagnetic environment.

(c) **Subtask of determining the half-widths, shifts, and the temperature dependences of half-widths and shifts (ET1.3).**

(d) **Subtask of determining the line mixing and interference parameters (ET1.4).**

(e) **Subtask of determining the Einstein coefficients (ET1.5).** The result is the Einstein coefficients for the transition frequencies.

2. **Spectral lines assignment (T5).** The result is the identified relation between the transition frequencies and quantum numbers.

3. **Determination of the energy levels of an isolated molecule (T6).** The result is the list of energy levels with attributed quantum numbers, uncertainties of the energy levels and number transitions defining level.

We suppose, the tasks listed are the main ones to be introduced to the information system as applications (executed program code). The measurement results and these tasks solutions are had to be placed to the data warehouse. Concerning description of the resources related to the domain, such hierarchy of tasks allows us to build the ontology of the problems of molecular spectroscopy, which is used in the ICS to form a knowledge base on this domain.

The second group of tasks involves the rules connected with preparation of resources for task solution. To such classes belong the following tasks (their full list can hardly be provided, for in every case it will be determined by the information system's functionality):

1. **Formation of complex data sources.** The input data are the resources of the data warehouse, which contains original resources. The output data is the complex data sources of spectral line parameters, spectral functions, and energy levels.

2. **Visualization of metadata for the sake of comparison of different user data and between user and warehouse data.**

3. **Data visualization** (spectral line parameters, energy levels, spectral functions).

4. **Relations between data and metadata in molecular spectroscopy.**

3 DATA STRUCTURES

In molecular spectroscopy, the subject of study is the spectra of molecules. The objects of study are molecules and radiation. The properties of these objects determine data intensions in the information model of the domain. The values of the properties are data extensions. In our work on implementation of the ICS on molecular spectroscopy we use two data models: structured data model and formal data model[14].

The manipulations with data typical of any information system are its upload, storage, and representation for users. The problems of molecular spectroscopy determine the structures of resources for any manipulation. References 11 and 12 give descriptions of the structures of spectroscopy resources for some problems. The data and metadata schemes used for uploading can be found on the Internet at

- http://saga.atmos.iao.ru/xml/xsd/2004-absorption-use-ver2.xsd,
- http://saga.atmos.iao.ru/rdfs/absorption-coefficient.rdfs.

Molecular spectroscopy data are arranged in three groups in the DICS: fundamental molecular characteristics, spectral line parameters, and values of spectral functions.

Fundamental molecular characteristics determine energy of a molecule. Depending on the method of description, these can be either the parameters of the molecular Hamiltonian (potential energy, dipole moment) or the effective Hamiltonian (rotational, centrifugal distortion and resonance constants, effective dipole moment parameters). Here should also be added quadrupole and octupole moments and other parameters characterizing intermolecular interactions in gases.

Spectral line parameters are grouped as follows: parameters of an isolated spectral line (intensity, line centre, lower energy level, statistical weight of the upper and lower states, transition moment, etc.); identification parameters (vibrational and vibration-rotational identifications); collision parameters (half-width, pressure-induced shift, temperature dependence of half-width, etc.).

Spectral functions involve absorption coefficient, transmission function, absorption cross-section, emission spectra, etc.

The hierarchy of these data structures is of a nesting type. Theoretical spectral studies are based on the knowledge of fundamental molecular characteristics. The latter allow calculation of spectral line parameters, which, in their turn, help to determine spectral functions, for example, the absorption coefficient.

Spectral line parameters have a more complex structure and the largest data volume among all the three listed domains.

In molecular spectroscopy, we have defined three mechanisms of automatic annotating the ICS resource content area for structured data. The first one is user input of subject oriented data and data source creation. When the user inputs data, he enters the part of metadata which cannot be mechanically inserted by the ICS tools. Other statements are created dynamically (number of records in the data source, their volume, etc.). The second mechanism is conditioned by the data manipulation processes, when the user creates new data sources on basis of those already existing in the ICS.

Note that manipulations with the data accessible for all the ICS users are only those that have been annotated. The third annotating mechanism is connected with the tasks which the user solves within the ICS. The output document that the user gets upon task solution, contains the annotation that involves the RDF-description of the original task parameters, solution methods, results, etc. The necessary components in the annotations are task ontology and domain ontology. The variety of annotations accumulated through data entries and task solutions makes up the knowledge base of molecular spectroscopy.

4 CURRENT STATUS OF DIS

Currently, the distributed information-computational system (DICS) on molecular spectroscopy has three servers located in Tomsk (http://saga.atmos.iao.ru), St. Petersburg (http://saga.atmos.molsp.phys.spbu.ru), and Nizhny Novgorod (http://saga.atmos.appl.sci-nnov.ru). The whole system is built on basis of the same typical software which we used for implementation of the ATMOS portal[15]. Each organization contributing to DICS development can develop their information and computational resources on a self-governing basis. The resources to be shared by all the organizations are the knowledge bases that hold annotations to the downloaded data and to the task solutions obtained within this system. To write annotations, we use the RDF and OWL.

The news about fresh resources and applications appearing in the DICS are automatically delivered to each local IS. To do this, we use the RSS format.

5 CONCLUSION

This paper describes the main potentials of each typical server of the distributed information-computational system on molecular spectroscopy that we are now creating in Russia. We have explicitly formulated our preferred approach to structuring the molecular spectroscopy data taking into account its functional use. Here we have given examples of data input, calculation of the absorption coefficient, and comparison between calculations and experiments.

By now, we have structured the data and prepared the subject domain metadata for tasks T2, T3, T4, and E1. For the remaining tasks, data structuring and metadata description is in progress.

6 ACKNOWLEDGEMENT

The authors acknowledge the Russian Foundation for Basic Research for financial support (grant No. 05-07-90196, 06-07-89201).

REFERENCES

1. G. Yu. Golubiatnikov, A. V. Lapinov, A. Guarnieri and R. Knöchel, J. Mol. Spectrosc., **v. 234** (2005) 136-140.

2. M. Yu. Tretyakov, M. A. Koshelev, and V. V. Dorovskikh, *et al.*, J. Mol. Spectr, **v.231** (2005) 1-14.

3. N. N. Filippov, V. P. Ogibalov, and M. V. Tonkov. JQSRT, 2002, **v.72**, No 4, 315-325.

4. O. K. Voitsekhovskaya, A. V. Rosina, and N. N. Trifonova, Information System on High Resolution Spectroscopy, Novosibirsk, Nauka, 1988, 150p. (in Russian).

5. V. F. Golovko, A. V. Nikitin, A. A. Chursin, and Vl. G. Tyuterev, Information system AIRSENTRY for modeling atmospheric IR-spectra and radiation transmission in the atmosphere, Proceedings of the 2^{nd} Int. Workshop ADBIS'95, **v.2**, Moscow, p.12-14, 1995.

6. Yu. L. Babikov, A. Barbe, V. F. Golovko, S. M. Mikhailenko, and Vl. G. Tyuterev, Internet Collection for Molecular Spectroscopy, Proceedings of 3^{rd}

National Russian Research Conference “Digital Libraries: Advanced Methods and Technologies, Digital Collections”, Petrozavodsk, p.183-187, 2001 (in Russian). (Spectroscopy of Atmospheric Gases http://spectra.iao.ru)

7. Data banks of CO_2, ftp://ftp.iao.ru/pub/CDSD-296, ftp://ftp.iao.ru/pub/CDSD-1000

8. S. Mikhailenko, Yu. Babikov, Vl. G. Tyuterev, and A. Barbe, The DataBank of Ozone Spectroscopy on WEB (S&MPO), Computational Technologies, 2002, **v.7**, p.64-70

9. D. De Roure, N. Jennings, and N. Shadbolt, A Future e-Science Infrastructure, Report commissioned for EPSRC/DTI Core e-Science Programme, 2001, 78p.

10. M. R. Kogalovskii, Scientific Collections of Information Resources in the Digital Libraries, Proceedings of 1^{st} National Russian Research Conference “Digital Libraries: Advanced Methods and Technologies, Digital Collections”, S.-Petersburg, 1999, p. 16-31.

11. A. D. Bykov , B. A. Voronin , A. V. Kozodoev , N. A. Lavrentiev , O. B. Rodimova , and A. Z. Fazliev, Information system for molecular spectroscopy. 1. Structure of information resources, Atmosph. and Oceanic Optics, **v.17**, 2004, No.11, p.816-820. ICS “Atmospheric spectroscopy”, http://saga.atmos.iao.ru

12. A. Z. Fazliev, Description of information resource in molecular spectroscopy using XML platform facilities, Computational Technologies, Special Issue, **v.10**, part 1, p. 39-46, 2005 (in Russian).

13. O. B. Rodimova, S. D. Tvorogov, and A. Z. Fazliev, Ontology for molecular spectroscopy of atmospheric gases, Proceedings of 5^{th} National Russian Research Conference “Digital Libraries: Advanced Methods and Technologies, Digital Collections”, S.-Petersburg., 29-31 October 2003, p. 211-215 (in Russian).

14. M. R. Kogalovskii, Perspective Technologies of Information Systems, Moscow, DMK Press, IT Company, 2003, 288p. (in Russian).

15. E. P. Gordov, V. N. Lykosov, and A. Z. Fazliev, Web portal on environmental sciences “ATMOS”, Adv. Geosci., v.8, p.33–38, 2006.