Distributed web system oriented to atmospheric aerosol data

N.V. Golikov, I.A. Sudarikova, and Yu.I. Molorodov

Institute of Computational Technologies,
Siberian Branch of the Russian Academy of Sciences, Novosibirsk

Received November 27, 2005

Brief description of the information and computation Internet-system "Atlas of atmospheric aerosols of Siberia" is presented. The technology and methods for developing the distributed web-system are considered. The system provides systematization and structuring of initial data on atmospheric aerosols, allows maximal unification of access to heterogeneous data resources (contents of which can change in time).

Introduction

Atmospheric aerosol (AA) being an essential part of the atmosphere plays a crucial role in many atmospheric processes (cloud and precipitation formation, radiation heat transfer, and visibility). It significantly affects an environmental quality, climate, chemistry and physics of the atmosphere. AA is notable for variety of forms and physico-chemical properties, as well as for an extremely wide size range of its particles.

Strong spatiotemporal variability of aerosol parameters and their correlations with all atmospheric processes stimulate in-depth study the role of atmospheric particles in global climate changes and their effects on geospheric and biospheric processes. The ever-increasing anthropogenic impact on the Earth atmosphere is another important motivation for aerosol study, which allows a consistent quantitative assessment of probable ecological backlashes.

For the most part, grounds for environmental sciences are data bulks. Therefore, organizing the collaboration work with data repositories is a key problem in informatization of these sciences. Development of Internet–technologies allows a new organization of data storage and access.

The world community has accumulated a great bulk of data on spatiotemporal variability of atmospheric aerosols.[1] Storing data in different and often incompatible formats makes them difficult to process and compare. A lot of algorithms for data processing, representation, and visualization are working out. To unify receiving data, information and information-computation systems are developed. Authors of every such system solve a definite range of problems, often reproducing the same algorithms and functions.

Within the integration project "Atmospheric aerosols of Siberia" the public information system "Atlas of atmospheric aerosols of Siberia" has been designed with the following features:

1) its main database includes aerosol monitoring results from different data sources all over Siberia;

2) it allows processing of stored data with some mathematical algorithms;

3) it is possible to represent data in schematic (plots and diagrams) and tabular forms;

4) the system is accessible via Internet (http://web.ict.nsc.ru/aerosol/);

5) discretion of user rights to data adding and modification is provided;

6) two languages (Russian and English) are supported both at the interface and data levels.

Among the disadvantages of the system are impossibility of data processing with foreign algorithms and the strictly defined structure of the database (strictly relational).

In view of the above-mentioned, we have developed a distributed system meeting the following requirements:

1. The data bank is to be extensible and accessible, constituted of a set of databases located on different servers and consisted of files of different formats.

2. A unified glossary of basic notions related to AA is to be created, which is virtually the ontology of the subject field "Study of atmospheric aerosol."

3. Data are to be processable with algorithms from different application servers.

4. For adequate data understanding by a person and intelligent computer processing, documentation or metadata are to be enclosed, which is necessary to facilitate data sharing.

The use of web-service technologies was proposed for solving the problem and the subject field ontology as a logical connecting link. This allows virtual integration of heterogeneous data, located on differently organized servers, into a common database on the base of open international standards. This also gives an opportunity to use computational algorithms in data processing.

1. Constructing technology

"Web service" is a software system for maintaining net communications between computers. The service interface is described with a computer language, e.g. WSDL.[2] Other systems connect to the web service in a way specified in its description by means of SOAP messages using HTTP and XML in combinations with other web standards. Physically "web service" is a software fragment called "agent." An agent is capable to message and message back; it realizes service functionality. One service can be supported by different agents.

WSDL describes functional capabilities of a web service and groups the interaction operations to the interfaces defining ways of executing operations as well as sets input and output parameters. The problem of interpreting the information described in WSDL terms to the ontology terminology thus arises. This allows the notion on the subject field to be extended. For instance, some ontology describes the web service calculating the coefficient of correlation between ion and elemental aerosol fractions. Mass concentrations of aerosol fractions are used as input parameters. If a user needs some information on input data (what it is, what values it takes, how it is calculated, and so on) he can get it via semantic links in the ontology.

Today the use of ontology essentially extends. The timeliness in evaluating its properties is important when developing and maintaining. However, a standard definition of ontology as well as agreement on uniqueness of its properties estimation are lacking in literature. Give several examples.

Ontology of a subject field is a set of agreements. It defines terms of the subject field and their interpretation, contains assertions restricting implications of the terms, and interprets the assertions.

Ontology is a logical theory bounding admissible models of a logical language[3]; that is, it is to provide for axioms limiting values of nonlogical symbols (predicates and functions) used as "primitives" for purposes of representation.

The ontology of a subject field is the part of knowledge on the subject, which is supposed to be changeless. Another part of the knowledge is supposed to be changeable but should remain compatible with the ontology.

One subject field can have several ontologies, which have no fundamental differences. Ontologies of different subject fields can be similar. Some ontologies can be simplifications of others. Finally, ontologies of some subject fields can be constructed of ontologies of other subject fields. For instance, the ontology of science includes the ontology of aerosol chemistry (Fig. 1).

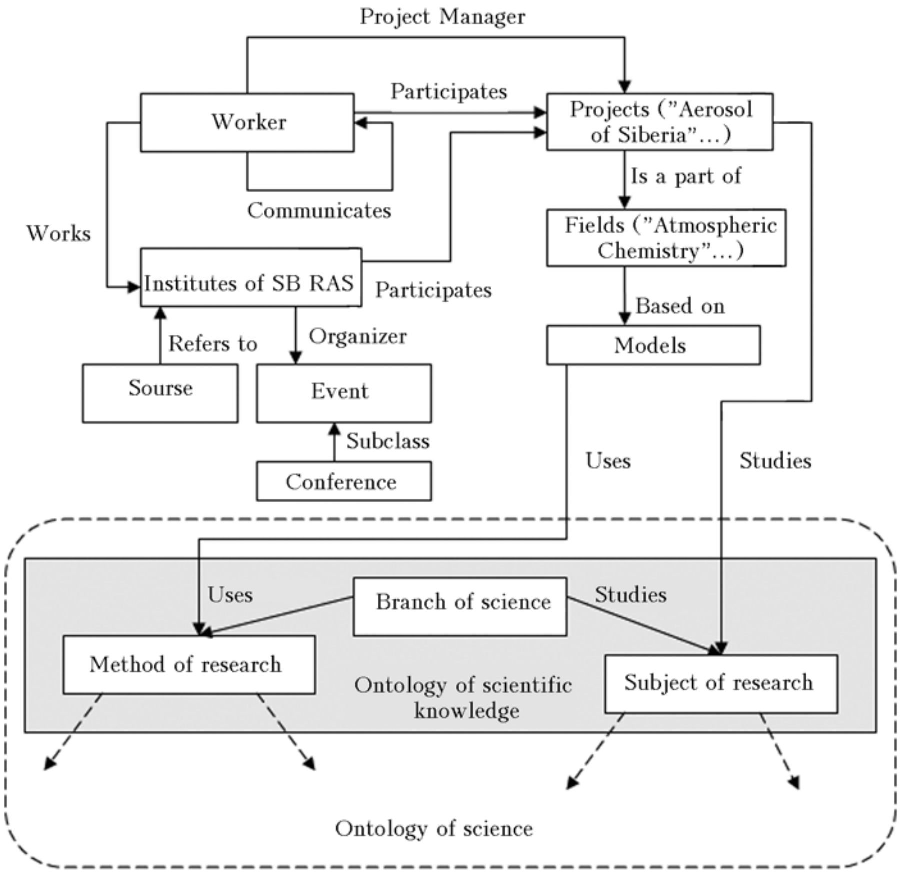

Fig. 1. Example of ontology relations.

Ontology can be presented as a graph of its basic structural components. A graph ontology model is a directed multigraph (a graph with loops and repetitive arcs) where each vertex corresponds to an entity or an entity class while directed arcs connecting the vertexes reflect different relations between the entities. The part of the AA ontology is shown in Fig. 2 in the graph model form.

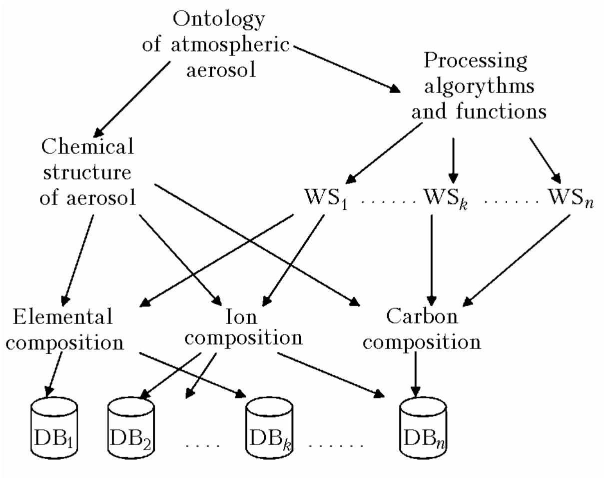

Fig. 2. Graph model of the atmospheric aerosol ontology.

Here WS_1, ..., WS_n is the set of web services which are processing or calculating programs. Each web service is associated with a group of AA chemical structures for operation. Database (DB_1, ..., DB_n), in its turn, assign the structures where data on respective aerosol composition are stored.

At present, a user points a database with which a web service is to operate but in future we suppose to change such the ontology solution. Following the web service description, it is necessary to decide which database is to work with. At that, a selection mechanism of the most complete and relevant data is required. While source item processing, the degree of data relevance to notions is to be monitored and the list of sources most appropriate for the notion is to be compiled. Thus the "decision problem" arises.

Ontology should include descriptions of both a subject field (SF) and related sources. The SF description includes a set of terms and relationships semantically significant for the SF, as well as rules for building assertions about SF elements. The source description contains a list of web links and description of their communications. To fix significant relationships between terms, principal relationships are pointed out, which can be represented in the graph form using binary relation diagrams. Such relationships can further serve a basis for integration of different ontologies. In our case, the ontology is to include complete data on all "classes" (laboratories, expeditions, scientific teams, chemical elements, etc.) which are possible subjects of the portal operation, references to similar notions in other databases, properties and restrictions related to these classes, and complete descriptions of web services able to work with the classes.

Thus, facility of any web service will be described in the portal ontology as well as notions, with which it operates, restrictions to input and output parameters (type of input and calculated data), description of executing procedures, its location, and so on. A "flexible" user interface can be built on such a basis. For instance, selecting a notion of interest in the menu (generated by the ontology) a user gets the information on web services able to work with the data.

2. System implementation

Two different variants can be distinguished in the distributed system operation. In the first case, the web service operates with a data bulk and there is no need to load all the data necessary for calculations to the main server hosting the user interface therefore web service is installed on the same server where the database is. In such a way the web service for albedo calculations operates in the "Atlas" (albedo is the reflectivity of the Earth surface). The user interface–web service communication is shown schematically in Fig. 3.

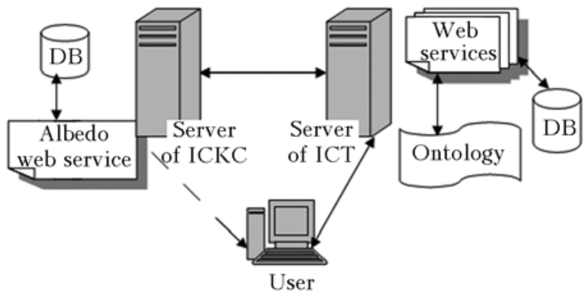

Fig. 3. The distribution of SB RAS servers' data and web services.

A user inputs geographic coordinates and time points using the system interface; according to the input parameters, the web service calculates albedo at the preset site and time via processing space images stored in the file directory on the server of the Institute of Chemical Kinetics and Combustion SB RAS. Each file from the directory contains an array of albedos values calculated at a certain time for Earth surface points and is of 300 MB in size after dearchieving.

In the second case, the web service operates with small data arrays, and there is no sense to distribute databases and web services over individual servers. The structure of such distributed system is given in Fig. 4.

In this case, the main system server hosts basic applications and the subject field ontology which contents are involved in the system interface (the ontology structure is shown in Fig. 2). On user demand for processing algorithms, being web services, a list of corresponding chemical structures and databases containing the information necessary for calculations is given according to the ontology. As was mentioned above, at present the user makes choice, but in future we plan to solve the "decision problem."

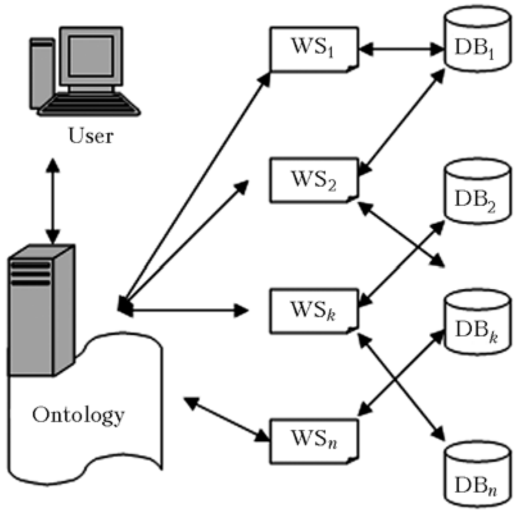

Fig. 4. The web services–distributed databases chart.

Conclusion

In the last years, environmental problems are in the highlight. For them, information on aerosol parameters is extremely important. Measurements, monitoring, and forecast of different aerosol parameters are essential for evaluation of the environmental quality and for working out stringent ecologically-sound requirements for industrial gas cleaning, clearness of industrial premises, control for sanitary and residential zones of industrial centers, as well as background observations aimed at detecting and studying negative anthropogenic impacts on atmosphere and environment.

A growing amount of investigations made actual the problem of accumulating and publishing information on AA properties in Internet. To do this, some difficulties are to be overcome. Data stored in parts in different systems, on the one hand, inevitably duplicate each other, but on the other hand are insufficient. Approaches to their structuring can differ in different systems. Standardization of sources, data systematization and structuring are important steps toward the common integration. For this purposes, the maximum possible unification of the access to heterogeneous data sources is to be realized.

As a rule, information systems (IS) essentially differ in using technologies, performances, access methods, and so on. This requires an individual approach to each source. However, there are standards and technologies allowing unification of heterogeneous IS communications; use of these standards makes the architecture of the common integration transparent. As a whole, the distributed system is to allow estimating the influence of anthropogenic and natural sources on variation of AA parameters and thus solving many environmental problems.

References

1. K.P. Kutsenogii, P.K. Kutsenogii, Yu.I. Molorodov, and A.M. Fedotov, in: *Proc. of Intern. Conf. on Mathematical Methods in Geophysics* (Novosibirsk, 2003), Part II, pp. 438–444.
2. Web Services Description Language: http://www.w3.org/2002/ws/desc.
3. M.R. Genesereth and N.J. Nilsson, *Logical Foundation of Artificial Intelligence* (Morgan Kaufmann, Los Altos, California, 1987).