Geophysical Research Abstracts,
Vol. 11, EGU2009-10783, 2009
EGU General Assembly 2009

{Semantic metadata application for information resources systematization in water spectroscopy}

A.Fazliev (1), A.Privezentsev (1), J.Tennyson (2)

(1) Institute of Atmospheric Optics SB RAS, Tomsk, Russia, (2) University College London, London, UK

(faz@iao.ru /Fax: +7 3822 492086 / Phone: 7 3822 492277)

A. Fazliev (1) and the A.Fazliev Team

(1) Institute of atmospheric optics SB RAS, Integrated information systems center, Tomsk, Russian Federation (faz@iao.ru), (2) University College London, London, UK

The information and knowledge layers of information-computational system for water spectroscopy are described. Semantic metadata for all the tasks of domain information model that are the basis of the layers have been studied. The principle of semantic metadata determination and mechanisms of the usage during information systematization in molecular spectroscopy has been revealed. The software developed for the work with semantic metadata is described as well.

Formation of domain model in the framework of Semantic Web is based on the use of explicit specification of its conceptualization or, in other words, its ontologies. Formation of conceptualization for molecular spectroscopy was described in Refs. 1, 2. In these works two chains of task are selected for zeroth approximation for knowledge domain description. These are direct tasks chain and inverse tasks chain. Solution schemes of these tasks defined approximation of data layer for knowledge domain conceptualization. Spectroscopy tasks solutions properties lead to a step-by-step extension of molecular spectroscopy conceptualization. Information layer of information system corresponds to this extension.

An advantage of molecular spectroscopy model designed in a form of tasks chain is actualized in the fact that one can explicitly define data and metadata at each step of solution of these molecular spectroscopy chain tasks. Metadata structure (tasks solutions properties) in knowledge domain also has form of a chain in which input data and metadata of the previous task become metadata of the following tasks. The term metadata is used in its narrow sense: metadata are the properties of spectroscopy tasks solutions. Semantic metadata represented with the help of OWL [3] are formed automatically and they are individuals of classes (A-box). Unification of T-box and A-box is an ontology that can be processed with the help of inference engine.

In this work we analyzed the formation of individuals of molecular spectroscopy applied ontologies as well as the software used for their creation by means of OWL DL language. The results of this work are presented in a form of an information layer and a knowledge layer in *W@DIS* information system [4].

1 FORMATION OF INDIVIDUALS OF WATER SPECTROSCOPY APPLIED ONTOLOGY

Applied tasks ontology contains explicit description of input an output data of physical tasks solved in two chains of molecular spectroscopy tasks. Besides physical concepts, related to spectroscopy tasks solutions, an information source, which is a key concept of knowledge domain information model, is also used. Each solution of knowledge domain task is linked to the information source which contains a reference on published task solution, molecule and task solution properties. Each information source allows us to identify a certain knowledge domain task solution contained in the information system. Water spectroscopy applied ontology classes are formed on the basis of molecular spectroscopy concepts taxonomy. They are defined by constrains on properties of the selected concep-

tualization.

Extension of applied ontology in *W@DIS* information system is actualized according to two scenarios. Individuals (ontology facts or axioms) formation is actualized during the task solution upload in the information system. Ontology user operation that implies molecular spectroscopy taxonomy and individuals is performed solely by the user. For this purpose Protege ontology editor was used.

For the formation, processing and visualization of knowledge domain tasks individuals a software was designed and implemented. Method of individual formation determines the sequence of steps of created ontology individuals' generation. Tasks solutions properties (metadata) have qualitative and quantitative values. Qualitative metadata are regarded as metadata describing qualitative side of a task such as solution method or other information that can be explicitly specified by object properties of OWL DL language. Quantitative metadata are metadata that describe quantitative properties of task solution such as minimal and maximal data value or other information that can be explicitly obtained by programmed algorithmic operations. These metadata are related to DatatypeProperty properties of OWL specification language

Quantitative metadata can be obtained automatically during data upload into information system. Since Object-Property values are objects, processing of qualitative metadata requires logical constraints. In case of the task solved in *W@DIS* ICS qualitative metadata can be formed automatically (for example in spectral functions calculation task). The used methods of translation of qualitative metadata into quantitative is characterized as roughened representation of knowledge in knowledge domain.

The existence of two ways of data obtainment is a key moment in the formation of applied ontology of molecular spectroscopy task.

- experimental method (metadata for experimental data contain description of equipment, experiment conditions and so on) on the initial stage and inverse task solution on the following stages;
- calculation method (metadata for calculation data are closely related to the metadata used for the description of physical and mathematical models of molecular spectroscopy)

2 SOFTWARE FOR ONTOLOGY OPERATION

Data collection in water spectroscopy information system is organized in a form of workflow that contains such operations as information source creation, entry of bibliographic data on publications, formation of uploaded data schema an so on. Metadata are generated in information source as well. Two methods are used for their formation: automatic metadata generation and manual metadata generation (performed by user). Software implementation of support of actions related to metadata formation is performed by META+ module.

Functions of META+ module can be divided into two groups. The first groups contains the functions necessary to software developer while the second one the functions necessary to a user of the information system.

META+ module functions necessary to the developer are:

1. creation of taxonomy (T-boxes) of applied ontology classes of knowledge domain tasks;

2. creation of instances of task classes;

3. creation of data schemes of tasks in a form of an XML-pattern and based on XML-syntax. XML-pattern is developed for instances generator and created according to certain rules imposed on software generator implementation.

4. implementation of metadata values calculation algorithms;

5. creation of a request interface and additional knowledge processing function for the solution of these task;

6. unification of the created functions and interfaces into one information system

The following sequence is universal for the generation of task classes' individuals that form chains.

Special interfaces for user operations management are designed for software developer in META+ module. There are means for qualitative metadata values updating during data reuploading to information source.

The list of functions necessary to end user contains:

- data sets visualization and editing, taking into account their metadata, e.g.: display of unique number of bands in transitions for a certain data source;

- export of OWL/RDF models from information system to the environment in XML-syntax;

- visualization of instances of classes of applied ontology tasks on molecular spectroscopy;

- import of OWL/RDF models into the information system and their integration with domain vocabulary;

- formation of additional knowledge of knowledge domain for the construction of ontological instances of task classes using GTML-formats and their processing;

- formation of additional knowledge in knowledge domain for the construction of instances of task classes, using software algorithm for data sets processing;

- function of semantic search implementation using an interface that formulates questions in a form of related triplets in order for getting an adequate answer.

3 STRUCTURE OF META+ MODULE

META+ software module that provides the above functions contains the following components:

- a knowledge base that stores semantic metadata and taxonomies of information system;

- software libraries POWL and RAP [5] created by third-party developer and providing access to ontological storage;

- function classes and libraries that form the core of the module and perform the tasks of formation, storage and visualization of classes instances;

- configuration files and module patterns that allow one to adjust and organize operation of different functional blocks;

META+ module also contains scripts and patterns implemented according to the rules of *W@DIS* information system development environment.

- scripts for interaction with environment by means of the software core of information system. These scripts provide organizing web-oriented interactive communication;

- patterns for the formation of functionality visualization realized by the scripts

Software core of scientific information-computational system *W@DIS* is created with the help of MVC (Model – View – Controller) design pattern that allows us to separate logic of application from its representation. It realizes the interaction of three logical components, actualizing interactivity with the environment via Web and performing its preprocessing. Functions of «Controller» logical component are realized with the help of scripts designed according to the rules imposed by software core of the information system. Each script represents a definite object-oriented class with obligatory class method of script initiation called “start”. Functions of actualization of domain application operation results representation (i.e. “View” component) are sets of HTML-patterns that allow one to visualize the results of domain applications operation with the help of additional constructions processed by software core of the system.

Besides the interaction with the software core of the scientific information system this module also deals with configuration files of software core and its database. Such organization of work provides closer integration with software core and deeper and more adequate connection in operating system support.

4 CONCLUSION

In this work the problems of semantic metadata creation in information system oriented on information representation in the area of molecular spectroscopy have been discussed. The described method of semantic metadata and functions formation as well as realization and structure of META+ module have been described.

Architecture of META+ module is closely related to the existing software of “Molecular spectroscopy” scientific information system. Realization of the module is performed with the use of modern approaches to Web-oriented applications development. It uses the existing applied interfaces. The developed software allows us to:

- perform automatic metadata annotation of calculated tasks solutions directly in the information system;

- perform automatic annotation of metadata on the solution of tasks on task solution results uploading outside the information system forming an instance of the solved task on the basis of entry data;

- use ontological instances of task solution for identification of data in information tasks of viewing, comparison and search solved by information system;

- export applied tasks ontologies for the operation with them by external means;

- solve the task of semantic search according to the pattern and using question-answer type interface.

5 ACKNOWLEDGEMENT

The authors are grateful to RFBR for the financial support of development of distributed information system for molecular spectroscopy.

REFERENCES

1. A.D.Bykov, A.Z. Fazliev, N.N.Filippov, A.V. Kozodoev, A.I.Privezentsev, L.N.Sinitsa, M.V.Tonkov and M.Yu.Tretyakov, Distributed information system on atmospheric spectroscopy // Geophysical Research Abstracts, SRef-ID: 1607-7962/gra/EGU2007-A-01906, 2007, v. 9, p. 01906.

2. A.I.Prevezentsev, A.Z. Fazliev Applied task ontology for molecular spectroscopy information resources systematization. The Proceedings of 9^{th} Russian scientific conference "Electronic libraries: advanced methods and technologies, electronic collections" - RCDL'2007, Pereslavl Zalesskii, 2007, part.1, 2007, P.201-210.

3. OWL Web Ontology Language Semantics and Abstract Syntax, W3C Recommendation 10 February 2004, http://www.w3.org/TR/2004/REC-owl-semantics-20040210/

4. W@DIS information system,
   http://wadis.saga.iao.ru

5. RAP library,
   http://www4.wiwiss.fu-berlin.de/bizer/rdfapi/.