Development of Information System Based on Ontological Design Patterns

Yurii I. Molorodov
Institute of Computational Technologies
Siberian Branch of the Russian Academy of Sciences
yumo@ict.sbras.ru

Abstract

In this paper, we present an intellectual scientific Internet resource (INIR) built on ontological patterns applied to scientific knowledge based on classes that set the structure for the description of concepts included in any scientific subject area (SSA), including the most interesting for us. Our subject area is focused on the field of atmospheric physics, which deals with the study of the chemical composition of atmospheric aerosols (AA) and their space-time variability. An important characteristic of AA is the dispersed composition, counting and mass concentration of chemical elements that are the basis of atmospheric aerosols. Ontology serves as the information basis of the resource, supports the integration of these resources into a single information space and provides access to them via the Internet. INIR provides meaningful access to this information, to the methods of its processing and methods of solving problems adopted in this field of knowledge and related Internet resources.

1 Introduction

The infrastructure of megacities is extremely saturated with the production of potentially dangerous objects, which leads to man-made and natural disasters, including the emission of aerosols. Aerosol pollution is the most dynamic and poses an immediate threat to health and the environment. It is these circumstances that have made it urgent to solve the problem of monitoring the atmosphere of large industrial centers, assessing the consequences and protecting against man-made disasters and accidents, fires and natural disasters. This range of problems is actively studied by teams of specialists working in various fields of environmental Sciences. These teams successfully work in the field of physics, chemistry, hydro-and gas dynamics, biology, ecology, etc. This implies the need for interdisciplinary approaches in environmental Sciences, and the inherent desire of any science to understand the phenomena and their prediction leads to the widespread use of mathematical modeling as a computational tool.

The accumulated large amounts of factual material have led to the problem of their systematization and providing them with effective meaningful access to a wide range of users. For this purpose, information systems (is) are developed, which are based on different approaches and principles that reflect the specifics of the areas of knowledge they represent and the tasks they solve [Zag16, Zag19]. Large demands are placed on the created IP. They should take into account the heterogeneity, distribution, and continuing growth of both experimental data and unstructured information in the form of publications, reports, images, and other types. The creation of such systems is an important, urgent and complex task. In their development, they have passed several stages, based on technology development tools and techniques for the development of knowledge portals (KP), intellectual scientific Internet resources (INIR), intellectual information and analytical Internet resources. This approach has been successfully used in such areas as active seismology [Zag18], thermophysical properties of substances [Zag19], decision support, research in energy. Initially focusing on weakly formalized KB, on the representation of semantic dependencies of concepts and systematization of heterogeneous information, it was supplemented by means of processing this information and solving problems typical for the considered KB.

To fully represent areas of knowledge that have large amounts of numerical experimental data stored in distributed sources, technology must provide a means of managing these data. The increased flow of information has led to the need to use new ways of its storage, presentation, formalization, systematization and automatic processing. There were ways to create knowledge bases (Knowledge Base - KB), allowing them to be used for various practical purposes [Zag14, Kov14]. Among them are Semantic Web (SW) technologies that contain hypertext pages with additional markup. It often contains information about the semantics of page elements. An important component of SW is the concept of ontology, which describes the meaning of semantic markup.

As a rule, ontology is understood as a system of concepts of some subject area. It is represented as a set of entities connected by various relationships. The creation of an intellectual scientific Internet resource or portal on certain topics, based on a developed ontological structure, will ensure effective access to information and its subsequent processing [Kov14]. Ontology, as the basis of the resource, allows, in addition to structuring data, to make initial verification of data, based on the rules specified in it [Hit09]. This helps to improve the quality of input information, and, in some cases, eliminate conflicting facts.

Ontologies are similar in many ways to thesauruses and taxonomies, but are actually broader in that they provide additional means to describe the structure of the data being described. At the center of most ontologies are classes, each of which can have subclasses that represent more precise concepts than the original class. All ontology classes are arranged in one or more hierarchies and describe the concepts of the domain. In this case, classes can contain attributes that describe the properties and internal structure of the concepts underlying the classes. All subclasses inherit the attributes of their parent classes. In addition to the name, each attribute of a class has a value type, allowed values, and a number of values. An attribute value type describes what types of values an attribute can contain, such as a string or an integer. There is also a limitation to the value of an attribute, which is that it can only accept certain classes or instances of certain classes.

Ontologies can be used wherever data processing that takes into account their semantics is required. Due to the initial orientation of the OWL language to machine processing, the correct application of ontologies can, on the one hand, significantly simplify and, on the other hand, open up new opportunities in the development of applications that solve the problems of automated processing and access to data.

2 Technology of development of intellectual scientific Internet resources

Intelligent scientific Internet resource (INIR) is a system with a Web-interface that contains systematized information related to the field of knowledge (KB), focused on atmospheric aerosols. It provides meaningful access to this information, to the methods of its processing and methods of solving problems adopted in this field of knowledge, as well as to related Internet resources. The main component of INIR is the ontology of KB. On its basis, the systematization of information about the lake and the functioning of INIR is carried out.

Development technology provides:

- the methodology used to build ontologies,
- set of basic ontologies,
- INIR sheath,
- means the specification of the user interface,
- data editor,
- a set of services that provide the functionality of the resource.

3 A set of ontological design patterns

To support this methodology, we used a set of OP patterns [Zag17] implemented in OWL. This set includes three types of patterns: structural logic patterns, representation patterns, and content patterns. The need to use structural logical patterns arose due to the lack of expressive means in the OWL language to represent complex entities and constructions that are relevant in the construction of ontologies, in particular, areas of valid values, multi-place and attributed relations (binary relations with attributes).

The valid value domain representation pattern is designed to specify constructs that are called domains in a relational data model and are characterized by a name and a set of elementary values. [Zag17] class properties when the entire set of such values is known in advance. In this pattern, a domain is defined by an enumerable class that inherits a specially introduced service class Domain and consists of a finite set of different individuals (objects) that define the possible values of some property.

As a base we use the pattern intended for the description of the applied problems solved within a scientific subject area (see Figure. 1). The following set of qualification questions is associated with this pattern:

1. What methods are used to solve the problem?
2. What data is used to solve the problem?
3. What is the result of solving the problem?
4. To what section of science does the problem belong?
5. Who formulates the task? etc.

To develop the ontology, CmapTools tools, the Protégé editor and the technique proposed in the ISI SB RAS inir development technology are used. The methodology is focused on the use of basic ontologies and ontological design patterns. When constructing the ontology, the concept structure based on the sequence ***Aerosol* = > *Measurements* => *Researcher* = > *Publications*** was used.

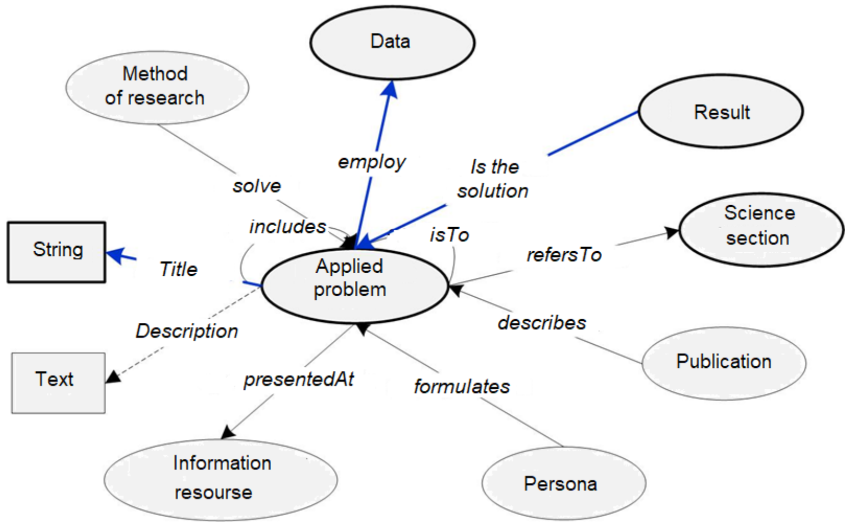

Figure 1: Pattern for describing applied tasks

4 The ontology of measurement systems

When measuring any physical quantity, it is important to know not only the measurement result itself, but also its spatial-temporal reference, the conditions in which the measurements were made, the characteristics of measuring instruments, etc.in one form or another, any system that makes measurements stores such metadata along with the measurement results.

The process of measuring physical quantities is well studied and regulated within the science of Metrology. Its main provisions are set out in the standards and regulations, so at the stage of modeling the subject area, we rely on them.Metrological support of measuring systems [Gos02] defines a measuring system as a set of components (measuring, connecting, computing) forming measuring channels. The measuring channel is a logical entity that combines the whole complex of measurements and transformations to obtain the measurement result of a given physical quantity.

The ontology of measuring systems was developed on the basis of this metrological standard. The General scheme of this ontology is presented in Figure. 2.

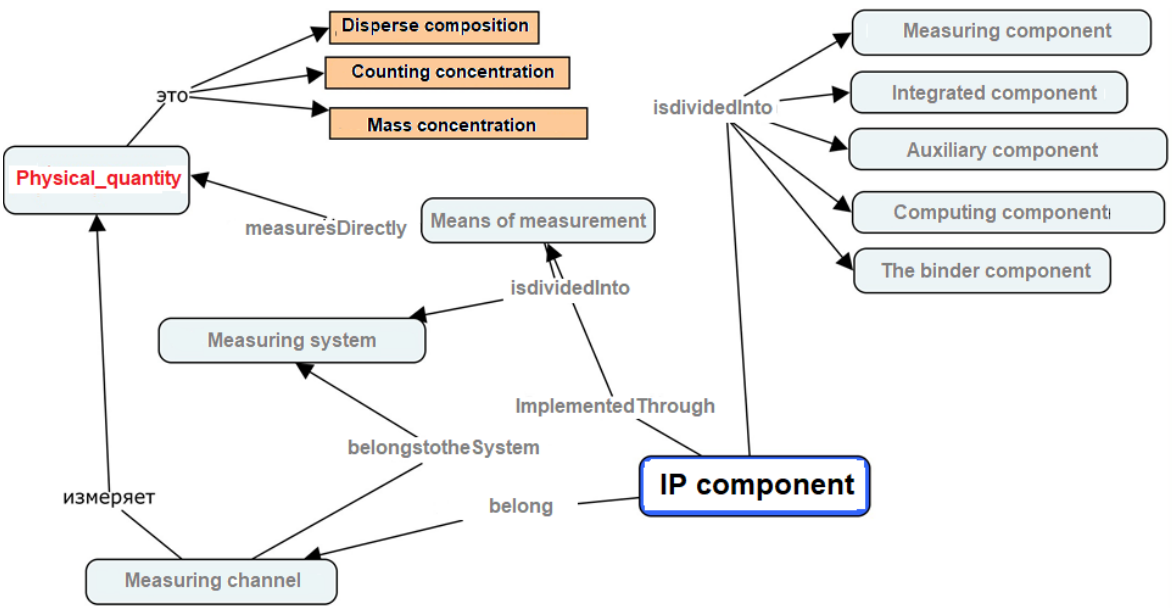

Figure 2: The measurement systems ontology

Important properties are the measured characteristics of the measurement object. They have at least two parameters: the name of the measured characteristic (concentration of any type of aerosol) and its dimension (for example, mg / m3). The values of the mass concentration of the submicron fraction of atmospheric aerosols in the form of experimental numerical data series are stored in an external data storage system. There can also be stored information, for example, on the spatial and temporal variability of concentrations of organic and inorganic carbon, nitrogen dioxide. Thus, users of INIR have a need not only to view external data directly in R & d, but also to use them to solve the problems of OZ resource. Therefore, special tools have been created to work with data from external sources.

5 External data access system architecture

To access external data, a system has been developed that provides INIR users with the following functionality:

1. Organization of interaction with external data sources. This can be third-party databases (DB), or databases created by developers of specific INIR.
2. Description of information objects with values from external databases.
3. Import property values of specified objects from external sources.
4. Visualization of object property values as tables or graphs.
5. Starting services analysis of the imported data.
6. Use imported data to solve problems.

The main component of this system is the data download service-a *Loader* that directly interacts with external storage. To connect to the database system, *the administrator Panel* is used, which has a web-based user interface that allows you to register new data sources and generate query templates to access them. The loader has its own database, which contains the addresses of the databases registered in it and the information required to build a request to a specific resource. Both SQL query templates for relational databases and other query formats (REST API, SOAP, SPARQL, etc.) for external resources can be used here. To build specific queries, the necessary parameters are passed to the Loader, which are extracted from the INIR ontology.

Services for working with data allow you to show them to the user, perform their analysis or use to solve the problems of resource management.

A special plug – in *Manager* was developed to organize the interaction of INIR with the Loader And services for working with data. This plugin is designed to extract from the inir ontology parameters required by the Loader to build a query to an external database. The Manager passes the parameters to the loader, receives the request ID from It, which then passes it to the required Service.

Let's consider the scheme of functioning of the system of access to external data. In order to be able to use external data in the INIR, it is necessary to register in the database Loader the template of the query that makes the selection of the necessary data through the administrator Panel. In this case, each three (template, connection string, database type) is assigned a unique identifier, which is reported to the knowledge engineer. The knowledge engineer must define a class of objects in the ontology of OZ, whose properties will take values from the external database, and associate the resulting identifier with this class. In addition, it should take care to associate the properties of such objects with the parameters of the query template to the external database (their names and order in the template).

When the properties of objects belonging to such classes are displayed in the INIR, the Value attribute appears with a hyperlink">>>". Clicking on this link initiates the work of the Manager, which is passed the name of the displayed object. By the name of the object, the Manager makes a request to the ontology and R & d, which will result in an identifier associated with the class of this object and a set of parameters of the query template with their names, values and order of occurrence in the query. The identifier and parameter set are passed by The Manager to the Loader. The loader writes this information to the temporary cache and also assigns a unique identifier to each such new request. Depending on at what point in the work of INIR and for what purposes the Manager was called, it transmits the request ID to a certain Service that can access the Loader, and presenting the request ID, receive and process the data received from the Loader.

6 Conclusion

As part of the technology of creating intelligent scientific Internet resources, a system of access to data from external sources was developed. The use of Semantic Web technology and tools made it possible to simplify the establishment of communication between INIR content objects and the values of their properties stored in external databases.

Service-oriented approach was used in the basis of shell INIR development. According to this approach, all functionality INIR is implemented with the help of services – local or distributed, loosely connected, replaceable components equipped with standardized interfaces for interaction on standardized protocols. This approach allows resource developers to create different services for processing information stored both in INIR content and in external storage, as well as to use third-party services.

Our decisions are based on the development of ontologies related to a particular subject area. Presented knowledge portal on atmospheric aerosols allows you to link the data obtained in the course of field and computational experiments with detailed information about the activities, persons relevant to the research, text documents that can be useful in the further use of data and interpretation of simulation results.

Ontology development is an iterative process. The compiled anthology will be supplemented with new concepts as the number of experts increases. In our opinion, the work performed is the next step towards solving the problems of integrating knowledge into specialized Internet resources focused on environmental protection, in General, and monitoring of atmospheric aerosols in particular.

Acknowledgements

The work was supported by a RFBR grant N 18-07-01457\19 and projects N 0316-2019-0002 and N 0316-2018-0009

References

[Zag16] Zagorulko Yu.A., Zagorulko G.B., Borovikova O.I. *The technology of creating thematic intellectual scientific Internet resources based on ontology* // Theoretical and Applied Scientific and Technical Journal "Software Engineering". 2016. V.7, No. 2. P. 51-60. (In Russian)

[Zag19] Zagorulko Yu. A., Borovikova O. I. *Problems of constructing ontologies of scientific subject areas on the basis of ontological design patterns* // Pr. ITiS-2019 Information technology and systems, March 12-16, 2019, P. 157-161, Khanty-Mansiysk, 2019. https://itis.uriit.ru/files/sbornik_full.pdf (In Russian)

[Zag18] Zagorulko, Y., Borovikova, O., Zagorulko, G., 2018. *Development of Ontologies of Scientific Subject Domains Using Ontology Design Patterns.* In Communications in Computer and Information Science, 822: 141–156.

[Zag17] Zagorulko Yu.A., Borovikova OI, Zagorulko G.B. *The use of ontological design patterns in the development of ontologies of scientific subject areas* // Proceedings of the XIX International Conference DAMDID / RCDL'2017. Moscow, Moscow State University, Russia. 2017. S. 332-340. (In Russian)

[Bra15] Braginskaya L.P., Grigoryuk A.P., Kovalevsky V.V. *Scientific information system "Active seismology" for complex geophysical research* // Bulletin of the KRA-UC, Earth Sciences, 2015. No. 1, Issue. No. 25, S. 94-98. (In Russian)

[Zag14] Zagorulko GB, Molorodov Yu.I., Fedotov A.M. *Systematization of knowledge on the thermophysical properties of substances* // Bulletin of Novosibirsk State University. Series: Information Technology. 2014. T. 12, No. 3. S. 48-56. (In Russian)

[Kov14] Kovalev S.P., Prokopov N.A. *Automation of measurement processes of physicochemical quantities based on ontology* // Computational technologies. 2007.V. 12, Special. release. S. 79-87. (In Russian)

[Who] WHO, Air quality and health, http://www.who.int/mediacentre/factsheets/fs313/en/index.html (In Russian)

[Tpr] https://tproger.ru/articles/mvc/

[Gos02] GOST R 8.596-2002. Metrological support of measuring systems. The main provisions. M .: Gosstandart of Russia, 2002 (In Russian)

[Hit09] Hitzler P., Krötzsch V., Rudolph S. *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 2009. 455 p.

[Pro] http://protege.stanford.edu/products.php#desktop-protege