

Developing Of High-Performance Technologies For Creating And Supporting Web Ontologies

Valentin Filatov

Professor of Artificial intelligence department, Kharkiv National University of Radioelectronics, prospect Lenina, 14, UKRAINE, E-mail: val_filatov@ukr.net

Sergey Shcherbak

Associate professor of Information control systems department, Kharkiv National University of Radioelectronics, prospect Lenina, 14, UKRAINE, E-mail: ontolog@gmail.com

Antonina Khairova

Undergraduate of Artificial intelligence department, Kharkiv National University of Radioelectronics, prospect Lenina, 14, UKRAINE, E-mail: antonina.khairova@gmail.com

Abstract – The subject of the article is examination of storing large and stable ontologies in the spatial database, and also development of information storage and retrieval system of construction and supporting ontologies.

The article describes the method of storage large and stable ontologies with using systems of management of spatial data.

Key words –Ontology, RDF(Recourse Description Framework), RDFS (RDF Schema), OWL (Ontology Web Language), Oracle Database 10g, Oracle Spatial

I. Introduction

Due to the widening of Internet the work of searching systems has become harder, to provide sufficient coverage they had to index and process bigger volume of information. But the main difficulty was not even the increasing number of indexed web-sites, but providing users' queries with relevant answers, i.e. giving them links to the resources they needed.

Present-day reality in individual areas of scientific exploration dictates the necessity of applying methods on knowledge engineering for solving wide class of practical goals. As an example there may be used the initiative of Semantic Web, the main aim of which is to give to large massive sod data, printed in the Internet, more sensible, to raise accommodation of working with the information.

One of the main achievements of Semantic Web project became the developing of ontology standard description language – OWL (Ontology Web Language). Thanks to this many knowledge engineers, programmers and experts got the possibility of using common rules of ontology conception, storage and processing.

There are different meanings of word ontology. In this article ontology is the specification of some data domain, it's conceptual description as formalize conception, which includes vocabulary of terms of data domain and logical expressions describing intercommunications of these concepts. Thus, ontology of some data domain represents concepts thesaurus of this domain, providing the possibility of interpretation data domain terms via interpretation paradigmatic relations like "the whole and the parts", "class - subclass" and some kinds of associative relations.

Together with the interests to the ontology the instrumentals for working with them were created, specially aligned for wide common use of ontology in the issues of intellectual searching, classification, and exposure of data non-coordination, modeling of behavior intellectual agents and processing of data. However, even presence of good instrumental environment doesn't decline problems

of ontology designing and constructing. And the problem of knowledge extraction automation just like the whole the problem of ontology extraction still doesn't have an effective solution nowadays. Already developed ontologies and the experience of using them for solving different problems are becoming become more valuable.

The process of creating modern intellectual information systems often requires integration of knowledge from different sources, and as consequence effective solving of problems of knowledge replication. The problem of automation of choosing process still doesn't have a satisfactory solution. That's why studies on the developing such an approach for providing and replication of knowledge, which on one hand could allow to consider the specificity of data domain the most sufficiently and on the other hand – to provide and use knowledge in some uniform kind – are very important to present day.

II. Urgency and goal of work

Ontology models for the time of researching in this area have undergone considerable development. In the present time there are a lot of instruments for creating and supporting ontologies, which besides common functions of editing and oversight perform support of documentation ontology, import, export of ontologies in different formats, support of graphic editing, management of ontology libraries etc.

These tools of ontology constructing have several drawbacks. Most of the tools store their ontologies in text files, which limits size of ontology, the have low productivity, have superfluity of functions, that makes user's work more difficult. The additional development of algorithms for comfort of working with stored metadata is needed.

Basing on the features of analysis of existing tools drawbacks, we can say that subject of the article is examination of storing large and stable ontologies in the spatial database, and also development of information storage and retrieval system of construction and supporting ontologies.

The article describes the method of storage large and stable ontologies with using systems of management of spatial data. During the realization f this module the following features of ontologies were taken into account: in ontologies knowledge is formalized as a description of data domain by means of class hierarchy; a separate set of features and objects is given for each class, features in

ontologies have definitional domain – class, for which this feature is set, and also value area.

Depending on the value area the features are divided into two types: T-features (values of this data type or set of fated values) and o-features (values of which are objects of fated class).

We can say, that the hierarchy structure on ontologies is projected on spatial structure of database. It is necessary to note that ontologies can be very big, in some cases besides complicate hierarchy with amount of classes and features there can be stored millions and millions of objects. In addition to this situation, when in one database great number of ontologies are stored, the speed of queries execution to database becomes lower. As the considered database is one of the parts of complicate project, in which access to database is performed very often, so that in general the speed of work execution depends on the speed of queries execution to database simpliciter.

Methods of application should allow putting information about data domain like ontologies in the spatial database, and also allow high speed of query execution. There should be methods for processing ontologies, classes, objects and features of objects, and also administration of database session in it.

As a result, it is necessary to draw attention, that the suggested methods of storage and support of ontologies in the spatial database and methods of their management are universal, which allows wide using of the results of this work in research and applied work.

III. Semantic data description

Ontologies have been developed and used for solving different problems, including combined usage by people and program agents, possibility of accumulation and second knowledge use in data domain, creating models and programs, operating with ontologies, and not neatly specified structures of data, analysis of knowledge in the data domain.

For more intellectual generalization of sections of the information it is necessary for web-portals to define the ontology, which should describe the terminology used in contents of a web-portal, and the axioms setting rules of use of these terms in a context of other terms. Set of ontology and axioms is the model of the data description.

The base building block of data model is the statement representing a three: a resource, the called property and its value. In terminology RDF (Resource Description Framework) these three parts of the statement accordingly refer to: the subject, a predicate and object [5].

Everything that is described by means RDF is called a resource. It can be ordinary Web-page or its any part, for example, separate element HTML or XML, being a part of the described document. Also a resource can be the whole collection of pages, for example, separately taken Web-site. And, at last, something, not being accessible through the Internet can be resource, for instance, any subject from the world of things. In a word, everything that can attribute some URI (the universal identifier) or URI with addition of an internal name of object (a name

of an anchor in HTML) can become a resource and be described by means of RDF.

It is necessary to understand a certain aspect, the characteristic, attribute or the attitude used for the description of a resource as property. Each property has the specific sense, admissible values, type of resources to which it can be applied, and also attitudes with other properties.

According to the specification, value of property can have one of two types. The first is the resource set by some URI. The second type – a literal – is some text value of the characteristic. However, the literal can express value of any primitive type of the data which are present in XML. It's test also can comprise a certain marking, for example, XML, but distinctive feature of such marking is that it is not processed by the RDF-processor and is perceived as a usual line [6].

Real value RDF cannot be estimated, while it is used for the internal purposes of a separately taken application. There will be the use of RDF introduction, when it becomes means of interprograms interaction, data exchange when machines receive ability to combine the information received from various sources, thereby, receive any new information. The more applications on the Internet can work with data, the higher their value will be.

The description of technologies for development and support of ontologies.

The status of recommendation W3C and the presence of ready interoperability program decisions make Semantic Web technologies more attractive, than other technological decisions of knowledge engineering [7]. The key moment of the given approach is that besides the resources in our database metadata for the description of objects of storehouse and management by them, metadata, described by means of RDFS (Resource Description Framework Schema) language (fig. 1) are stored.

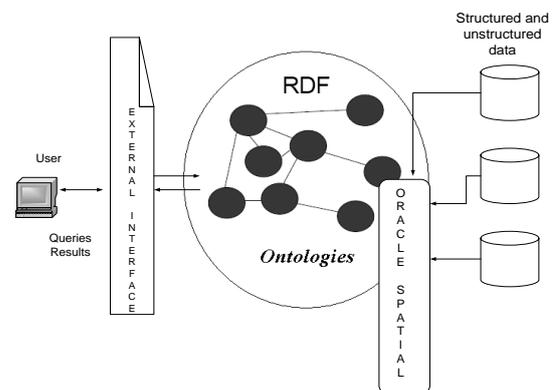


Fig.1 – Model of storage of metadata

The given approach is based on Oracle Spatial - data manager technologies Oracle Database 10g, including additional opportunities on processing spatial data for support of spatial services, different kinds of the GIS – applications intended for processing or granting of the information about the location of objects and other information systems.

Database Oracle 10g includes RDF/RDFS languages support, enabling developers of applications to use advan-

tages of a platform of a semantic data structure. Applied developers can supplement value to data and metadata, defining new sets of terms and attitudes between them. These sets of terms ("ontologies") are more adapted for realization of queries and the analysis based on the semantic approach, than usual data sets. The ontologic data sets often contain millions of data elements and attitudes between them which can be grouped in triplets, using new RDF model of data. Oracle supposes expansion to trillions of triplets for satisfaction of requirements of the majority of applications. What kind of storage principles of RDF has Oracle Spatial 10g?

- RDF data are stored as directed, logic graph;
- Subjects and objects are displayed as units, and predicates as ties at which the subject is the initial unit, and the object is final;
- Ties are a full RDF triplet;
- RDF the Data model supports three types of objects of a database:
- Model (RDF graph, that consists of a set of triplets);
- Base of rules (the set of rules);
- the Index of a rule (directed RDF graph).

Use of the proposed technology allows developers of a portal to create the uniform unified data presentation in all applications that will allow to find precisely the necessary information, simplify corporate data integration, reduce redundancy of data and provide unity of semantic values in all applications. All this, in turn, facilitates development, support and updating of applications within the limits of corporation.

The basic advantages of use Oracle Spatial 10g are:

- Integration of data from different sources without using of programming;
- Support of the decentralized management by data;
- Support of all RDF types of data;
- SQL search and restoration RDF of models;
- Realization of queries to RDF Models, with use of the scheme the graph;
- the Combination of queries RDF to others SQL operators;
- the Logic conclusion based on RDFS (RDF schemes) rules;
- The logic conclusion based on rules, defined by the user in the application.

IV. Development of elements of the software of system

In a basis of the proposed approach is active use of me-taknowledge not only for the description of syntax and semantics of language of knowledge representation, but also for ontologies constructing, describing the basic kinds between concepts of problem area, and also model of the user of the system. However, as it was already marked in the introduction, designing and development of ontologies, that is ontologic engineering, is not a trivial problem. It requires the developers to have professionally the knowledge engineering technologies — from methods of extraction of knowledge before their structurization and formalization [8].

Nowadays for the majority of ontologies construction tools the following principles are typical: first of all, though the main part of similar systems has a visual component, it is necessary to type some constructions manually, that raises a level of requirements to the ontologies developer – before starting the work directly, the expert is compelled to waste time for studying the language of representation of knowledge; secondly, a part of tools realizes the certain functionality for performance of ontologies queries, but, unfortunately, have no unified interface for formation and performance of queries from external applications; thirdly, practically there are no ontology editors freely distributed and focused on the final user, which, naturally, slows down the development of all direction of ontology engineering.

During the development of the tool environment for the ontology developer we have tried to level minuses of analogues and to adopt their advantages. Any object of ontology has graphic representation (not only classes and individuals, but also properties, ties, etc.). The system is the independent application which is capable to represent itself as an ontology server. Now the editor completely supports designs of language of the ontology description RDF, work on expansion of its opportunities to language OWL is conducted. On fig. 2 the architecture of the appendix is resulted.

In this article we have paid special attention to the development of the effective application of processing, creation and management of ontologies.

Let's note, that each copy in the world of ontologies is a member of class THING, so that each class defined by us automatically is subclass THING.

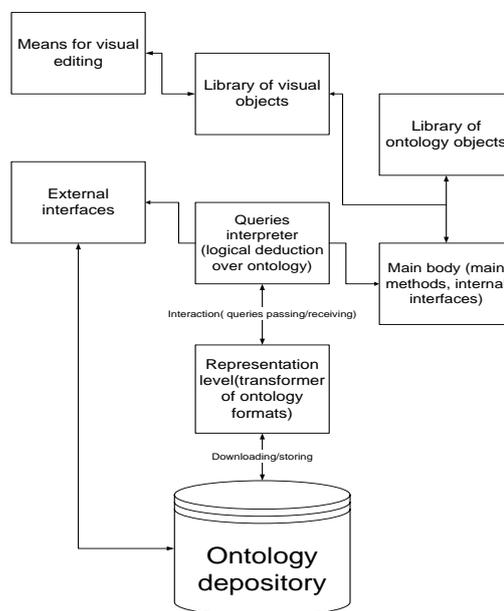


Fig.2 – Architecture of the developed system

Work of the user in system begins with creation of new data model. In the given model the user can create own ontology of a subject domain and in the further to edit it not mentioning other models. At the same time he has an opportunity at any moment to change current model of

data and to start supporting another ontology or in general to remove model from base (fig. 3).

For ontology creation after model constructing the user passes on next tab "Classes". In the left part of page the hierarchy of classes in the form of a tree realized by means of technology ajax is located.

When choosing any class in the right part of the window the properties belonging to the given class, and also specimens concerning to them (fig. 3) are displayed. Here there is an opportunity of addition and removal of a class. Also on this page the user can attach properties to the chosen class and to add or delete specimens of a class. All changes are brought in base as RDF – triplets.

On a tab "Property" the tree of properties is displayed, besides realization is based on technology ajax. In the given tree parental attitudes of properties to each other, i.e. hierarchy of properties are displayed. When choosing any property the tree of the instances showing their attitudes among themselves on given property.

On the given page there is an opportunity of addition of new property, removal of property and its editing. In property edition the user can change type of property (float, string, int, class), classes to which the given property can be attached, and also parents of property (fig. 4).

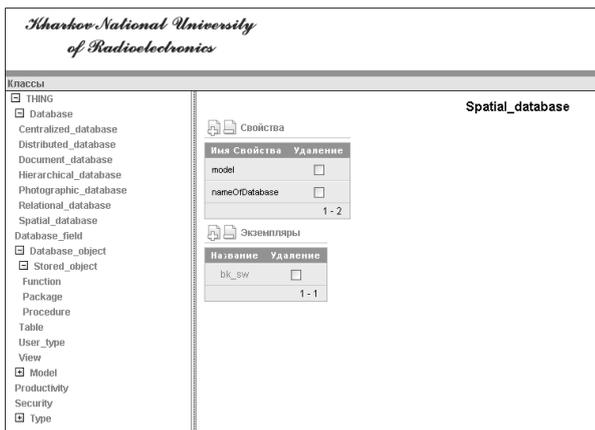


Fig.3 – Representation of classes, their copies and properties

The opportunity of the user independently to develop queries to a database is realized. Here we can combine usual sql queries with sparql queries. We shall consider a following example: to recognize all databases concerning to a class to spatial databases and to receive date of creation of base from other table. The given inquiry will look like that:

```
SELECT database_inf.createdate FROM database_inf,
TABLE(SDO_RDF_MATCH(
    '(?m rdf:type :Spatial_database)
    (?m :nameOfDatabase ?name)',
    SDO_RDF_Models("bk_sw_model"),
    null,
    SDO_RDF_Aliases(
        SDO_RDF_Alias(",sm_pkg.user_alias),
        null)
    )t WHERE
sm_pkg.user_alias||database_inf.name=t.NAME
```



Fig.4 – the Window of editing of property

Using the function SDO_RDF_MATCH we get access to our database RDF of triplets. The first parameter is the basis sparql query, in which we particularly specify what we wish to receive, the second parameter is a model to which we address. The third parameter is the base of rules using which we get a possibility to make a logic conclusion. Fourth parameter is a space of model's names. The fifth parameter is a filter, one of the parameters of sparql query. At first it is recommended to use parameter like "null" because for filter using it needs to be added into the rules base, which will be realized in the next version of the application.

The whole ontology of the defined data domain, as it was mentioned earlier, is stored in a database like RDF triplets. But triplets that are stored in any database do not bear any advantage, because the basis part of Semantic Web are search agents, on queries they receive some kind of the certain resource description in the form of the text file described on owl or as in our case on RDF language. In our system the module which gives out part of RDF document, describing required object, by query of the intellectual agent or other system (the query is made using the http protocol) is realized. The module also gives the list of all objects, the information about which is stored in system. Agent requests formed RDF document from database by the following address:

```
HTTP://<HOST_NAME>/<DATA_BASE_ACCESS_DE
SCRIPTION>/
<PROCEDURE_NAME>
<PARAMETERS_OF_QUERY>
```

The result of query of http://localhost/apex/swagent?p=bk_sw the following rdf-document comes back:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-
schema#
xmlns:model="http://swhost.kture/swstore/dbl#">
<rdf:Description
rdf:about="http://swhost.kture/swstore/dbl#bk_sw">
```

```

<rdf:type
rdf:resource="http://swhost.kture/swstore/dbl#Spatial_dat
abase"/>
<model:model
rdf:resource="http://swhost.kture/swstore/dbl#relational_
mode"/><model:nameOfDatabase>bk_sw
</model:nameOfDatabase>
</rdf:Description>
</rdf:RDF>

```

Conclusions

In this article the way of storage of large and stable ontologies, using technology Spatial Database Oracle 10g is considered. The use of Oracle Database 10g for data management which are marked by semantic marking language. It allows to allocate a set of advantages in comparison with management approaches based on files or on specialized databases. First of all there is low risk, high quality, productivity and safety in this approach.

As a result of the analysis of ontology storage and development existing systems of knowledge bases the following minuses [9] have been revealed:

- Data is stored in files;
- low productivity;
- development of additional algorithms for convenience of metadata storage;
- redundancy.

The information storage and retrieval system of ontology development and support has been developed for elimination listed mines above.

As a result of the full description of objects and their properties, the data domain is presented as the complex of hierarchical knowledge base. It is possible to carry out "intellectual" operations on it, such as semantic search and definition of integrity and reliability of data.

The basic advantages of the developed system are:

- convenient ontology storage in a spatial database;
- access to ontologies granted by a server via web-service, ontology preservation and extraction from depository;
- absence of converter from format RDF to the relational scheme and on the contrary realization necessity;
- use of objective ontology model, which represents concepts and attitudes from ontology in a user-friendly object-oriented interface;
- the convenient user interface;
- granting of the object description in a text kind on user query.

One of the minuses of system is the complexity of introduction. Format RDF has high complexity and it is not good for using by ordinary Internet users. Also this format does not allow to describe a data domain in corpore, therefore, support of the ontology description by OWL [5] will be stipulated in the future.

For many web-developers and programmers can be difficult to study RDF and OWL. Besides the main aim of the concept still is not known for many users. Work on

popularization Semantic Web still is not finished, there are no practical examples.

The use of the proposed technology will allow ontology developers to create the uniform unified data presentation in all applications. It will allow to find precisely the necessary information, will simplify corporate integration of data, will reduce redundancy of data and will provide unity of semantic values in all applications. All this, in turn, facilitates development, support and updating of applications.

References

- [1] Filatov V.A., Khairova A.A. Technology of the educational web-services organization on the base of XMLDB, HTMLDB, ORACLE SPATIAL // *International scientific and technical magazine « Informational technologies and computer engineering »* – Vinnitsa: VNTU. – 2007r. – Ed. 1 (18) – p. 240 – 247.
- [2] Collins H. Enterprise knowledge portals: next generation portal solutions for dynamic information access, better decision making and maximum results. – N.Y.: AMACOM, 2003. – p. 403.
- [3] Sherback S.S., Khairova A.A. Developing of education web-services as effective strategy of development network study// *Scientific – practice forum «Informatization of business by yang people: progressive technologies,science, business undertakings » (17th – 18th of may 2007).* – Kharkov: KNEU, 2007. – p. 73 – 74.
- [4] Filatov V.A., Khairova A.A. Research of methods and tools for development educational web – services. // *11-th International youth forum « Radio electronics and youth in XXI a century ».* – Kharkov: KNURE, 2007. – p. 387.
- [5] Berners-Lee, T. , Hendler, J., Lassilla O. The Semantic web - a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities // *Scientific American*, May 2001.
- [6] L.Stojanovic, J. Schneider, A. Maedche, S. Libischer, R. Studer, Th. Lumpp, A. Abecker, G. Breiter и J. Dinger. The role of ontologies in autonomous computing systems. - *IBM Research Journal*. 2004.
- [7] Xavier Lopez, Susie Stephens, Jeam Ihm, Jayant Sharma, Melliya Annamalai, Omar Olonso. Semantic Data integration for the Enterprise. March 2006.
- [8] Ternier, S., Duval, E., Vandepitte, P. LOMster: Peer-to-peer Learning Object Metadata. In: P.Barker and S. Reblsky (eds.) *Proceedings of ED-MEDIA'2002 - World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Denver, CO, June 24-29, 2002, AACE - pp.1942-1943.
- [9] Gavrilova T.A., Horoshevsky V.F. Knowledge base of intellectual systems: the Textbook for high schools. – SPb: "Peter", 2000.