Information model portal of scientific knowledge

Jamalova Jala

Abstract— **Information model portal of scientific knowledge which should provide substantial access to scientific information resources is proposed. Information model combines the models of domain and problem fields of the portal, as well as describes the types of the presented information. On the base of this model, the internal portal data storage is constructed, its information content, navigation and search are organized.**

***Index Terms*— Ontology, knowledge portal, information model, scientific information resources**

I. INTRODUCTION

Currently, the problem of the effective use of the vast amount knowledge and information resources accumulated in various areas of human activity is very acute. However, access to these knowledge and resources is significantly limited due to the fact that they are poorly structured, poorly systematized, and dispersed across various Internet sites, libraries and archives. To solve this problem, an approach to the construction of specialized Internet knowledge portals [1,2] ensuring the integration of accumulated knowledge and information resources in a certain field of scientific knowledge and substantial access to them is proposed.

An information model is proposed to provide a unified presentation of diverse knowledge of data, accounting for their connectivity, as well as supporting the functionality of such portals, the basis of which is ontology. The paper is devoted to the description of this model.

II. INFORMATION MODEL PORTAL OF THE KNOWLEDGE

Information model portal of the knowledge combines the models of domain and problem fields, as well as describes the types presented in the information portal. Formally the information model of the portal M_p is described by two:

$$M_p = \langle O_p, IC_p \rangle,$$

where O_p is the ontology of the portal; IC_p is information content of the portal.

Ontology is the core, basic component of information model of the portal. Not only it describes the knowledge system of the portal, but also gives a formal structure for the presentation of its content.

In the paper ontology conception is used in the sense that how it is applied to the computer science and artificial intelligence. In particular, we adhere to the determination of the ontology given in the paper according to which the ontology is an exact specification of conceptualization. Here, some abstraction is realized by the conceptualization, i.e. a simplified view of the world constructed for a specific purpose. Conceptualization includes the objects, conception and other entities of the studied fields, as well as the relation between them.

For the presentation of ontology, the formalism ensuring the description of the problem and domain fields conception of the portal and diverse semantical relations between them is necessary. An important demand for it is the ability to construct the concepts of the domain field (DF) in the hierarchy of "general-private" and support for the inheritance of properties in this hierarchy. This formalism should also provide the ability to set constraints on the properties of domain fields and the organization of logical inference.

The formalism meeting the above mentioned needs, which is presented as a met ontology form, is proposed in the paper.

$$O = < C, R, T, DA, F, Ax) \qquad (1)$$

where

$C = \{C_1, \ldots, C_n\}$ is the final non-empty set of classes describing the conception of some domain and problem fields:

$R = \{R_1, \ldots, R_m\}$, $R_i \subseteq C \times C$, $R = R_T \cup R_p \cup R_A$ is the final set of binary relations given in the classes (concepts); here, R_T is antisymmetric, transitive, unreflective binary relations of the inheritance giving partial order in the set of concepts C; R_p is binary transitive relation of the inclusion (partial-whole); R_A is the final set of associative relations; T is a set of standard types; $D = \{d_1, \ldots, d_n\}$ is a set of domains; $d_i = (s_1, \ldots, s_k)$ is the value of standard types s_i;

$TD = T \cup D$ is a generalized data type including a set of standard types and a set of domains;

$A = \{a_1, \ldots, a_w\}$, $A \subseteq C \times TD \cup R_A \times TD$ is the final set of attributes, i.e. binary ratio of the form $a_i(C_j, td_k)$ or $a_i(R_j, td_k)$ describing the properties of the concepts C and ratio R_A;

F is a set of restrictions on the value of concepts and ratio attributes, i.e. predicates of the form $p_i(e_1, \ldots, e_m)$, where e_k is the name of either attribute $(e_k \in A)$ or constant $(e_k \in td_j,\ where\ td_j \in TD)$;

A_x is a set of axiom determining semantic groups and ontology ratio.

The ratio R_T serves to define the hierarchies on concepts. Its peculiarity is that while inheriting from a parent

Jamalova Jala Siyavush, graduated from the Azerbaijan State Technical University on mechanical engineering specialtyand. work as an assistant on my specialty. The direction of her scientific work - systems and networks. Scientific direction area - computer complexes, systems and networks

class, its class-descendant is passed not only to all the attribute, but also their relations.

The inclusion relation R_p is endowed with the transitivity property, due to while searching for the objects (direct instance of classes), it is possible to carry out a transitive closure in this relation.

A set of associative relations R_A is determined by need for the knowledge about the links between the objects in specific problem and subject areas. The proposed formalism enables to specify such relations attributes specializing the relationship between the arguments (objects).

Information content portal IC_p is constructed on the base of formal structures specified in meta-ontology O (formula (1)):

$$IC_p = \langle I, R_I, V, A_I, R_{IC}, R_{IR}, R_{IA}, R_{VTD} \rangle, \quad (2)$$

where

$I = \{i_1, \ldots, i_n\}$ is the final set of direct instance classes of ontology;

$R_I = \{ri_1, \ldots, ri_k\}$ is the final set of specified relations (instances of relations) $ri_i(i_j, i_k)$ between instances of classes.

$V = \{v_1, \ldots, v_q\}$ is the final set of specific value of the generalized type TD;

$A_I = \{ai_1, \ldots, ai_w\}$ is the final set of specified attributes, i.e. binary relations $a_i i_i(i_x, v_y)$ or $a_i i_i(ri_x, v_y)$ between instances of classes or the ration of specific values;

$R_{IC} \subseteq I \times C$ is a binary incidence relation between a set of instances I and a set of classes C;

$R_{IR} \subseteq R_I \times R$ is a binary incidence relation between a set of instances R_I and a set of relations R;

$R_{IA} \subseteq A_I \times A$ is a binary incidence relation between a set of specified attributes A_I and a set A;

$R_{VTD} \subseteq V \times TD$ is a binary incidence relation between a set of values V and generalized type TD.

III. Knowledge Portal Ontology

From the substantive point of view, the ontology of the portal knowledge serves to present the concepts needed for describing both scientific activity and scientific knowledge on the whole, and particularly specific knowledge area.

In order to simplify the configuration of the portal to the selected area of the knowledge, the ontology of the portal is divided into subject-independent (basic) ontologies and the knowledge portal domain ontology (subject ontology).

The ontologies of scientific activity O_{RA} and scientific knowledge O_{SK}, which doesn't depend on the domain of knowledge portal, are chosen as a base. These ontologies are constructed on the base of the meta-ontology formalism O described in section 1.

The portal knowledge domain can include several related subject areas. In this connection, the ontology of the knowledge portal domain O_{KD} can include the ontologies of several subject domains O_{SDi} constructed on the base of scientific knowledge ontology O_{SK}.

Thus, the knowledge portal ontology O_P has the following form:

$$O_P = \langle O_{RA}, O_{KD} \rangle,$$

where O_{RA} is scientific activity ontology; $O_{KD} = \{O_{SDI}, \ldots, O_{SDm}\}$ is knowledge portal domain.

The scientific activity ontology acting as an ontology of the problem domain of the knowledge portal is constructed on the base of the ontology proposed in [6] for the description of research projects, and in fact it is an upper level ontology. It includes basic classes of concepts related to the organization of scientific and research activities, such as the *Explorer*, *Organization, Event, Activity (project), Publication*, and so on. This ontology also includes the *Information resource*, which serves to describe the information resources presented on the Internet.

The scientific knowledge ontology fixes the basic content structures used to construct subject-matter ontologies. In particular, this ontology contains meta-concepts defining the structures to describe the concepts of a particular knowledge area as: *Science Section, the Research Method, the Object of Research, the Subject of Research, the Scientific Result*.

The concepts of each subject ontology are realizations of the meta-concepts of scientific knowledge ontology and can be ordered in the "general-private" hierarchy.

The concepts of basic ontologies are related to each other by associative relations, the choice of which was made not only based on the completeness presentation of the problem and subject domains of the portal, but also taking into account the ease of navigating through its content and information retrieval therein.

The most important associative relations are as follows:

"scientific direction" - connects events, publications, organizations, researches by science sections;

• "describes" - specifies the relationship of the publication with the scientific result, object or research method;

• "uses" - connects the research method to the activity, researches or science section;

• "applies to" - relates the research method to the research object;

• "result" - connects the scientific result with the activity;

• "resource" -connects an information resource with events, publications, researches, methods and research objects;

• "works in" - connects the researcher with the organization which he works for.

It should be noted that, the last relation has three additional attributes: *position, acceptance date and dismissal date*, serving to clarify, in which position and since when the researcher worked in the organization.

The axioms included in the basic ontologies allow to derive additional associative relations between objects.

As a sample of such an axiom, an axiom that derives a new associative relation "works _ in" is given:

$A1$: *if works _ in* $(И_1, O_1)$, && *includes* (O_2, O_1) *then works in* $(И_1, O_2)$, $И_1 \in$ *Researcher,* $O_1, O_2 \in$*Organization.*

IV. Presentation of Data and Information Resources

By introducing formal descriptions of the knowledge domain concepts in the form of object classes and relations between them, the portal ontology defines the structures for representing real objects and the relationships between them. In accordance with this, the data on the portal is presented as a set of diverse information objects and connections, which together constitute the information or portal content.

Formally, each information object (IO) is an element of the set *I*, and the connection between information objects - the elements of the class R_I (see the formula (2)).

The content of the IO is an instance of any ontology class and presents the description of some objects of the modeled knowledge area or the information resource relevant to it. The semantics of the connections established between the IO is determined by the relationships specified between the corresponding classes of the ontology.

An important component of the portal content is the description of information resources. As mentioned above, each resource presented on the portal corresponds to such an ontology concept as the *Information resource*, the set of attributes and the connections of which is based on the standard Dublin Core [7]. The description of the resource includes an instance concept of the Information resource and a set of instances of the relations connecting this concept with other concepts of ontology (*Organizations*, *Researchers*, *Publications, Events, Sections and so on*).

V. Organization of Substantial Access

On the base of proposed information model, not only the content of the knowledge portal is constructed, but also substantial access to the systematized knowledge and information resources of the modeled knowledge area is organized. This function ensures advanced navigation and search facilities.

A. Navigation

While navigating through the portal, it is possible to select information objects of a certain class, a detailed view of each IO, as well as information resources, connections in which are contained in the viewed IO.

Navigation through the portal begins with the selection of a certain class in the ontology concepts tree constructed on the base of relation R_T . In this case, the user is given a list of information objects of the selected class $L(C_X)$, which is shown in the form html-page containing a set of links to these objects. For large lists of objects, a composite page is formed including a list of pages with navigation elements on this list.

It should be noted that while forming a list of the objects of the selected class C_X , a transitive closure with respect to R_T takes place. In consequence of this resulting list $L(C_X)$ will include both objects of the required class C_X and the objects of its classes of successors i.e.

$$L(C_X) = \{i_i | i_i \in C_X \cup C_T\},$$

where $C_X \in C$,

$$C_T = \{C_Y \in C | \exists R_T (C_X, C_Y)\}$$

If on information objects of class C_X the ratio of inclusion R_p is given at the user's request, a set of the objects included in $L(C_X)$can be represented either as a list or as a tree.

Information about a particular object and its connections are also displayed in the form html-page. (See the figure), the format and content of which depend on the class of this object and the relations defined for it and visualization template. In this case, the objects associated with this object are presented on its page in the form of hyperlinks, by which one can go to their detailed description.

Further navigation through the portal is a process of transition from one information object to another according to the relationships established between them - instances of associative relations R_A.

For example, while viewing information about a particular project (see the figure), one can see not only the values of its attributes, but also its connections with other objects. Using the presented connections as navigation elements, one can proceed to view detailed information about it as for direct relations (about the research object, about the research methods used in the project and the scientific results obtained in the course of its implementation), and on the reverse (about the participants of the project, describing its publications and information resources).

B. Search

The search for information objects is based on the ontology due to which the user is given an opportunity to set the query in terms of the portal knowledge area.

Search queries are specified through a special graphical interface controlled by the knowledge portal ontology. When a user selects a class of required information objects, a search form is automatically generated in which you can set constraints on the values of the attributes of the objects of the selected class, as well as on the values of the attributes of the objects associated with this object by associative relationships. In this case, the conditions for the values of the attributes of the corresponding relations can also be specified.

Formally, the search query constructed in terms of the ontology is as follows:

$$Q = \{i_X \in C_X | P(i_X) \& Rd(i_X) \& Rb(i_X)\},$$

where $C_X \in C$ is the class of required objects.

$P(i_X) = \underset{i=1}{\overset{n}{\&}}(a_i(i_X, v_i) \& \underset{j=1}{\overset{m}{\&}} p_j(v_i, c_j))$ is the description of the required object properties i_X class, where n is the number of user-defined attributes used, a_i; m is the number of user-defined predicates p_j on the value v_j attributes a_i; c_j are constants $c_j \in td_j$, where $td_j \in TD$;

$$Rd(i_X) = \underset{k=1}{\overset{d}{\&}}(ri_k(i_X, i_Y), ri_k \in R_k, i_Y \in C_Y | P(ri_k) \& P(i_Y))$$

is the description of the objects consisting of the objects of class C_X in "direct" relations, where d is the number of user-defined "direct" relations ri_k – are descriptions of the instance relation properties and properties of the object $R_k \in R_p \cup R_A$; $P(ri_k)$ and $P(i_Y)$ consistently, the structure of which is similar to the description structure $P(i_X)$;

$$Rb(i_X) = \underset{i=1}{\overset{b}{\&}}(ri_1(i_Z, i_Y), ri_1 \in R_1, i_Z \in C_Z | P(ri_1) \& P(i_Z))$$

is the description of the objects consisting of the objects of class C_X in "inverse" relations, where b is the number of user defined "reverse" relations $R_1 \in R_p \cup R_A$; $P(ri_k)$ and $P(i_Z)$- descriptions of the relation instance properties ri_i and the properties of the object i_Z respectively.

As a result of the query implementation Q, the search for the objects of a given class C_X can be performed (at the user's request) taking into account the transitive closure with respect to the ratio R_T. In this case, when processing the request, a set of objects will be considered $L(C_X) = \{i_i | i_i \in C_X \cup C_T\}$, where $C_T = \{C_K \in C | \exists R_T(C_X, C_K)\}$.

As a result of the search, or while passing through a specific connection of any IO, can be obtained by a sufficiently large list of objects $L(C_X)$ (for example, a list of all the participants in a major project or conference). To manage this situation, a filtering mechanism is introduced, which allows to select from the list $L(C_X)$ by specifying a filter.

The content of the filter is a set of conditions determining the permissible values of the attributes of information objects and the requirements for the existence of connection with other IO. Formally, the filter is specified by an expression similar to the search query.

This method allows, for example, to filter out the set of project participants both by age or scientific degree (conditions for the attribute), and by the research methods they use (conditions on the associated object).

Conclusion

The article considers the information model of the scientific knowledge portal providing the user with substantial access to the systematized knowledge and information resources of a certain knowledge field. This model not only provides a formal description of the system concepts of the problem domain of the knowledge portal, but also supports all its functionality. On the base of his model, the knowledge system and internal data storage of the portal are constructed, its content, navigation and search are organized.

By using ontologies as a base for an information model, the knowledge portal is not just another resource catalog for a given topic, but a knowledge and data network that supports convenient navigation and substantial search.

Separation of portal ontology into subject-independent and subject-matter ontologies makes the portal customizable to any field of scientific knowledge.

On the base of the proposed model, together with the Institute of Archeology and Ethnography of the SB RAS, an archaeological knowledge portal [8] was developed, providing substantial access to a wide range of users for systematic knowledge and information resources in archeology and ethnography. Later, this approach was applied in the development of a knowledge portal on computer linguistics [9].

Reference

[1] Borovikova O.I., Zagorulko Y.A. Organization of knowledge portal on the base of ontology//Computational linguistics and intellectual technology. International seminar "Dialogue-2002". Moscow. 2002. P.76-82.

[2] Zagorulko Y.A., Borovikova O.I. Construction technology of ontology for scientific knowledge portal. "Information technology" 2007. 2 ed. P. 42-52.

[3] Gavrilova T.A., Khoroshevckiy V.F. Knowledge base of intellectual system. Training aid. Peter.2001. 384 p.

[4] Guarino N. Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam: IOS Press. 1998. P.13-15.

[5] Gruber T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing// International Journal of Human-Computer Studies. November 1995. Vol.43. Issues 5-6. P.907-908.

[6] Benjamins V.R., Fensel D. Community is Knowledge in (KA)2/B.R.Gaines, M.A. Musen (eds.) // Proceedings of the 11 th Banff Knowledge Acquisition for Knowledge-based Systems work-shop, KAW'98. Calgary. SRDG Publications, Department of Computer Science, University of Calgary, 1998.

[7] Using Dublin Core. http://dublincore.org/documents/usageguide/

[8] Andreeva O.A., Borovikova O.I., Bulgakov S.V., Zagorulko Y.A, Sidorova E.A., Sirkin B.G., Kholoshkin Y.P. Archeological portal knowledge: substantial access to knowledge and information resource on archeology//National conference on artificial intellect with international collaboration. M. 2006. P.832-840.

[9] Borovikova O.I., Zagorulko Y.A. Zagorulko G.B. Conenco I.S., Sokolova E.G. Development of knowledge portal according to computational linguistics. National conference on artificial intellect with international collaboration. M. 2008. P. 380-388.

I, **Jamalova Jala Siyavush** was born in Baku in 1965. In 1995 I graduated from the Azerbaijan State Technical University on mechanical engineering specialtyand I work as an assistant on my specialty.

The direction of her scientific work - systems and networks. Scientific direction area - computer complexes, systems and networks.