

## Methodology for knowledge portals development: background, foundations, experience of application, problems and prospects\*

O. I. Borovikova, L. S. Globa, R. L. Novogradskaya, M. Yu. Ternovoy,  
G. B. Zagorulko, Yu. A. Zagorulko

**Abstract.** The paper discusses an experience of using the methodology for the development of knowledge portals which provide systematization and integration of scientific and engineering knowledge and information resources as well as the content-based access to them. To provide a sufficiently complete and consistent representation of knowledge and information resources, their systematization and integration are performed on the basis of ontology. The suggested methodology has been successfully applied to the development of knowledge Internet portals on archaeology, computational linguistics, strength of materials and Antarctic data. **Keywords:** scientific information resources, knowledge portal, ontology, content-based access, knowledge-driven navigation, semantic search

### 1. Introduction

Recently, a great amount of information resources relating to various areas of knowledge has been accumulated in the Internet. However, access to these resources and their use is rather complicated as they are disembodied and ill-structured, or distributed over various Internet sites, electronic libraries and archives. At the same time, researchers need an efficient content-based access to scientific papers, research reports and other information resources containing descriptions of methods and approaches developed in the framework of the branch of science or engineering interesting to them. (Here by the content-based access we mean advanced semantic search and knowledge-driven navigation through information resources relating to a certain topic.)

Traditional search engines are not able to provide efficient access to information resources. They usually return to the user hundreds and thousands of irrelevant documents that hinder comprehension and choice of the information needed. The reason for this is that modern search engines use primarily keyword search mechanisms, which are insensitive to query semantics, and index Web resources with virtually no tools for analysis of the information presented in them. Besides, as a rule, search engines do not provide convenient navigation through information resources that have been found.

---

\*Partially supported by Russian Academy of Sciences, Integration Project No.15/10.

In recent years, there have been some attempts to provide semantic access to information allocated in the Internet by using ontologies [1] for describing the semantics of Web resources. So, the Semantic Web approach [2] suggests to supply Web resources with annotations in the RDF [3] or OWL [4]. Using such annotations, intelligent search agents can provide more relevant responses to a user query as compared with traditional search engines. Some progress has been made in this direction, though the general situation has not improved, since Web resources annotated in such a manner are an infinitesimal drop in the sea of the Web.

To provide content-based access to information resources related to a given area of scientific or engineering knowledge, we have suggested a methodology for the development of specialized Internet portals – knowledge portals [5, 6] supporting:

- a sufficiently complete and consistent representation of some branch of science or engineering, its components and various aspects of research activity (persons, organizations, events, methods, objects and results of research, etc.);
- integration of knowledge and information resources relating to a given branch of science or engineering into a uniform information space;
- advanced semantic search and knowledge-driven navigation through the integrated knowledge and information resources.

An important feature of these portals is the possibility of their declarative adjustability to any area of knowledge both at the development and operation stage. This possibility allows us to track up the dynamics of appearance of new knowledge and information resources related to the topics of portals and in that way to secure maintenance of their topicality and utility.

The knowledge portals support the functionality described above and are declaratively adjustable to any area of knowledge due to using ontology as their conceptual basis and information model.

In this paper, we would like to present a methodology for ontology-based development of knowledge portals providing a content-based access to knowledge and information resources. Also, we are going to discuss an experience of applying this methodology, problems arising in the process of its applications and ways to cope with them.

## **2. Ontology-based development of knowledge portals**

In the suggested methodology, an ontology is used both at the time of knowledge portal development and at its run time. It serves to describe the knowledge area of the portal and to specify its data storage structure. It supports systematization and integration of relevant information resources and content-based access to them.

## 2.1. Formalism for building portal ontology

To support an effective representation of a portal ontology, a formalism providing description of both concepts of its problem and subject domains and diverse semantic relations between them is required. An important requirement to this formalism is the possibility of ordering concepts in a “generic-specific” hierarchy and supporting inheritance of properties through this hierarchy. Also, this formalism should allow us to specify a set of axioms defining semantics of ontology classes and relations, as well as to impose constraints on the values of properties of objects which are instances of ontology concepts:

We suggest a formalism meeting the requirements described above. It is described as follows:

$$O = \langle C, R, T, D, A, F, Ax \rangle \quad (1)$$

where  $C = \{C_1, \dots, C_n\}$  is a finite nonempty set of classes describing concepts of some subject or problem domain;

$R = \{R_1, \dots, R_m\}$ ,  $R_i \subseteq C \times C$ ,  $R = \{R_T\} \cup \{R_P\} \cup R_A$  is a finite set of binary relations defined on classes (concepts)  $C$ :

$R_T$  is a strict partial order on the set of concepts  $C$  defining an inheritance relation,

$R_P$  is a binary transitive relation of inclusion (“whole-part” relation),

$R_A$  is a finite set of associative relations;

$T$  is a set of standard data types;

$D = \{d_1, \dots, d_k\}$  is a set of domains  $d_i = \{s_1, \dots, s_d\}$ , where  $s_j$  is a value of the standard type *string*;

$TD = T \cup D$  is a generic data type containing the set of standard data types and the set of domains;

$A = \{a_1, \dots, a_w\}$  is a finite set of attributes, i.e. binary relations like  $a_i(C_j, td_k)$  or  $a_i(R_j, td_k)$ , where  $td_k \in TD$ , describing the properties of concepts  $C$  and relations  $R_A$ ;

$F$  is a set of constraints imposed on values of attributes concepts and relations, i.e. predicates like  $p_i(e_1, \dots, e_m)$ , where  $e_k$  is either a name of an attribute ( $e_k \in A$ ) or a constant ( $e_k \in td_j$ , where  $td_j \in TD$ );

$Ax$  is a set of axioms defining additional semantics of ontology classes and relations.

A feature of the relation  $R_T$  is providing inheritance of all attributes and all relations passed on from a parent class to a descendant class.

The set of associative relations  $R_A$  is defined by a developer of a concrete portal ontology. Availability of these relations allows us to organize content-based search and navigation through portal content. An important feature of relations  $R_A$  is that they can have their own attributes which specify a link between arguments.

## 2.2. Ontology structuring and building

To meet the portal objectives, the ontology must be well-structured. Therefore, the portal ontology is divided into domain-independent (basic) ontologies and subject domain ontology.

A portal ontology is built on the basis of the formalism described above (see Formula 1), according to the methodology suggested in [7]. The main principle of this methodology is building an ontology by means of completion and evolution of basic ontologies, which considerably simplifies the creation of portal ontology and its maintenance.

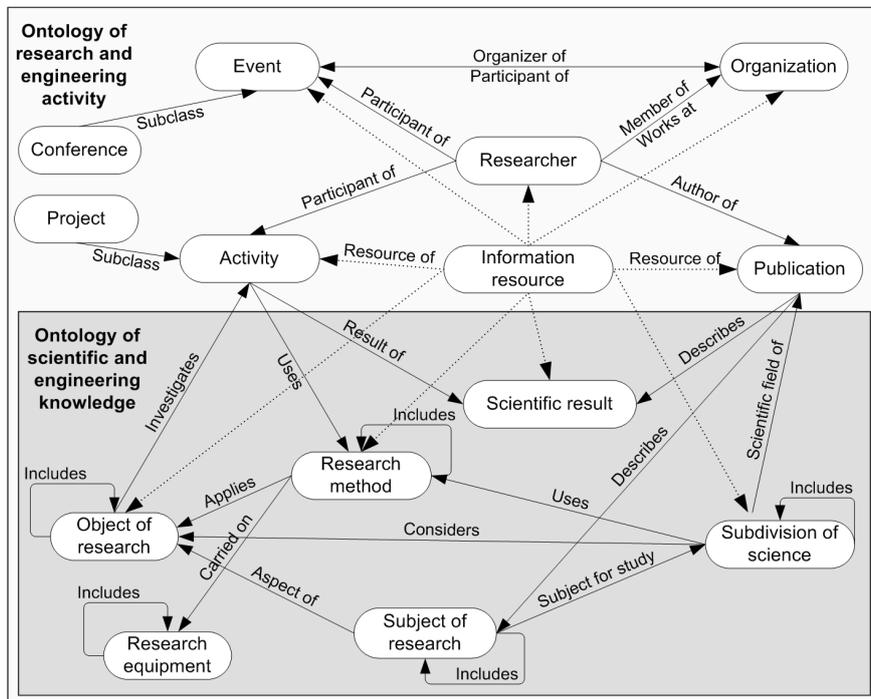


Figure 1. Basic ontologies

Two ontologies were selected as basic (see Figure 1): the ontology of research and engineering activity which constitutes the basis of the portal's problem domain ontology and the ontology of scientific and engineering knowledge used as a basis for building a subject domain ontology.

The *ontology of research and engineering activity* is based on the ontology suggested in [8]. Practically, it is a top-level ontology including base classes of concepts related to the organization of research and engineering activities such as *Person (Researcher)*, *Organization*, *Event*, *Activity*, *Publication*. These classes are used for describing participants of research and engineering activity, scientific events, research programs and projects, vari-

ous types of publications and materials represented in printed or electronic format (such as monographs, articles, reports, proceedings of conferences, periodicals, photo and video data, etc.). This ontology also includes the class *Information resource* that serves for describing information resources presented in the Internet. The set of attributes and relations of the *Information resource* is based on the Dublin Core standard [9].

*The ontology of scientific and engineering knowledge* is virtually a meta-ontology. It states basic structures used for building the subject domain ontology describing a specific area of knowledge or branches of science or engineering. In particular, this ontology contains meta-concepts specifying structures for the description of concepts of specific subject domains, such as *Subdivision of science*, *Research method*, *Research equipment*, *Object of research*, *Subject of research* and *Scientific result*. Using these meta-concepts, we can describe divisions and subdivisions significant for a given science or kind of engineering, determine classification of methods and objects of research or engineering, as well as equipment used to conduct research, and describe the results of research and engineering activity.

The concepts of basic ontologies are interconnected with each other by associative relations (see Figure 1), selection of which is actualized taking into account not only preciseness and completeness of representation of the subject domain of a portal, but also convenience of navigation through portal content and information search.

Axioms included into basic ontologies make it possible to derive additional associative relations between objects (instances of a definite class of ontology) and have the following form:

$$\text{If } r_{p1}(c_1, c_2) \wedge \dots \wedge r_{pm}(c_{m-1}, c_m) \text{ then } r_c(c_j, c_k),$$

where  $r_{pi} \in R, pi \in 1..n, c_i \in C_i \subseteq C, i \in 1..m, r_c \in R_A, c_j \in C_j \subseteq C, c_k \in C_k \subseteq C, j, k \in 1..m$ .

Let us give an example of an axiom which derives a new instance of the associative relation “Works at” (this relation links a researcher with an organization where he works; see Figure 1):

*Ax1: if Works at (P1, O1)  $\wedge$  Includes (O2, O1) then Works at (P1, O2), P1  $\in$  Researcher, O1, O2  $\in$  Organization.*

The main difficulty in building a portal ontology is development of subject domain ontology. This ontology describes a given branch of science or engineering, and therefore its building requires great efforts from experts in this branch of science or engineering.

The concepts of subject domain ontology are at the same time realizations of meta-concepts of ontology of scientific knowledge and can be ordered in the “generic-specific” hierarchy.

### 2.3. Structure of a portal content

The ontology developed within the framework of our methodology not only describes the problem and subject domains of a portal but can be used as a specification of its content structure. Indeed, by introducing formal description of concepts in the form of classes of objects and relations between them, a portal ontology defines structures for representing real-world data and relations between them. In accordance with this, data in the portal content are presented as a set of interrelated information objects (IO) of various types.

Formally, each IO corresponds to some concept of ontology (is an instance of a definite class of ontology) and has structures defined by it. Syntax and semantics of connections between information objects are defined by relations established between the corresponding classes of ontology. A collection of such information objects constitute the content of a portal.

An important part of the portal content is the descriptions of information resources integrated into the portal's information space. Here each resource is represented by information object that is an instance of the concept *Information resource* and a set of instances of relations linking this object with other information objects.

The information object representing an information resource has attributes such as "Title of the resource", "Address in the Internet" (URL), "Resource type", "Language", etc., and can be linked with a person (by relation "Author of"), organization (by relation "Resource of organization"), event (by relation "Reference to event"), results of research (by relation "Describes"), etc.

### 3. Adjustment of the portal to a given area of knowledge and management of its content

Adjustment of the portal to a given area of scientific or engineering knowledge is performed with the help of an ontology editor.

The ontology editor has been selected and designed in such a way that it is easy to understand and use for experts not experienced in computer science and programming. In particular, to meet these requirements, we refused to use such a popular editor as Protégé [10].

Moreover, the ontology editor has been designed taking into account its use in the distributed development of ontologies. Therefore, it was implemented as a Web-application and has a procedure for granting privileges to experts of different levels.

Using the ontology editor, an expert can create, modify and delete any elements of the ontology: classes of concepts, relations, and domains.

When a class is created, it takes its name, a set of attributes defining

various properties of concepts, as well as constraints on the attributes values. A parent of the class under creation can be selected from a set of the classes that have already been created. Thereby this class inherits from the parent class not only all its attributes, but also its relations, whereas the parent class gets linked to a new class by a “subclass” (“is-a”) relation.

Classes of ontology can be linked by associative relations ( $R_A$ ). The peculiarity of these relations is their ability to have their own attributes that specify the nature of the connection between arguments of relation.

To make representation of information more convenient for the user, the possibility of adjustment of knowledge and data view is provided. For this purpose, templates of view of objects for each class of ontology and templates of view of reference to such objects are specified.

A template of view of class objects contains all its attributes and all relations associated with it. By default, attributes of class and its relations are depicted in the same order and mode as defined in the ontology; however, the user can change this order.

A template of view of reference to a class object can include both values of attributes of this class and values of attributes of classes linked with it by relations and attributes of these relations. Values of attributes included in this template are used for building text presentation of hyperlink to a class object.

Management (extension and edition) of portal content is performed by an expert with the help of a data editor allowing one to create, modify and delete information objects and relations between them.

Operation of a data editor is based on portal ontology. Therefore, when a new information object is created, the expert firstly selects from a tree of concepts the necessary class of ontology (usually a hierarchy of ontology classes is represented as a tree). Then, based on a description of this class presented in ontology, a form for data input is automatically generated.

Simultaneously with the object creation, the expert can specify its connection with other objects already existing in the portal content. These connections and their attributes are defined by the corresponding relations of ontology; and a form for their input is automatically generated on the basis of a description of these relations.

Correct creation and deletion of information objects and relations between them as well as modification of values of their properties are supported by a data editor on the basis of constraints and axioms defined in ontology.

#### **4. Providing content-based access to portal content**

Content-based access to systemized knowledge and information resources integrated in the portal content is provided by a user interface which includes advanced facilities for searching and navigation.

#### 4.1. Knowledge-based navigation

Navigation through portal content is based on ontology. For this purpose, both taxonomic relations  $R_T$  defining the hierarchy of concepts and a set of associative relations  $R_A$  connecting information objects (instances of classes) are used.

During navigation through portal content, ontology enables the choice of information objects of a required class, filtration of the list of such objects, detailed browsing of each information object and transition along the ontological links (relations) from this IO to other IOs, as well as browsing of information resources references to which are contained in this IO.

Navigation starts with the choice of a definite class  $C_X$  in the tree of concepts of the ontology built on the basis of the relation  $R_T$ . As a result of this action, a user gets a list of information objects  $L(C_X)$  corresponding to the selected class  $C_X$  represented as an html-page containing a set of hyperlinks to these objects.

It should be noted that, when the list of objects is formed, the transitive closure of  $R_T$  relation is performed. Due to this fact, the resulting list will include both the objects of the sought class  $C_X$  and the objects of its class-successors, i.e.

$$L(C_X) = \{i_i | i_i \in C_X \cup C_T\} \quad (2)$$

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

If the relation of inclusion  $R_P$  is defined on information objects of the class  $C_X$ , a set of objects from  $L(C_X)$  can be represented as either a list or a tree.

Information on a concrete object and its connections is presented in the form of an html-page (see Figure 2). The format and content of this page depend on the class of the object and on the relations defined for this class. At the same time, objects linked with a given object are presented on its page by their hyperlinks. Using these hyperlinks, the user can proceed to a detailed description of these objects.

Further navigation through portal content is a process of transition from some information objects to others along the ontological links (instances of associative relations  $R_A$ ) defined between them.

For example, when we are looking through information on a specific project (see Figure 2), we can see both values of its attributes and its connections with other objects. Using the connections presented as elements of navigation, we can proceed to browsing detailed information about the project, using both direct relations (about the object of research, methods used in the project, and results obtained in the project) and inverse relations (about the participants of the project (persons and organizations), publications and information resources describing this project).

<b>Activity name</b>	AGILE	
<b>Activity description</b>	Automatic Generation of Instructions in Languages of Eastern Europe	
<b>Start date</b>	1 January 1998	
<b>Finish date</b>	31 December 2000	
<b>Project status</b>	completed	

Object relations		
Investigates_object		
Objects of investigation	Language	
Manual		

Activity_result		
Scientific result		
AGILE System		

Activity direction		
Subdivision of science		
Automatic generation of text		

Scientific result		
KPML		

Relations with object		
Person_Participant activities		
Person	Rol	
Bateman, J.A.	participant	
Hana, J.	participant	
Hartley, T.	coordinator	
Kruijff, G.-J.	participant	
Kruijff-Korbayová, I.	participant	
(Total: 11)		

Organization_Participant activities		
Organization		
Information Technology Research Institute, University of Brighton, ITRI		
Institute for Applied Linguistics, University of the Saarland		
Institute of Formal and Applied Linguistics (Charles University), ÚFAL		
Institute of Information Technology, Bulgarian Academy of Sciences, IIT BAS		
РосНИИ искусственного интеллекта, РосНИИ ИИ		

Publications related to activities		
Publications		
Bateman, J.A., Hana, J., Hartley, T., Kruijff, G.-J., Kruijff-Korbayová, I., Skoumalová, H., Stavkova, K., Teich, E., Соколова, Е.Г., Шапов, С.А., A multilingual system for text generation in three slavic languages, 2000		
Bateman, J.A., Kruijff, G.-J., Kruijff-Korbayová, I., Skoumalová, H., Teich, E., Шапов, С.А., Resources for multilingual text generation in three Slavic languages, 2000		

Activity Information resource		
Internet resources		
AGILE Home page		

Figure 2. Presentation of an information object and its connections

As we move along a certain link of any information object, we can obtain quite a long list of objects (for example, a list of printed works of a well-known scientist). To solve this problem, the procedure of list filtration has been implemented.

The filter is a set of conditions that define admissible values of attributes of an information object and requirements imposed on the existence of connections with other IOs. This method allows one to filter a set of publications by date (condition on attribute), by the scientific result described in this publication or by the object of research (conditions on the object connected with a given object).

#### 4.2. Semantic search

Since search, like navigation, is based on ontology the user can formulate his query in terms of a portal knowledge area. Basic elements of such a query

are concepts and relations of ontology, as well as constraints imposed on the data being searched.

To define constraints put on the objects connected by associative relations with the object sought, the user has a possibility to define conditions on the values of any attributes of these objects. Thereby the user can define conditions on the values of attributes of the corresponding relations.

Formally, a query specified in terms of ontology looks as follows:

$$Q = \{i_X \in C_X | P(i_X) \wedge Rd(i_X) \wedge Rb(i_X)\} \quad (3)$$

where  $C_X \in C$  is a class of objects sought;  $P(i_X)$  is a description of properties of the sought object  $i_X$  of the class  $C_X$ ;

$Rd(i_X)$  is a description of objects connected with the objects of the class  $C_X$  by “direct” relations;

$Rb(i_X)$  is a description of objects connected with the objects of the class  $C_X$  by “inverse” relations.

The result of processing the query  $Q$  will be a set of objects of the sought class  $C_X$  which satisfy the properties (constraints) defined in the query.

The retrieval queries are formed by means of a special graphic interface driven by portal ontology. When a user selects a class of the objects sought, a retrieval form is generated automatically, in which the user can define constraints on the values of attributes of the sought object, as well as on the values of attributes of the objects connected with the sought object by associative relations.

Note that the search of objects of the class  $C_X$  can be executed taking into account the transitive closure of relation  $R_T$ . In this case, when processing the query  $Q$ , objects from the set  $\{i_i | i_i \in C_X \cup C_T\}$ , where  $C_T = \{C_K \in C | \exists R_T(C_X, C_K)\}$ , will be considered.

For example, the query “Find finished projects carried out from year 2000 to year 2005 that used a method of semantic analysis of text for the processing of Russian business letters” will be presented as follows:

**Class Project:**

**Attribute Project stage** = “finished”

**Attribute Start date**: ( $\geq 2000$ )

**Attribute Finished date**: ( $\leq 2005$ )

**Relation Investigate:**

**Class Object of research:**

**Attribute Object name** = “Business letter”

**Attribute Language** = “Russian”

**Relation Apply Method:**

**Class Research method:**

**Attribute Method name** = “Semantic analysis of text”

Retrieval queries are formed by means of a special graphic interface driven by portal ontology. When the user selects a class of the sought information objects, a retrieval form is generated, in which the user can define constraints on the values of attributes of the object sought and of the objects connected with the sought object by associative relations (“direct” and/or “inverse”) selected by the user from the list of relations of the class  $C_X$ .

## 5. Methodology application to knowledge portals development

The methodology has been used for the development of knowledge Internet portals on archaeology, computational linguistics, strength of materials and Antarctic data.

### 5.1. Knowledge portal on archaeology

A knowledge portal on archaeology [11, 12] was developed at A.P. Ershov Institute of Informatics Systems in cooperation with the Institute of Archaeology and Ethnography, Siberian Branch of the Russian Academy of Sciences. The portal was built with a view to systemizing and integrating a great amount of accumulated knowledge and information resources on archaeology, as well as in order to provide a content-based access to them.

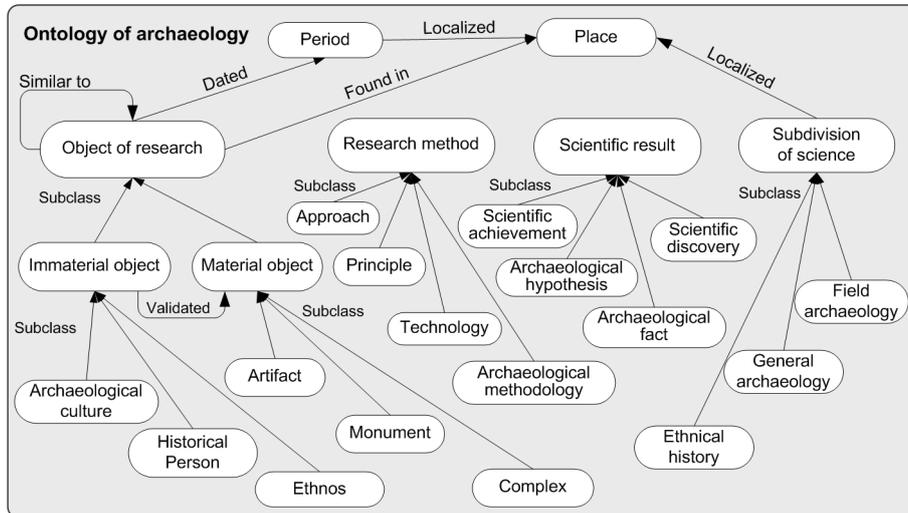
This portal is intended for a wide circle of users, from scientists and lecturers to students and schoolchildren interested in archaeological achievements. Users of these portals can get an idea of the modeling area of knowledge and find information about the activities carried out in this area. The portal primarily contains information about scientists, organizations, research groups and their activities. An important part of the portal’s content is description of Internet resources: the Internet sites of organizations, projects, conferences, topic portals and catalogues, as well as separate Internet pages containing information of graphic, multimedia and text types.

The foundation of the portal is the ontology of archaeology built on the basis of the scientific and engineering knowledge ontology.

This ontology (see Figure 3) includes four basic hierarchies:

*The subdivision of archaeology hierarchy* presents the structure of the main directions of scientific activities taking place in archaeology. Archaeology can be subdivided into *General archaeology*, *Field archaeology*, and *Reconstructive archaeology* which, in turn, are subdivided into more specific subdivisions of archaeology.

*The objects-of-research hierarchy* systematizes objects of research in archaeology. These objects have such properties as description of an object, date of discovery, accuracy of dating, etc. Examples of objects of research are *Archaeological Culture*, *Historical Person*, as well as *Artifact*, *Complex*, and *Monument* discovered during archaeological excavations.



**Figure 3.** Ontology of archaeology

*The research methods hierarchy* serves to describe approaches, research methods, principles, technologies, and methods applied to a certain type of archaeological objects.

*The scientific results hierarchy* is intended for the description of results of research activities conducted in archaeology, such as discoveries, hypotheses, new laws, theories, archaeological facts, etc.

It was found that meta-concepts presented in the ontology of scientific and engineering knowledge were not sufficient to present chronological and geographical features of archaeological knowledge. Therefore, the ontology of scientific knowledge has been complemented by new classes – *Archaeological period* and *Place* - typical for the historical sciences.

The class *Archaeological period* serves for dating the objects of research and binding the subdivisions of science to time domains. Information objects corresponding to specific archaeological periods constitute a hierarchy of nesting and historical sequences are described by the time domain.

The class *Place* allows one to associate the subdivisions of science with geographical locations and to define historical and chronological connections between archaeological periods and their geographical locations.

Thus, the portal presents knowledge on the main subfields of archaeology, its objects, methods and results of research, united in accordance with chronological and geographical principles and supplied with text descriptions and references to publications. Information saved in the portal content allows one to organize the content-based access to resources presenting data bases and electronic archives of graphic and bibliographic documents, reports of archaeological activity, as well as descriptions of museum collections.

## 5.2. Knowledge portal on computational linguistics

The knowledge portal on computational linguistics [13] provides effective access to linguistic information resources. The users of this portal are researchers, lecturers, and students involved in this branch of science, as well as specialists who develop systems for natural language processing, analysis and synthesis of speech.

The ontology of computational linguistics built on the basis of the ontology of scientific and engineering knowledge plays the role of the ontology of a portal's area of knowledge. The ontology of computational linguistics includes about 200 concepts organized into five basic hierarchies: *Subdivisions of science*, *Research methods*, *Objects of research*, *Subjects of research*, and *Scientific results*.

*The subdivisions-of-computational-linguistics hierarchy* is based on the classification of the main theoretical directions of computational linguistics and determines significant divisions and subdivisions of computational linguistics. Divisions of computational linguistics are *Machine translation*, *Text processing*, *Speech analysis* and *Speech synthesis*, etc. These general divisions are divided into more specific subdivisions. For example, *Machine translation* includes *Automatic machine translation* and *Automatized machine translation*.

*The objects-of-research hierarchy* sets classification of the objects of research. A *Discourse* as the base object of research is considered as a form of natural language existence and use. In particular, phonetic, syntactic and other linguistic phenomena are taken into consideration, as well as such forms of discourse as *Text* and *Speech*.

*The research methods hierarchy* serves for systematized description of various models and methods that are applied in computational linguistics. Here such groups of methods as *Linguistic methods*, *Computer Science methods*, and *Mathematical methods* are distinguished.

*The scientific results hierarchy* is designed to structure and describe the results of research activities. It includes such kinds of results as *Technologies* and *Software products*, *Applied systems* and *Linguistic resources* (*Lexical Ontology*, *Dictionaries* and *Text Corpora*).

*The subjects-of-research hierarchy* describes *Basic processes* and *Tasks* associated with functioning of units of language in the process of communication, as well as *Applied processes* and *Tasks* which are of practical value and satisfy certain social needs.

All hierarchies of computational linguistics are connected by means of associative relations. Some of these relations are inherited from basic ontologies, while the others are specific relations of a given subject domain.

As the portal is devoted to computational linguistics, its content presents primarily knowledge about the main divisions of computational linguistics,

its objects and subjects of research, its models and methods. As for information resources, they describe the results of activities of organizations and researchers obtained within the frameworks of scientific and commercial projects. These results are technologies, software products, applied systems and traditional linguistic resources, such as vocabularies, linguistic databases, and text and speech corpora.

### 5.3. Knowledge portal on the strength of materials

The knowledge portal on the strength of materials [14] was developed at the National Technical University of Ukraine Kyiv Polytechnic Institute. This portal provides effective access to information resources within a declared subject domain. Development of such kind of portal makes it possible to get access to knowledge and data in materials science domain as well as methods of solution of computational problems in this subject domain. Users of this portal are scientists, engineers and researchers interested in the strength of materials. The portal gives them an opportunity to operate with important theoretical and practical results obtained at specific scientific institutions.

The strength-of-materials ontology includes a lot of concepts organized into four hierarchies: *Research methods*, *Objects of research*, *Purpose of research*, *Research equipment*.

*The research methods hierarchy* serves to describe various methods, procedure, techniques, and practices that can be applied to the strength of materials subject domain. Here such methods as *Methods of strain analysis*, *Deflection method*, *Stress distribution method*, *Evaluation of stress-strain state of construction method* and others are distinguished.

*The objects-of-research hierarchy* classifies objects of research in the strength of materials subject domain. *This hierarchy* describes materials in terms of (due to) their relation to the material groups and specific materials properties. The objects have such properties as types of materials, composition, mechanical properties, technological characteristics, etc. Examples of objects of materials are titanium, aluminum, cast iron and so on.

*The purpose-of-research hierarchy* describes research objectives and purposes such as *Exploration of stability*, *Resistance to deformation*, *Stress sustainability*, etc.

*The research equipment hierarchy* determines *Units*, *Machines*, *Machine-tools*, *Ballers* and other types of equipment and facilities used for the determination of material properties and for conducting experiments.

Thus, the portal content presents knowledge about such branch of science as the strength of materials, its objects, methods and equipment used, united in accordance with material types and their characteristics, supplied with text descriptions and references to publications. Information saved in the portal content allows one to organize the content-based access to infor-

mation resources which present data bases and electronic archives of graphic and tabulated data manuals and reports, information on experiments results description and experiments conducting.

#### 5.4. Knowledge portal on Antarctic data

The knowledge portal on Antarctic data [15] has been developed at the National Technical University of Ukraine Kyiv Polytechnic Institute in collaboration with the National Antarctic Scientific Center of Ukraine (NASC). It is used to provide content-based access to a wide variety of Antarctic data to NASC researchers and scientists from various institutions whose work is related to Antarctic research and researchers from the Academic Vernadsky Station in Antarctica.

The ontology of Antarctic data, built on the basis of the ontology of scientific and engineering knowledge, plays the role of the ontology of the portal's area of knowledge. It includes such hierarchies as *Fields of research hierarchy*, *Research methods*, *Objects of research*, and *Research results*.

*The fields-of-research hierarchy* describes the classification of research areas typical for the Antarctic stations, and includes formal and informal descriptions of the concepts and relations between them. For example, it includes *Meteorological research*, *Geological and Geophysical research*, *Oceanographic research*, *Medical physiology research*, *Biological research*, etc.

*The research methods hierarchy* serves to represent a wide variety of investigation methods used in the subject domain described: *Gravimetric tomography*, *Seismographic analysis*, *Electromagnetic sensing*, *Satellite photograph* and so on

*The objects-of-research hierarchy* includes concepts that define typical objects of investigation. It is important to correlate the name of a research object with the person who explores it. The examples of concepts from Objects-of-research hierarchy are *Bottom surface relief*, *Land surface relief*, *Ice cap*, *Mineral resources base of the shelf zone*, etc.

*The scientific results hierarchy* describes the results of Antarctic activities and their typing. It contains the results of the investigations that have been obtained by scientists. Research results are reflected in publications. Usually the results of scientific research are some models, structures or cards of objects such as *Models of deep structures*, *Oil and gas field structures*, *Mass balance models*, *Geological models*, etc.

*The location hierarchy* serves to describe geographic and administrative territorial location of the objects of research. Availability of this kind of description makes it possible to integrate data and knowledge about the names of objects of research and their geographic location.

All hierarchies of Antarctic data are connected by means of associative relations. Some of these relations are inherited from basic ontologies, while

the others are specific relations of a given subject domain.

The portal content includes, apart from knowledge about the fields of the research conducted at the Antarctic station, information on research objects, methods used and results obtained.

## **6. Related work**

In recent years, there have been attempts to use ontologies to describe the Web resource semantics. There are many tools being developed for semantic annotation of Web-pages and documents, when each document is linked to its semantic content. Using such annotations, the intelligent search agents provide more relevant responses to a user query as compared with the existing engines. For example, to do this, the SHOE system [16] supplies HTML documents with a set of special tags for knowledge presentation, and the above mentioned Semantic Web initiative [2] presumes documents to be supplied with annotations in the RDF language [3]. Some progress has been achieved in this direction; however, this does not improve the situation in general, since Web-pages annotated in such a manner are an infinitesimal drop in the sea of the Web.

It is interesting to consider the so-called Semantic Web Portal [17] developed by using ontologies and other components of Semantic Web technologies.

For example, the Esperanto Portal is a case study of the ODESeW knowledge portal generator [18] developed by the Ontology Group at Facultad de Informatica, Universidad Politecnica de Madrid. It satisfies the information needs of participants of the European project called Esperanto. The information contained and its reliability makes the Esperanto Portal one of the best sources on ontology research. For the Esperanto Portal five different domain ontologies have been developed, namely Project, Documentation, Person, Organization and Meeting. These ontologies describe R&D Projects, especially IST-Projects, and can be reuse. Though these ontologies are connected by several relations, the Esperanto Portal allows the search for objects only on the basis of the attributes of classes presented within one ontology. In addition, this portal is aimed at presenting information on research projects only and cannot be adjusted to another application domain.

The OntoWeb Portal [19] is referred to the so-called community portals. It was also constructed within the framework of a European project of the same name to facilitate the transfer of experience gained on ontologies and Semantic Web technology, from the researchers' community to industry. The OntoWeb ontology provides presentation of all types of information objects necessary for portal operation (e.g. Document, Image, Project, etc). Due to this fact, the project participants can place any information on organizations, events, projects, persons, and documents (scientific papers) in the portal's

content. The drawback of this portal is systematization of information based on the project needs rather than by the topical principle, which makes search for necessary documents difficult.

Another portal using the ontology-based approach should be mentioned. This is the semantic portal MuseumFinland, intended for publishing heterogeneous museum collections on the Semantic Web [20]. It is declared that this portal allows navigation and semantic search for museum exhibits in these collections on the basis of seven culturological ontologies in nine different dimensions: author, owner, material, type of artefact, collection, time of creation, place of creation, place of using, and situation of using. Our inspection of this portal, however, has shown that its search and navigation facilities are rather limited: the user can only search for museum exhibits proper and indicate them as the main element of navigation and search, i.e., this portal does not allow the user to indicate the author or the material of the exhibit as the object sought, whereas portals developed within the frameworks of the methodology considered allow for a search on the basis of arbitrary ontological concepts and navigation using all associative relations.

## **7. Problems and prospects**

Application of the methodology to the development of knowledge portals has not only demonstrated its productivity but has revealed a set of problems.

Firstly, this is lack of facilities for advanced visual analysis of portal ontology and content for experts and users. Closely related to this problem is the problem of visualization of the hierarchy of concepts and objects built on different classification bases.

Though knowledge portals built by the instrumentality of the methodology provide users with facilities for search request formulation in terms of subject domain ontology, an ordinary user finds this kind of communication with the knowledge portal difficult. In this connection, the task of providing the possibility for formulating a search request in a limited natural language arises.

Another important problem is the absence of feedback with users of the portal as well as lack of possibility allowing the users having sufficient skills in the portal subject domain to extend the portal content with new knowledge and information resources.

Though the methodology provides developers of knowledge portals with an ergonomic data editor, filling the portal content is a laborious process. Therefore, development of facilities for an automatic filling of the content portal is a burning problem.

Besides, the problem of “off-site” ontologies integration and their use in the knowledge portal has not been solved. In order to reuse some ontology developed for another application or by means of another methodology, it

should be rewritten using formalism provided by the suggested methodology.

Finally, the most important problem is lack of facilities for comprehensive support of collective development of ontologies. At present several experts can develop the same ontology simultaneously, but during this process the cooperation between them is not supported by any formal or software facilities. There is either no means for automatic and automatized reconciliation of single parts of ontology developed by different experts. To cope with this problem, visual support of collective development of ontologies is required. Each of the experts should be able to see, apart from the common part of an ontology and its fragment developed by him/her, the results of an automatic correctness check of his/her fragment of the ontology and an automatic comparison of this fragment with the ontology taken as a whole and with the fragments developed by other experts.

To deal with the problems discovered, the methodology will evolve in the following directions:

- Development of advanced facilities of graphic visualization; this will allow us to present, in the form of a graph, not only a hierarchy of ontology concepts but also the portal content;
- Inclusion in the portal of additional facilities for analysis of the portal content and inference of new knowledge;
- Development of an ontology editor effectively supporting collective development of ontologies;
- Providing the possibility for users having sufficient skill in the portal subject domain of extending the portal content with new knowledge and information resources;
- In addition to information search using the ontology concepts, ontology-based search methods using queries presented in a limited natural language are going to be developed.

## **8. Conclusions**

The paper presents and discusses the methodology for the development of knowledge portals providing content-based access to scientific information resources on the basis of ontology. This methodology is supported by software environment for the creation of knowledge portals that contain both ontology and data editors and program modules providing search and navigation through the portal information space.

Ontology provides facilities for an effective representation of diverse information about a given area of knowledge, supports systematization and integration of relevant information resources and semantic access to them.

Ontology is used for an automatic generation of the following information structures of knowledge portals: a scheme of the internal data warehouse of the portal, forms for filling the internal data base, and forms for formulating a query in terms of ontological concepts and relations. Also, ontology supports knowledge-driven navigation through portal content.

The structuring of portal ontology into domain-independent and subject domain ontologies makes the knowledge portal easily adjustable to any area of knowledge.

Successful applications of the methodology suggested to the development of knowledge Internet portals on archaeology, computational linguistics, strength of materials and Antarctic data has demonstrates its soundness and productivity.

## References

- [1] Guarino N. Formal Ontology in Information Systems // Formal Ontology in Information Systems / N. Guarino (ed.). – Proc. of FOIS'98, Trento, Italy. – IOS Press, 1998. – P.3–15.
- [2] Berners-Lee T., Hendler J., Lassila O. The Semantic Web // Scientific American. – 2001. – Vol. 284, No.5. – P.34–43.
- [3] Brickley D., Guha R. RDF Vocabulary Description Language 1.0: RDF Schema W3C Recommendation 10 February 2004. – Available at <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>
- [4] OWL Web Ontology Language Guide, 2009. – Available at <http://www.w3.org/TR/owl-guide/>
- [5] Zagorulko Yu., Borovikova O., Bulgakov S., Sidorova E. Ontology-based approach to development of adjustable knowledge internet portal for support of research activity // Joint Bull. of NCC&IIS. Ser.: Comput. Sci. — 2005. — Is. 23. — P. 45–56.
- [6] Zarogulko Yu., Borovikova O. An approach to constructing knowledge portals // Optoelectronics, Instrumentation and Data Processing, 2008. –Vol. 44, – No.1. P. 75–82.
- [7] Zagorulko Yu., Borovikova O., 2007. Methodology of Building and Using Ontology for Providing Content-Based Access to Scientific Information Resources. // New Trends in Software Methodologies, Tools, and Techniques. Proc. of the six SoMeT\_07, Rome, Italy. / Hamido Fujita, Domenico M. Pisanelli (Eds.). – Amsterdam: IOS Press, 2007. – P. 105–115.
- [8] Benjamins V.R. and Fensel D. Community is Knowledge! in (KA)2. // B.R. Gaines, M.A. Musen (eds.), Proceedings of the 11th Banff Knowledge Acquisition for Knowledge-based Systems workshop, KAW'98 (Banff, Canada, April 1998). – Calgary: SRDG Publications, Department of Computer Science, University of Calgary, 1998.

- 
- [9] Using Dublin Core, 2009.  
– Available at <http://dublincore.org/documents/usageguide/>
- [10] Protégé. Web site, 2009. – Available at <http://protege.stanford.edu/>
- [11] Andreeva O., Borovikova O., Bulgakov S., Kholushkin Yu., Sidorova E., Tcirkin B., Zagorulko Yu. Archeological knowledge portal: Content-based access to archeological knowledge and information resources // Proc. of 10th National conference on Artificial Intelligence (CAI-2006). – Moscow: Fizmatlit, 2006. –Vol.3. –P. 832-840 (In Russian).
- [12] Zagorulko Yury A. On Experience of Building Knowledge Portals on Humanities // First Russia and Pacific Conference on Computer Technology and Applications, 6-9 September, 2010 – Vladivostok, Russia. – P. 336–339.
- [13] Zagorulko Yu., Borovikova O., Zagorulko G. Knowledge Portal on Computational Linguistics: Content-Based Multilingual Access to Linguistic Information Resources // Selected topics in Applied Computer Science. Proc. of the 10th WSEAS International Conference on Applied Computer Science (ACS'10). Hamido Fujita, Jun Sasaki (Eds.). (Iwate Prefectural University, Japan, October 4-6, 2010). – WSEAS Press, 2010. – P. 255–262.
- [14] Globa L.S., Novogrudskaya R.L., Mamuzich R. L. Development of a model for the internet portal “strength of materials” // Materials and technology. – 2012. – Vol. 4 (46). – P.407–410.
- [15] Globa L.S., Novogrudskaya R.L. Ontology Based Approach to Antarctic Knowledge Portal Designing // Proceedings of 18th international multi-conference ACS'2012, 30 May – 1 June 2012, Poland.
- [16] Heflin J., Hendler J. Searching the Web with SHOE // Artificial Intelligence for Web Search. – Menlo Park: AAAI Press, 2000. – P. 35–40.
- [17] Lausen H., Stollberg M., Hernández R. L., Ding Y., Han S.-K., and Fensel D. Semantic Web Portals - state of the art survey. Technical Report TR-2004-04-03, DERI ([www.deri.org](http://www.deri.org)), 2004.
- [18] Corcho O., Gómez-Pérez A., López-Cima A., López-García V., MC. Suárez-Figueroa. ODESeW. Automatic Generation of Knowledge Portals for Intranets and Extranets. // Lecture Notes in Computer Science – Vol 2870. The Semantic Web - ISWC 2003. Springer-Verlag, October 2003. – P. 802–817.
- [19] Spyns P., Oberle D., Volz R., Zheng J., Jarrar M., Sure Y., Studer R., Meersman R.. OntoWeb a Semantic Web Community Portal. // Proc. Fourth International Conference on Practical Aspects of Knowledge Management (PAKM), December 2002, Vienna, Austria. – LNAI. – Vol. 2569. – P. 189–200.
- [20] Hyvönen E., Salminen M., Kettula S., Junnila M. A Content Creation Process for the Semantic Web // Proc. of OntoLex2004: Ontologies and Lexical Resources in Distributed Environments, May 29, 2004, Lisbon, Portugal.