doi:10.2498/cit.1002034

Ontology-based Competency Analyses in New Research Domains

Julia Rogushina[1] and Anatoly Gladun[2]

[1] Department of Intelligent systems, Institute of Software Systems, National Academy of Sciences, Kyiv, Ukraine
[2] Department of Intelligence Networks and Systems, International Research and Training Center of Information Technologies and Systems, National Academy of Sciences, Kyiv, Ukraine

Ontology-driven methods of competence management oriented on support of scientific research for new domains are proposed. Ontologies of research domain are matched with personal information about scientific researchers represented into Web (for example, at the social networks) and results of their work (publications, monographs, reports etc.) are processed by logical methods and ontological analysis. Web-services and multi-agent programming paradigm are used for their software realization.

Keywords: Semantic Web, knowledge management, ontology, expert competence

1. Problems of Competence Management

Competence-based management is a relatively new way of organizing human resources with high performance. The theory of competence-based management defines competence as the ability to sustain the coordinated deployment of resources in ways that help an organization achieve its goals [1]. In this work we consider the competence management as a part of knowledge management [2].

The competence concept has to reflect such aspects of human work in organization as the dynamic external environment of an organization, internal processes and interactions with other organizations, and, the most important for this work, cognitive processes in an organization and coordination of different human, informational and knowledge resources [3,4].

Now one of the important directions of development of intelligent informational systems is associated with the support of management decisions to ensure the persons who make decisions (PMD) by knowledge about subject domain and appropriate means of analysis. One of the directions of PMD work deals with personal selection for different tasks.

A common problem today is the situation when PMD have to solve the problem in some new domain (for example, make review of papers). For this solution they need in domain experts – specialists competent in problem domain. In general, an *expert* is a person who has special knowledge about problems that are directly associated with specific subject domain [5].

In new domains, commonly recognized authorities are usually absent. That's why PMD have to evaluate the competencies of potential experts relative to this subject domain [6] by analysis of accessible information about them – for example, from social nets, published articles.

The urgency of these problems is increasing in modern society by expansion of complex information processed, continuous emergence of new technologies, research directions, complex information-intensive goods, facilities and services.

1.1. Definition of Competence

Competence is the level of achievement (experience, knowledge, skills) of person in some domain. Competence can be defined based on the analysis of the specialist activities, level and breadth of awareness of the achievements of science and technology, understanding of the studied problems and possible ways of their solving.

To quantify the level of competence the competence coefficient is used. In general, *competence coefficient* R is a function F of K(A) – the characteristics of expert knowledge and experience (qualitative and quantitative) for some expert A and of S(B) – the description of expert problem for problem B:

$$R(A, B) = f\,(K\,(A), S\,(B)).$$

An expert A can have different evaluations and the problem can have different descriptions.

A is a person that can be an expert if he/she has competence coefficient R(A,B)>p for problem B, where p is a constant that specifies the minimum PMD requirements to expert.

The competence coefficient can reflect different personal data. For some scientific domains personal data of each expert can be represented by the triple $\langle s,r,h \rangle$, where $s \in S$, $r \in R$, $h \in H$. $S=\{s1,s2,s3\}$ – Higher education of expert (s1 – coincides with the profile priority, s2 – basic education in a related specialty, s3 – basic education in other specialties); $R= \{r1,r2,r3\}$ – scientific schooling (r1 – academician, r2 – professors, doctors, r3 – Ph.D., Senior Scientist, Associate Professor), $H=\{h1,h2,h3\}$ – experience with this priority (h1 – not less than ten years, h2 – at least five years, h3 – at least one year). The evaluations r and h are defined quite simply, but to evaluate s for new domains is more difficult (most of experts are evaluated by s2, and it is not possible to set a priority in their level of proficiency).

Also, if new activity has emerged quite recently, any specialist has great experience in this area, and high estimates obtained by the level of scientific schooling can be negative for new problem solving – specialist received knowledge when new research domain did not exist and therefore he/she probably are not versed in its specificity. On the other hand, over a long period of research activities specialists can fundamentally change the direction of their research, and high evaluation r indicates a high level of intelligence, persistence in scientific work and the ability to get interesting and useful results.

Therefore, besides the triple $\langle s,r,h \rangle$ we have to take into account information about current scientific and professional achievements of a specialist, his/her area of work and working capacity. For employees of many specialties such information appears in their scientific publications, statements of work, descriptions of developed products, prepared learning materials etc. The availability of these data (publishing over the Web, representation in the libraries) determines the degree of expert authority and his/her ability and willingness to credibility of his/her own knowledge. This is directly linked with the specialist's ability to be an expert.

1.2. Competence Management as a Special Case of Knowledge Management

Modern competence management is an independent field of research, but from the point of view of informational technologies it can be viewed as a special case of a knowledge management system (KMS) or as an information retrieval system (IRS), because, with their own specifics, these systems solve the problem of representing and matchmaking of the knowledge with regard to two different types of objects – tasks and people who can execute these tasks.

The task of competence management is a perspective area of implementation of Semantic Web technologies. The reasons for this are, on the one hand, the urgency and complexity of the problem (and the lack of automated solutions for common cases), and on the other hand – the need for use of distributed heterogeneous Web knowledge.

1.3. Competence Management in the Field of Scientific Research

In this work we consider a special case of competence management in the field of scientific research work. This subproblem appears to be the most complex and interesting for two reasons, the first of which being a weak formalization of competency profiles of researchers (almost every individual scientific topic requires the development of its own profile) and the great influence of the research domain specific that expect the need of automated processing of external knowledge bases contained into the Web.

The second reason is associated with the fact that the research activity and its results usually are well formalized and open for analysis: scientific publications, patents, reports and other materials represented as natural language texts with elements of structured data (tables and graphs) and multimedia that usually are accessible in electronic form that allows their automated analysis by corresponding software.

For competence management in the scientific research planning, some criteria – such as education, skills and experience – can be identified already at the beginning of the selection. Competencies emerge later, in the process of selection. But competencies can make an important contribution at any stage of selection. However, such studies are quite effective only for professions well formalized and described in detail, and, therefore, are not suitable for the selection and certification of personnel for scientific research.

This fact is caused, firstly, by the use of knowledge about the subject domain for evaluation of the work of researchers. This domain is constantly changed and supplemented. Moreover, in general, it may be controversial from the point of view of different theoretical concepts. Secondly, it is necessary to evaluate objectively the results of their work that are represented mainly in the form of natural language documents (articles, papers, reports, etc.).

In addition, evaluation of the overall scientific capacity of a person is often important for the decision about his/her participation in solving specific problems or in cooperation with particular team. The problem of semantic interpreting scientific papers is solved by the use of automation of semantic markup of NL-texts, creation and processing of their meta-description etc.

Though in this work we try to use the methods of knowledge management based on Semantic Web technologies for task of competence management in scientific subject domain.

2. Knowledge Management in Web Applications

Now, a lot of Web applications are intelligent, and have to use knowledge of some subject domain or produce some new knowledge. In such applications, knowledge is represented in interoperable form and can be reusable. Ontological approach for knowledge representation is widely used because ontologies have a fundamental theoretical foundation of their formal semantics – descriptive logic.

Ontologies typically provide some general vocabularies that describe different domains of user interest or specialization of informational resource and define the meanings of terms used in the vocabulary. The ontology representation contains data and conceptual models, for example, sets of terms, classifications or theories [7,8].

One way of domain modeling is creating its thesaurus. Often, the terms “ontology” and “thesaurus” are used as synonyms, but in IT thesaurus is often used to describe vocabulary in the projection of the semantics, and ontology – to model the semantics and pragmatics in the projection of language representation [9]. For this task, thesaurus can be seen as a special case of ontology.

2.1. Problems of Knowledge Management for Web

Main problems of knowledge management (Figure 1) for Web deal with [10]:

- Integration of knowledge from different informational resources (e.g. integration of ontologies built on the basis of different texts from one subject domain);

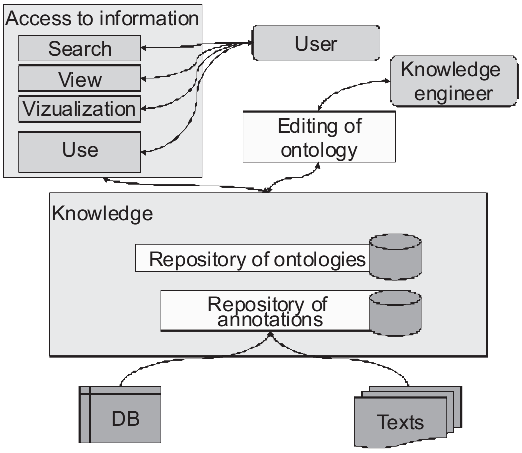

Figure 1. Main elements of ontological knowledge management.

- Search of inconsistency of knowledge acquired from the content of different informational resources and rating their adequacy and security;
- Knowledge acquisition from accessible information and its representation in the form understandable to user;
- Search of knowledge a user needs for the solution of some specific tasks;
- Automation of metadata creation and improvement that correctly describes the content of informational resources (textual or multimedia) on semantic level, and efficient search of such metadata.

A lot of other similar examples exist, but all of them come to the following ones [11]:

1. *Selection of the means for knowledge representation* (sufficiently powerful to satisfy the different requirements of users, but available for rapid processing and understandable to human): Now for these goals ontologies are widely used, but the problem deals with selection of ontology representation language version (OWL 1.0 versus OWL.2.0, OWL Lite, OWL DL, OWL Full, RDF, RDF Schema etc.) [12]. Domain ontology is a certain part of knowledge which describes important concepts and relations that can be used for the solution of problems at this domain;
2. *Methods of acquisition of new knowledge based on some informational resources* (for example, creation of metadescriptions of informational resources, generation of new rules by inductive inference or of new facts by traductive inference): new knowledge can be acquired from implicit, uncertain, contradictory textual representations, but large capacity of such information necessitates some automated methods of their processing. Availability of RDF – language for metadata representation – is a necessary, but not sufficient condition for it. For example, automated creation of metadata that describes the natural language document on semantic level requires to use: 1) methods of linguistic analysis; 2) knowledge of subject domain (e.g. domain ontology); 3) application-dependent methods of inductive, deductive or traductive inference oriented on processing of specific structures of knowledge (e.g. RDF triplets);
3. *Methods of matching of different informational objects on semantic level* (e.g. integration of two ontologies or detection of differences between them, matching of informational query and informational resource relevant to this query, discovery of subject domain of informational resource by analysis of its content): these problems are not trivial and don't reduce to traditional search because they have to analyze rules and knowledge of subject domain and their formal representations by special matching algorithms. The matching operations deal with the following challenges: large-scale evaluation, performance of ontology-matching techniques, discovering missing background knowledge, uncertainty in ontology matching, matcher selection and self-configuration, user involvement into the process of matching, explanation to user of matching results, collaborative ontology matching, alignment management and reasoning with alignments [2];
4. *Quality rating of new knowledge* (veracity, consistency, actuality, completeness). It needs to develop the different models of knowledge representation, to use the appropriate mathematical apparatus (e.g. first-order sentence theory, descriptive logics) and to evaluate the quality of ontologies related to real world and informal knowledge about real world.

2.2. Knowledge Management and Semantic Web

At the present stage of IT, in the majority of cases Web applications use standards and technologies of knowledge management developed by Semantic Web project. Knowledge management in Semantic Web environment needs in creation of adequate tools for retrieval, acquisition, store and use of knowledge subject to such properties of up-to-date Web as dynamics, heterogeneity, very large capacity and orientation on semantics.

The main component of Semantic Web concept is use of ontologies that allows to formalize knowledge about a subject domain. Semantic Web proposes – the DL-based language OWL for ontology representation [13]. Different instrumental tools provide the following possibilities: creation of ontologies and their linking with different informational resources, checking of ontology's consistency, refinement of ontology and executed of inference operations on ontologies but some important problems of KM are not supported by standard software tools. Figure 2 shows what elements of KM are now supported by Semantic Web technologies.

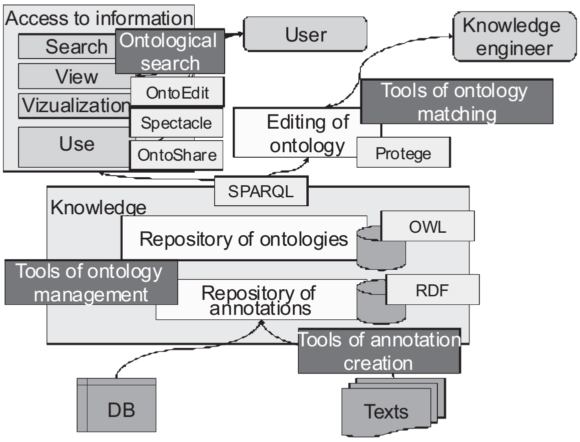

Figure 2. Semantic Web in knowledge management.

2.3. Use of Ontologies for Web Knowledge Representation

Analysis of publications shows that ontologies are an adequate and effective means for knowledge modeling about different subject domains, informational resources and other objects. Different authors represent various formal models of ontology, but all these models include [8,10]:

— the set of concepts that can be subdivided into the set of classes and the set of individuals;

— the set of relations between concepts where some subclasses of relations ("class-subclass", hierarchical, synonymy etc.) and functions (as special relation where the n-th element of relation is uniquely defined by other $n-1$ elements) can be separated;

— axioms and interpretation functions of concepts and relations.

Formal model of ontology is a triple O= $\langle$X,R,F$\rangle$, where X is a set of concepts, R – a set of relations between concepts from X and F – interpretation functions for concepts from X and relations from R. This is a general model, and in practice more precise models are used. For example, in [4] ontology is defined as a structure that includes identifiers of concepts, identifiers of relations, identifiers of attributes, data types and hierarchies of concepts and relations. Ontology can be defined as a tuple that, in addition to sets of classes, individuals, relations and data types, contains a set of values and some special relations (specialization, exception, creation of individual and assignment).

Existing technologies of the Semantic Web propose various means of ontology representation that differ one from another by their expressiveness and their complexity: RDF Schema is the simplest representation and OWL Full is the most powerful. Decision of ontology representation depends of problem specifics.

Languages for ontology representation can be viewed as syntactic variants of Description Logic (DL). The fundamental modeling concept of a DL is the *axiom* – a logical statement relating roles and/or concepts. There are many different Description Logics that differ in sets of properties and performance capabilities.

DLs have an informal naming convention, roughly describing the operators allowed: F – Functional properties; E – Full existential qualification (Existential restrictions that have fillers other than `owl:Thing`); U – Concept union; C – Complex concept negation; S – An abbreviation for ALC with transitive roles; H – Role hierarchy (subproperties – rdfs:subPropertyOf); R – Limited complex role inclusion axioms; reflexivity and irreflexivity; role disjointness; O – Nominals. (Enumerated classes of object value restrictions – `owl:oneOf`, `owl:hasValue`); I – Inverse properties; N – Cardinality restrictions (`owl:cardinality`, `owl:maxCardinality`); Q – Qualified cardinality restrictions (available in OWL 2.0, cardinality restrictions that have fillers other than `owl:Thing`); (D) – Use of datatype properties, data values or data types. The prototypical DL Attributive Concept Language with Complements (ALS) is a simply AL with complement of any concept allowed, not just atomic concepts. The description logic SHIQ is the logic ALC plus extended cardinality restrictions, and transitive and inverse roles. The naming conventions aren't purely systematic so that the logic ALCNIO might be referred to as ALCNIO and abbreviations are made where possible ALS used instead of the equivalent ALUE. The design of OWL is based on the family of DL. The Protégé ontology editor supports SHOIN(D). OWL 2.0 provides the expressiveness of SHOIQ(D), OWL-DL is based on SHOIN (D), and for OWL-Lite it is SHIF(D).

2.4. Semantic Search as a Part of Web Knowledge Management

We think that one of the most important tasks in knowledge management for Web deals with semantic search of information – in a lot of intelligent Web applications informational retrieval is a part of a system or is called as an external service. Retrieval of information about experts into the Web – people with some specific competencies – is a particular case of semantic search. Therefore, we analyze below the general approaches to this problem. The most promising of them deal with the Semantic Web project.

Semantic search is a superstructure on traditional retrieval procedure where for more efficient satisfaction of user's informational needs.

In semantic search of knowledge about a user, his/her personal informational needs and interests; about informational resources accessible for retrieval mechanism is processed for the purpose of increasing of search pertinence [14] – user receives the information that is really necessary for some task.

The result of semantic search can be not only a concrete Web document or fragment of such document, but some more complex informational object:

1. interesting to user information acquired from an accessible informational resource (textual or multimedia) where this information is contained implicitly;
2. a list of informational resources with some semantic annotations dealing with the user's query and user's personal preferences;
3. integration of the knowledge contained in different informational resources;
4. informational object specific for subject domain class (corresponding to some concept of domain ontology) – for example, organization, geographical object, human or scientific article;
5. composition of classified informational objects (e.g. human with some characteristics that work in organization of specific type and live in some concrete city).

Based on the analysis of current state of work in the sphere of informational content representation and methods of programming for Semantic Web, we can mark out some main problems that we have to solve in the process of designing of intelligent Web application realized in the semantic search procedure (i.e. the questions that have not now some universal standardized methods of solving and for which open software products are not realized):

— automated creation of metadescriptions of informational resources that reflect not only formal characteristics of documents, but their semantics that deals with some subject domain;

— generation of semantic markup of natural language documents by ontological concepts;

— automated creation and enhancement of ontology (at initial stage and for existing ontologies) on the basis of informational resources processing and by use of expert knowledge, particularly:

- formation of thesaurus of natural language informational resource;
- formation of initial ontology of subject domain by the set of natural language documents selected by user;
- enhancement of ontology of subject domain by the set of natural language documents;
- acquisition of ontological information from metadescriptions of informational resources;
- use of inductive inference for discovery of relations between the ontological concepts;

— operations on ontologies (the most necessary operations are consistency valuation of ontology, matching of pair of ontologies and integration of terminological base of different ontologies);

— semantic search that takes into consideration ontological knowledge about subject domain, the user and the task that he/she tries to solve.

It is not easy for a user to formalize the query for semantic search that reflects his/her informational need (as a user we consider either human or agent – software entity with some goals and intentions) because this formalization has to reflect:

1. Description of a problem that needs some information for its solving;
2. What information does a user have before this query?
3. What level of complexity and form of knowledge representation can a user understand?
4. How to acquire the necessary knowledge from accessible documents?

Semantic sears show some important differences from the traditional one realized in usual information retrieval systems (IRS) operating at Web environment:

	Traditional IRS	Semantic IRS
Query	The set of keywords	Informational need deals with some subject domain and problem
Information for search personification	History of user queries	Models of user and his informational needs
Search results	Document with keywords	Knowledge from relevant documents deal with user's interesting object
Information about IR	Index DB of IRS	Index DB of IRS and their metadata
Description of user domain	–	Domain ontology

2.5. Linguistic Methods in Creation of Ontologies of Natural Language Informational Resources

We use the algorithm of semantic markup of natural language texts that is described in detail in [15]. This algorithm factors into morphological and syntactical properties of natural language and knowledge about subject domain. As a result of this algorithm we receive the text where some fragments are linked with concepts and relations of domain ontology. The other result of this step is a set of rules that provide links between ontological entities and word forms of natural language.

The inputs of first stage are: O_0 – initial ontology of subject domain containing the concepts and relations most obvious for the user; T_0 – the set of natural language texts that describe a domain interesting for the user (texts from glossaries, manuals, textbooks, Wikipedia articles, other well structured definitions of domain terminology).

O_0 and T_0 can be empty. If O_0 is empty we have no knowledge about domain and therefore the retrieval procedure reduces to usual – non-semantic – search. If T_0 is empty then the search procedure stops and the user has to propose another request.

On the second step the rules of markup are used for new texts. If in one sentence two or more fragments are marked up by ontological concepts, but no fragments are marked by ontological relations, then we can add (if necessary) new relation to domain ontology.

If one fragment in one sentence is marked up by ontological concepts and the other one – by ontological relations then we can add (if necessary) new concept to domain ontology. An algorithm proposes those fragments sentences to user for extracting of new concepts if it is appropriate for subject domain.

An algorithm discovers links of text fragments deal with interesting to user concepts with other fragments of text only by their linguistic properties. Though that fragments can not deal with concrete task of user and then user don't include them to domain ontology.

This algorithm can mark up not only classes but individuals as well. In natural language the equivalents of individuals are named entities (names, titles etc.).

The results of such semantic markup can be used for development and improvement of ontologies together with linguistic approach. If some text paragraph contains two fragments linked with ontological concepts and a fragment linked with ontological relation and if linguistic analysis of the sentence shows that in this sentence these fragments are associated, but the domain ontology does not contain such relation of these concepts then the ontology can be enriched by this relation.

Ontology is enriched by the new concept: if one fragment of the text paragraph is linked with some other concept and other fragment – with some ontological relation and linguistic analysis helps to search the fragment that is semantically associated with these fragments, then user can determine a new ontological concept associated with this fragment.

In the process of linguistic analysis we propose to create and develop a lexical ontology of domain that contains information about natural text fragments that are associated with concepts and relations of domain ontology. This ontology is created in process of semantic markup of domain texts and then enriched in dialog with the user during the analysis of other texts.

2.6. Use of Thesauri in Semantic Search

Formal model of thesaurus is Th= $\langle$T,R$\rangle$, where T is the finite set of terms, and R – the finite set of relations between these terms. The term is a word or a verbal complex, which correlates with the concept of some organized field of knowledge (science, technology) which becomes into the system relations with other terms and forms with them some closed highly informative system.

Thesaurus is a special case of ontology $\langle$X,R,$\emptyset\rangle$. In some situations we can match in semantic search process the domain thesaurus with thesauri of available informational resources. The use of thesauri instead of the ontologies reduces the problems of their generation and matching because we can create thesaurus of natural language document much more easily (by lexical analysis) then ontology.

The thesaurus of domain is created as a union of sets that represent thesauri of natural language documents selected by a user to describe the sphere of his/her interests. Then the user can refine this thesaurus according to IDEF5 methodology for development of ontological models (`www.idef.com/IDEF5.htm`).

If we have an ontology of some domain or informational resource, then we can reduce it into the thesaurus. In some situations, for retrieval procedure we can take into account only the set X (concepts) and then matching of ontologies can be reduced to comparison of these sets (there is no deep semantic analysis, but this procedure can help to reject informational resources without corresponding terms).

For modeling of ontological relations mereological apparatus can be used. Mereology, as a formal theory about parts, marks out seven types of "the part of" relation, for example, component-object, part-mass, material-object.

This classification helps in refining of ontologies if a user in the process of adding a new relation to ontology explicitly states the mereological type of this relation.

For analysis of the lot of IR an algorithm of thesauri building is proposed: term vocabulary is building by the general list of document words, and then words from user list are thrown away from that vocabulary. User list can contain stop-words for some subject domain or natural language. If IR has some metadata describing its semantics (for example, in RDF), then words for vocabulary can be acquired from this metadata.

Then this vocabulary is matched with user thesaurus. User thesaurus can be built by extraction of concept names from domain ontology in OWL, as a union of vocabularies of IRs selected by user, manually by user or by combination of these methods (Figure 3).

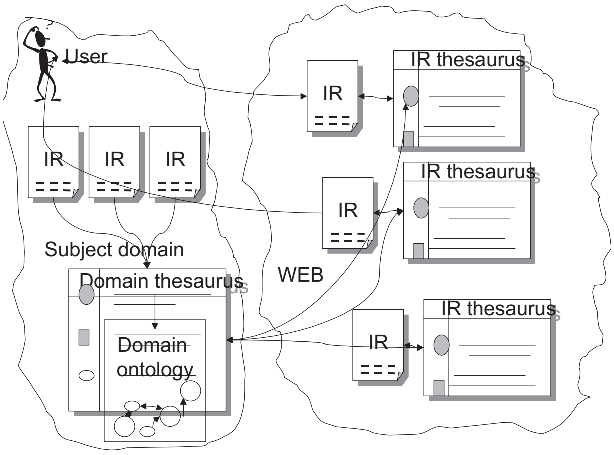

Figure 3. Informational retrieval on the basis of thesauri.

In forming of instances in the ontology of academic competencies a significant part of the knowledge can be acquired from the analysis of various open publications of researchers, that is from information resources available through the Web and represented by natural language (NL). If NL-texts are considered as a source of knowledge for constructing ontology of corresponding domain, then it is advisable to use thesauri.

Usually, thesaurus is defined as a dictionary that contains lexical items with explicit semantic links between them. Thesaurus can be viewed as a special case of ontology. We can explore the thesauri of individual specialists (e.g. experts) or thesauri of subject domains.

2.7. Semantic Search in IRS MAIPS

The results of research work described above were used in the realization of semantic search system MAIPS [16]. This IRS is oriented on users having stable informational interests in the Web and need regular acquisition of corresponding information.

In this system, ontologies and thesauri are used for formalized definition of the subject domain interesting for a user, and inductive inference methods provide acquisition of additional information about users by analysis of their permanent query history (e.g., preferences in informational sources, language and size of the text).

In addition, the search is personified with the help of individual indexes of natural language text readability that provides the most understandable and valuable information to the user.

MAIPS integrates ontological representation of knowledge, multiagent paradigm and Semantic Web technologies for the purpose of semantic search. The main features of IRS are:

- use of OWL language for domain ontologies and thesauri interoperable representation;
- realization of set-theoretic operations on thesauri;
- automated thesauri generation by natural language documents;
- use of Web 2.0 technologies (tag clouds – for search thesauri visualization, social services – for user cooperation;
- original sequencing algorithms for searched informational resources (IRs) with account of ontological concepts;
- use of natural language texts readability criteria for informational retrieval with account of personalized user needs;
- original inductive inference methods for generalization of MAIPS operation experience;
- use of multiagent paradigm for modeling of intelligent IRS behavior on the basis of BDI architecture [17];
- use of intelligent Semantic Web services paradigm for interoperable description of MAIPS functions [18].

3. Use of Inductive Inference in Semantic Search

3.1. Algorithms of Inductive Inference

IID3M Algorithm. A significant drawback of the well-known algorithm of inductive generalization ID3 [19] consists in the fact that it builds a classification rule only for the two classes. An original algorithm IID3M generalizes ID3 to an arbitrary number of classes and takes into account the level of accessibility of attribute values. This algorithm also detects the situation attributes which carry the most information about the result and thus help in constructing of the smallest decision tree. At each step the algorithm searches an attribute Ai with maximum information:

$$\begin{aligned} C(A_m) &= \sum_i \sum_j \frac{C(A_m = a_{mi}, R = R_j)}{T(A_m)} \\ &= \max_s C(A_s) \\ &= \max_s \sum_i \sum_j \frac{C(A_s = a_{si}, R = R_j)}{T(A_m)} \qquad (1) \end{aligned}$$

where C(X,Y) is the amount of information $C(X,Y) = \sum_i \sum_j p(X = x, Y = y)^* \log p(X = x, Y = y)$, where p(X=x,Y=y) is a probability of combined occurrence of the events X=x and Y=y, and $T(A_m)$ is the cost of obtaining the value of A_m.

The time for classification of the object by classification rule built IID3M upon the average is not exceeding the classification of the object in any other classification rule built on the learning sample. This follows from (1).

MID3 Algorithm. Attribute selection criterion (1) usually gives a good result, but the decision tree branching at every step for all possible attribute values causes a number of problems: the specialized rules are built and the number of examples in the nodes is reduced. Separation of the attribute values into two subsets increases the computational complexity by the choice of these subsets.

In this regard, we propose an algorithm MID3 – pseudo-binary generalization of IID3M that avoids complex calculations, but allows to remedy these deficiencies. Instead of branching for each value of attribute chosen by (1), it can branch some individual attribute value and other values in the form of a common branch and at each node of the decision tree attribute is a conditionally binary and accepts only two values – "X" and "not X".

For the same attribute these Xs may be different at different nodes of the decision tree. The choice of attribute values that is allocated to the separate branch is done based on information entropy measure (2). We choose the value of an attribute that carries the most information about the result:

$$\begin{aligned} \mathrm{a_{ki}} : \sum_{\mathrm{p}} \mathrm{C}(\mathrm{A_k} = \mathrm{a_{kp}}, \mathrm{R} = \mathrm{r_j}) \\ = \max_{\mathrm{m}} \sum_{\mathrm{j}} \mathrm{C}(\mathrm{A_k} = \mathrm{a_{ki}}, \mathrm{R} = \mathrm{r_j}). \end{aligned} \tag{2}$$

3.2. Processing of Incomplete Data in Inductive Inference

IID3M and MID3 algorithms are designed for processing of complete data during the consultation. But often it is necessary to classify objects where a full investigation is impossible (because of the complexity, cost and other reasons).

Data are incomplete (Maybe-data) if their values are currently unknown, but although they can be determined later. Based on these data it is not always possible to unambiguously classify the object, but we can select a subset of classes that object can belong on various methods of completions of incomplete data. We propose a method for constructing such subsets – a method of yellow-green branches.

The most adequate way of formalizing and processing of incomplete data is proposed by Codd method "Null Values", according to which data is incomplete if the property value for this object is currently unknown, although the property is inherent to the object and can be determined later. This unknown value can be defined by special constant, and any occurrence of such value may be substituted with the concrete value from the set of acceptable ones. The work with unknown values requires a special three-valued logic with the epistemic truth values (T-yes, F-no, W-maybe) and the corresponding truth table for all logical operations. The application of this logic to incomplete data sorts them into two classes: True-data that values are always accessible, and Maybe-data that values can be not available.

The following technology for inductive generalization of incomplete data is proposed:

Step 1: all *n* attributes are sorted by two classes according to a priori knowledge about their incompleteness: *m* attributes whose values are always available in the process of consultation, $m \leq n$, and *k* attributes whose values during the consultation can be unknown, $n = k + m$; then from the training set matrix X' obtained from the matrix X by reordering the columns so that the first *m* column of X' is formed by True-data;

Step 2: matrix X' is divided into a set of matrixes – matrix A containing *m* columns and matrixes B [h] containing *k* columns. The matrix A consists of such rows that for any row of the matrix A a row of the matrix X' exists where substring of the matrix A is a substring containing the first *m* attributes, and there is no other row of A where the first *m* values coincide with the values of this row, and each of the matrixes B[i], 0 <i≤h, consists of such lines that for any row of the matrix B[i] there is a row of X' that is a substring of it and the first *m* values of it a substring of the *i*-th row of the matrix A.

Step 3: decision trees built by the inductive inference algorithms for each of the obtained matrixes where another meaning – "unknown" (the attribute value is missing, can not be obtained, not known precisely, and so on – the data type Maybe) – is added to the list of possible values for each attribute of B[i] matrix. This value during the consultation is interpreted in a special way and is not considered in the construction of decision tree because the situation with an attribute value "unknown" is possible only in the consultation process.

Such inductive methods can be used for Semantic Web knowledge management in two ways: 1) for ontological knowledge acquisition from natural language documents (where the rows of learning sample are the occurrences of ontological concepts in some text and the results are the correlation of the text with some domain); 2) for ontology enhancement by new relations and concepts. In MAIPS, inductive inference is used also for acquisition of personal preferences of users (by generalization of system experience) and for clusterization of users with similar informational needs.

3.3. Inductive Inference as a Means of Knowledge Acquisition in Semantic IRS

Inductive inference algorithms can be used for the automated acquisition of ontological knowledge from semantically tagged natural texts about subject domain and from DB with user information.

For example, these methods are realized in intelligent IRS MAIPS oriented on users with permanent informational needs. MAIPS allows to personify the informational retrieval by inductive generalization of search experience and by taking into account personal readability of informational resources.

In competence management, the domain can be characterized by domain-specific ontology and by organizational ontology. From the general set of terms in analyzed texts, inductive inference acquires relations and terms important for the domain.

There are three main types of organization models based on ontologies that have structured information: 1) organizational ontology, 2) ontology of the organization subject domain, 3) the ontology of user activities. Organizational ontology provides semantic information about the organization structure. This complex structure often uses hierarchical decomposition into separate modules.

Organizational ontology [20] is an ontology that reflects knowledge about the organizational and functional structure of a particular subject of economic activities, i.e. its main components and connections between them. It contains information about employees, the hierarchy of production relations between them; resources used by the enterprise in the production process; products that is a result of enterprise functioning; structural units of the enterprise and the relationships between them (Figure 4).

The information about domain where somebody can be considered an expert is contained in the titles, UDC (Universal Decimal Classification) or ISBN and content of his/her publications and scientific reports and into the specialty passport, in the name of the organization and it's department where he/she works, and the most important – into the titles and abstracts of projects that he/she execute. The skill level of a person can also be evaluated by general and scientific experience, the number of publications (total and in recent years), the presence of graduate students etc. Another important source of information is an information about employees in the team with which projects were carried out or articles were published.

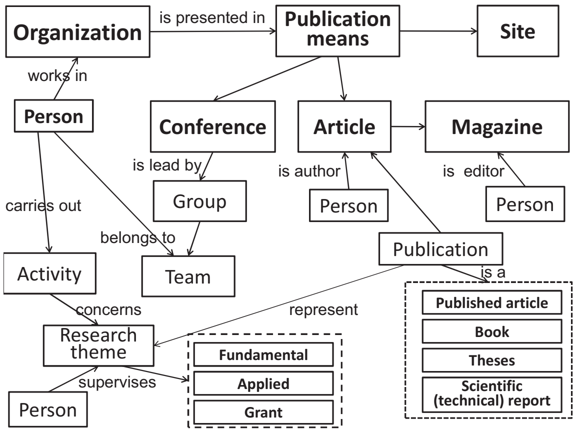

Figure 4. A fragment of structure of organizational ontology.

4. Use of Semantic Web Technologies in Competence Management

All proposed technologies can be used for task of competence identification of scientific researchers or learning courses as a part of research planning. This task is an example of a problem that needs an integrated use of different methods of Web knowledge management because knowledge about potential researchers and subject domain has to be attained from the available Web resources: structured descriptions of individuals (e.g., FOAF) and institutions (organizational ontologies) and their possibilities (for example, in form of Web services) with an account of their confidence level (with the help of Web 2.0 technologies and social networks) and from natural language and multimedia documents (and metadata describes their content) that fix the results of research work (articles, monographs, reports, presentations etc.).

Finally, methods of srmantic matchmaking have to be applied to founded information.

It is necessary to construct a theoretical model of competence management that includes:

1. DB of competence profiles (similar to IRS index and meta-descriptions of information resources) and set of services for their formation, comparison, storage and analysis;
2. Knowledge base that includes ontology of subject domain for solving problem, organizational ontology and the ontology of scientific activity;
3. Generalized model of the problem whose solution is formed by a set of actors, including the objectives to be achieved as a result of the work, the initial data, available resources and constraints on the process of the task realization (time, financial, legal, etc.) – an analog of the retrieval request;
4. Methods of analysis and semantic markup of NL-texts (texts containing the results of the scientific work of candidates), and a description of the goals and objectives of a research project – an analogue of the information resource indexing;
5. A set of methods and algorithms that can automatically form a set of performers (by formalized description of the scientific objectives and information about potential researchers) – search procedure. These methods should include a mechanism for the formation of rules of thumb and maintain a logical inference (similar to domain specialized IPS engine).

Future plans include the development of tools for optimal selection of team and predict the effectiveness of their collaboration.

Use of Semantic Web technologies provides interoperability of developed model and its software implementation, integration with other IR, KB and Semantic Web applications and the possibility of its extensions and modifications.

We intend to use the following Semantic Web technologies, Web 2.0 implementations and Open Source tools for solutions of such subproblems:

1. Intelligent Web-services, OWL-S, the information about people available through the Web (FOAF), as well as information found in the Web by semantic search engine MAIPS (author's certificates No32015 of 13.02.2010);
2. Ontologies in OWL format – available through the Web (e.g., organizational ontology of the National Academy of Sciences of Ukraine, various taxonomies and classifications) and designed specifically for a particular purpose;
3. Task ontology in OWL format and its extensions for the specific domains;
4. RDF and semantic markup means where relevant ontological terms are used as markup tags, as well as methods for assessing the IR significance similar to methods used in the IRS (for example, Google citation index);
5. Means of logical inference on ontologies such as Pellet, FaCT++, HermiT, Owlim;
6. Open Source software – for example, ontology editors and tools of visualization, means of semantic markup, linguistic processors, and tools for human resources management (e.g., OrangeHRM).

As a result of analysis of the generalized problem we can assign the following functions that are necessary in the systems of ontological knowledge management:

1. Methods of automated creation and updating of ontology: analysis of NL texts, processing of available meta-descriptions and semantic markup, on the basis of other ontologies;
2. Comparison of the two (or more) different ontologies: consecutive versions of one ontology; ontologies that describe one domain, but are created independently one of the other, ontologies which are an extension of one ontology for solving of different problems or by different developers;
3. Integration of the two (or more) different ontologies: describing one domain, but created independently; elaboration of one ontology for solving of different problems or by different developers, describing different domains with the intersected terminology (synergy);
4. Use of ontological knowledge: the confirmation or refutation of some fact about domain on the basis of the relevant ontology (task that in principle is not always solved because the systems of ontological knowledge use the paradigm of open world);

searching the relationship between certain objects which are described in the ontology and their properties.

4.1. Features of the Scientific Research Ontology

An important class in the ontology of scientific research is a summary of researcher. It contains information about publications, experience in research and teaching activities, educational level and history of research.

Standard of Resume RDF Schema can be used as a basis for describing the structure of a resume. However, the developers of this ontology have not paid adequate attention to the peculiarities of a summary of scientific researchers that, in addition to general information, includes publications, research experience, teaching, attending conferences, etc. Thus, they cannot be used widely for the ontology of scientific workers.

4.2. Algorithm of potential experts competency rating

Determination of competency assessment of specialists to some problem consists of the following steps (Figure 5):

- construction of thesauri of potential experts (by the organizational ontology) and the thesaurus of document under the examination (by the document content);
- normalizing of these thesauri using appropriate domain ontologies;
- comparison of the terminology (two sets of terms) of normalized thesauri.

By the content of the document under the examination the thesaurus is constructed. It is assumed that the document belongs to the domain corresponding to the area of the organization activities. Knowledge about this domain are represented in the form of domain ontology O_{domain}, $O_{domain} = \langle T,R,F \rangle$ where the set of concepts T contains *n* terms, $|T| = n$.

On the second step of the algorithm this thesaurus of document is normalized by its projection on domain ontology.
$ThN_n(d) = \{t_i : t_i \in T(O_{domain})\}$.

Suppose that the organizational ontology contains information about *s* employees. Thesauri of these employees are constructed by organizational ontology (this ontology contains knowledge about main scientific works of employers and references on their full texts or abstracts) [21]. This information for thesauri construction is acquired from the organizational ontology of research institute where they work and the administration of which may involve them as experts.

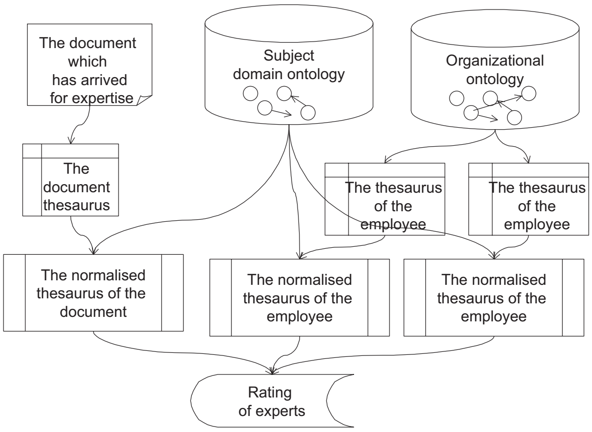
The document which has arrived for expertise
Subject domain ontology
Organizational ontology
The document thesaurus
The thesaurus of the employee
The thesaurus of the employee
The normalised thesaurus of the document
The normalised thesaurus of the employee
The normalised thesaurus of the employee
Rating of experts

Figure 5. Building of potential experts ratings.

It is very important that the employee thesaurus, in contrast to the user thesaurus (for example, in MAIPS), is not built manually by this person, but is generated automatically – and objectively – by proposed methods and based on the content of the Web and organizational ontology.

Some initial constraints can be imposed on the set of employees who can be used as experts (availability of scientific degree, the number of publications over the past year, age, experience of research, etc.). Thesaurus of employee $Th_i(p)$ is a set of words that is formed by the words from titles of his/her publications, passport of specialty, research projects which he/she executes, etc. The normalized thesaurus of employees is a projection of this thesaurus on set the terms of the domain ontology.

Each term t_j of the domain ontology O_{domain} obtains the set of values $m_{doc}(t_j)$ and weight of ontology terms $m_i(t_j)$, $i = \overline{1,s}$. For each employee, his rating $r_i = \sum_{j=1}^{n} m_{doc}(t_j)^* m_i(t_j)$, $i = \overline{1,s}$ for the examination of a particular document is constructed. Those employees who received the highest rating are potentially the most competent for examination of this paper (Figure 6).

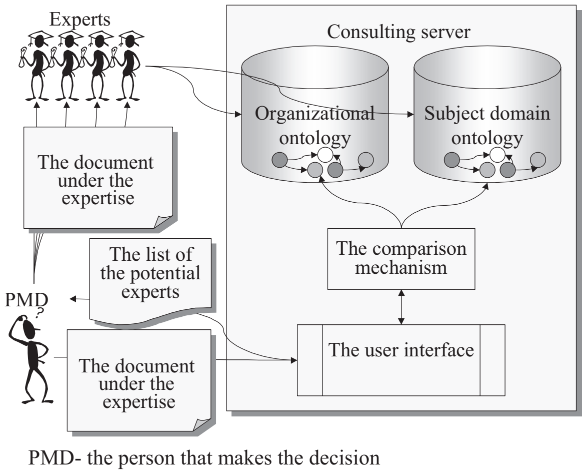
Experts
Consulting server
Organizational ontology
Subject domain ontology
The document under the expertise
The comparison mechanism
The list of the potential experts
PMD
The user interface
The document under the expertise
PMD- the person that makes the decision

Figure 6. The process of forming the initial set of experts by organizational ontology.

If the experts face the task not only to estimate an individual document (for example, write a review for an article in a scientific journal or for thesis project), but also make a comparative analysis and sort documents from some finite set of documents (for example, examination is carried out for the competition of research projects in some research domain), it is reasonable to use a union of thesauri of all documents under examination instead of individual document thesaurus.

5. Summary

This paper presents a new approach to the problem of competence management in the context of new IT technologies – in particular, the Semantic Web project and Web 2.0, as well as the results of research in psychology, competence management and so on.

The main idea of proposed approach is based on the use of ontologies for formalized and reusable representation of two main elements of competence management – knowledge about competencies of specialists and about the scientific problem that these specialists would solve. These two ontologies are aligned with the use of domain ontology and organization ontology.

We propose to build a task thesaurus and thesauri of potential experts by processing of expert task description and formal information about specialists (submitted online through the organizational ontology). Then we have to match them and determine based on this matchmaking the competence coefficient for each of these specialists – quantitative assessment of specialist's availability to solve this expert task.

This approach allows to form a group of experts for each task to improve the composition of which other methods can be used.

References

[1] G. Hamel, A. Heene, *Competence-based Competition*. John Wiley & Sons Ltd, 2000. – 358 p.

[2] G. Berio, M. Harzallah, Knowledge Management for Competence Management. *J. UKM 0(1)*. (2005), 21–28.

[3] F. Draganidis, P. Chamopoulou, G. Mentzas, A Semantic Web Architecture for Integrating Competence Management and Learning Paths. *Journal of Knowledge Management*, vol. 12, Iss: 6, pp. 121 – 136.

[4] E. Biesalski, A. Abecker, Human Resource Management with Ontologies. Presented in the *Proceedings of Professional Knowledge Management*, WM-2005, 3rd Biennial Conference, Kaiserslautern, Germany.

[5] R. Crowder, G. Hughes, W. Hall, Approaches to locating expertise using corporate knowledge. *Internet Journal of Intelligent Systems in Accounting, Finance & Management*, 11(4) (2002), pp. 185–200.

[6] I. Becerra-Fernandez, Searching for experts on the Web: A review of contemporary expertise locator systems. *ACM Transactions on Internet Technologies*, (6)4 (2006), pp. 333–355. ACM, New York, USA.

[7] P. Shvaiko, J. Euzenat, Ten challenges for ontology matching. Presented in the *Proceedings of the 7th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE)*, 2008.

[8] P. Cimiano *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications. Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006. – 347 p.

[9] A. Gladun, J. Rogushina, R. Martínez-Béjar, F. García-Sanchez, R. Valencia-García, Integration of Financial Domain Knowledge on Base of Semantic Web Technologies. *Information Models of Knowledge* (Edited by K. Markov, V. Velychko, O. Voloshin), (2010) pp. 106–112. ITHEA, Kiev-Sofia.

[10] J. EUZENAT, P. SHVAIKO, *Ontology matching*. Springer-Verlag, Berlin, Heidelberg, 2007. – 332 p.

[11] A. GLADUN, J. ROGUSHINA, Use of Semantic Web Technologies in Design of Informational Retrieval Systems. In Book *Building and Environment*, Nova Scientific Publishing, New-York, USA, 2009. pp. 89–103.

[12] I. HORROCKS, P. F. PATEL-SCHNEIDER, F. VAN HARMELEN, From SHIQ and RDF to OWL: The Making of a Web Ontology Language. *Journal of Web Semantics*, n. 1 (2003), pp. 7–26.

[13] OWL-S: SEMANTIC MARKUP FOR WEB SERVICES. `http://www.daml.org/services/owl-s/1.0/owl-s.html`

[14] A. GLADUN, J. ROGUSHINA, F. GARCIA-SANCHEZ, R. MARTINEZ-BEJAR, J. T. FERNANDEZ-BREIS, An application of intelligent techniques and Semantic Web technologies in e-learning environments. *Expert Systems with Applications*, *an International Journal*, v. 36 (2009), pp. 1922–1931.

[15] A. GLADUN, J. ROGUSHINA, Use of Semantic Web Technologies and Multilinguistic Thesauri for Knowledge-based Access to Biomedical Resources. *International Journal of Intelligent Systems and Applications*, v. 1 (2012), pp. 11–20.

[16] A. GLADUN, J. ROGUSHINA, V. SHTONDA, Ontological Approach to Domain Knowledge Representation for Informational Retrieval in Multiagent Systems. *International Jornal Information Theories and Applications*, v. 13, n. 4 (2006), pp. 354–362.

[17] T. BERNERS-LEE, J. HENDLER, O. LASSILA, The Semantic Web. *Scientific American*, (2001), pp. 34–43, May.

[18] SH. MCILRAITH, S. TRAN CAO, H. HONGLEI ZENG, Semantic Web Services. *IEEE Intelligent Systems*, Stanford University, vol. 9 (March/April 2001), pp. 46–54.

[19] J. R. QUINLAN, Discovery rules from large collections of examples: a case study. *Expert Systems in the Microelectronic Age*, Edinburg, 1979, – p. 87–102.

[20] AN ORGANIZATION ONTOLOGY, W3C Working Draft, 2012. `http://www.w3.org/TR/2012/WD-vocab-org-20120405/`

[21] A. GLADUN, J. ROGUSHINA, An Ontology-based Approach to Student Skills in Multiagent E-Learning Systems. *International Jornal Information Technologies & Knowledge*, v. 1, n. 3 (2007), pp. 219–225.

Received: December, 2011
Revised: July, 2012
Accepted: October, 2012

Contact addresses:

Julia Rogushina
Department of Intelligent systems
Institute of Software Systems
National Academy of Sciences
Kyiv
Ukraine

Anatoly Gladun
Department of Intelligence Networks and Systems
International Research and Training Center of Information Technologies and Systems
National Academy of Sciences
Kyiv
Ukraine

DR. JULIA ROGUSHINA was born in Kyiv in 1967. She received the B.Sc. from Kyiv Taras Shevchenko State University in 1989. She received her PhD degree in Computer Science from Glushkov's Institute of Cybernetics, Kyiv, in 1995. She is a senior researcher at the Institute of Software Systems, National Academy of Sciences of Ukraine.

Her research interests include the development and application of intelligent information systems; theory of software agents behavior, inductive knowledge acquisition, intelligent information retrieval, ontological analysis, Semantic Web technologies. She has published more than 140 publications in scientific journals and conferences. She is the coauthor of monograph "Agent Technologies" and several textbooks. Julia Rogushina has been involved in several national research projects, for example, "Research of intellectualization means for multiagent informaion retrieval systems".

She is an Associate Professor at the Department of Information Systems of Kyiv Slavistic University where she teaches the courses "Modern Internet Technologies", "Systems of Artificial Intelligence", "Data Mining".

DR. ANATOLY GLADUN was born in Rivne, Ukraine in 1961. He received the B.Sc. and M.Sc. degrees from Technical University in Lviv, Ukraine in 1984. He holds a PhD obtained from the Department of Computer Sciences at the Electrotechnical University (Saint-Petersburg, Russia). He is Head of Department of Intelligent Systems at the International Research and Training Centre of Information Technologies and Systems (National Academy of Sciences). He has been involved in several national and internal research projects (FP5), for example, INCO-Copernicus Project 960114 – EXPERNET "A Distributed Expert System for the Management of the National Network", Grant NATO NIG 971779 – "National Telecommunication Networks for Scientific and Educational Institutions" – URAN, ATM-Sat Project "ATM-based Multimedia Communication" (in GMD FOKUS, Berlin, 2000-2001) and "Research of intellectualization means for multi-agent information retrieval systems". He is the author of more than a 150 publications in conferences, journals and books. His research interests include the development and application of knowledge technologies to different fields such as e-Medicine, e-Commerce, e-Learning, retrieval systems, Semantic Web, network management, intelligent software agents (models, architectures, methodologies of development) and their application; Semantic Web services, ontologies, wireless networks. He is an Associate Professor at the Department of Computer Science (half-time) at the University "Kiev-Mogyla Academy", (`www.ukma.kiev.ua`) and at European University in Kiev, (`http://e-u.in.ua/eng/`).