Available online at *http://www.mecs-press.net/ijmsc*

Use of the Ontological Model for Personification of the Semantic Search

J. Rogushina

Senior researcher, Institute of software systems of National Acedemy of Sciences of Ukraine, Glushkov, 44, Kiev, Ukraine

Abstract

Semantic search is an important component of modern intelligent applications oriented on work in open information environment. The intelligence level of application depends of it's capabilities in knowledge processing and defines it's facilities. Now applications widely use ontologies for knowledge representation. Therefore criteria of intelligence level estimation (that can analyze ontologies) of applications and retrieval systems as their particular case are proposed.

In this paper, an ontological model of the intelligent interaction of the main objects and subjects of the semantic search (the Web information resources, information objects and information consumers etc.) is developed. Software realization of semantic search on base of this ontological model and integration of this search instrument with applied systems are describes.

Index Terms: Semantic search, information object, ontological model, thesaurus, collaborative search.

1. Introduction

Information retrieval in the most general understanding is a complex problem which includes two sub-problems – matching of user beliefs about required knowledge with content of available information resources (IR) and building of the information objects (IO) by content of relevant IR with a finite set of properties that are interesting to user. A lot of modern intelligent applications oriented on work in open information environment include the search component. Information retrieved by this component is a base of these applications work, and it`s pertinence. completeness and actuality influence on the quality of applied results. Therefore this article analyze such corresponding questions: – the complex task of the semantic search; – criteria of intelligence level

* Corresponding author – Rogushina J.
E-mail address: ladamandraka2010@gmail.com

estimation of applications; use of these criteria for retrieval systems; – ontological model of semantic search that explicitly defines the roles and relations of the main objects and subjects of retrieval process; – use of this model in different intelligent applications and perspectives of further research.

2. Semantic Search

Special cases of this problem are pattern recognition, speech and image recognition, semantic analysis and translation of natural language texts, service discovery and composition.

The problem of semantic search is extremely difficult in practice and therefore it is advisable to consider it's simplified subproblems with some limitations on the user needs and generated information objects. User has some part of information about IO and tries to supplement it by information extracted from various IR.

On the basis of analysis of research of the existing information retrieval systems (IRS) we can identify the main trends in improvement of information retrieval systems:

- from formal – to semantic;
- from unified – to personified;
- from individual – to collaborative;
- from closed – to controlled.

Semantic search is an information search where the matching of request with IR and IO generation are performed on the semantic level, i.e. with the help of knowledge. Semantic search is based on achievements in the field of artificial intelligence, in particular – on the general theory of knowledge representation and processing, pattern recognition and inference, methods of mathematical statistics and social psychology.

The main difference of semantic search from the traditional one is the use of the knowledge about the searched object, users, IR and subject domain of search, as well as the ability to detect not the data but knowledge.The subject of study of this work is the process of semantic search in the distributed Web environment, it's objects, entities and relations between them, as well as methods of comparison of information resources with the information needs of users.

Now knowledge discovery of the Web resources is an integral part of many intelligent applications that are designed to work in an open information space and need permanent updating of information. This is caused by the fact that most of modern applications are focused on the use and processing of knowledge about the interesting for user subject domain. Such domain knowledge aren`t laid into these applications during their development and are retrieved dynamically from the available IRs. As a rule, access to domain IRs is provided by the Web where a variety of formalization models, representation forms, store formats, access conditions and processing methods can be used.

The information that is represented structurally (ontologies, thesauri, metadescriptions, semantically marked up files, etc.) can be process much easier and effectively but, unfortunately, only a small part of the Web IRs has such representation. Therefore, the search among such IRs often doesn't allow to find the desired information.

A significant part of the knowledge accumulated as a result of development of human society that describes the different subject domains is contained in the documents in the form of natural language text. Also the Web contains a lot of multimedia documents and with the growing of various connected to the Internet devices, their volume increases much faster but multimedia documents contain much less useful information (for example, the vast majority of the various photos and videos usually are interesting only to those who filmed them.

An *information object* (IO) is a model of a real or virtual domain object (object, being, event, process, etc.) in the information space that defines the structure, attributes, constraints and, possibly, the behavior of this object [1]. For example, person, publication, Web-site, organization, city are objects, and their descriptions are information objects. By the Semantic Web conception such immaterial entities of the Web information space as ontologies, software agents, Web-services, information resources, metadata, databases, etc. are IO too.

A lot of IO have some predetermined complex structure. Examples of such IO are organizations, educational institutions, people, Web-services, business processes. If we have some domain ontology then the appropriate classes of this ontology can be used as a resource of knowledge about domain IO structure, and then the values of these IO properties are acquired from the relevant IRs. IO taxonomy can be used by the SSS user to specify more formally what kind of IO he tries to find and what information about it (and other IO connected with the searches one) he has. For example, in process of search of IO classified as person we can specify some search conditions where "Rogushina" is a name of this person, "Kiev" is a place of person's residence, and the "Kelly Lada de Mandraka" is the name of this person's dog. The structure of the IO provides semantic markup of the request.

This approach raises the problem of finding an ontology that represents the IO structure and is relevant to search domain. It provides an intelligent analysis of the structure and content of IR. In general, the system of semantic search (SSS) can be considered as a building-up of the traditional IRS that implements an intellectualization of these IRS.

3. Problem Definition

Now the efficient development of the information retrieval deals with the intellectualization of traditional search engines, with personification of the retrieval process that provides user the tools of control, and with use of the external knowledge bases accessible by the Web. Formal ontology-based model that adequately describes the interaction between core subjects and objects of semantic search (users and their information needs and personal preferences, on the one hand, and information resources that contain information of interest to users of information about the objects, on the other hand) is necessary for development of intelligent SSS that provides has to use the knowledge derived from both internal and external knowledge bases for matching of required and available information objects.

4. Related Work

Intelligent system is a system for solving the creative problems that traditionally require some human intelligent efforts and the use of knowledge from the corresponding domain. They are discovered by the artificial intelligence (AI) that analyses the methods, models and techniques of computer-assisted human mental activity [2, 3]

The important properties of intelligent information systems (IIS) are adaptation, self-organization and self-improvement [4]. Intelligent of system is estimated – by analogy with human behavior – through the complexity of the processed domain objects, space-time and causal relationship between them and the algorithms that are used for their processing [5, 6]. Intelligence level of IIS is directly related to the degree of information structurization and methods of its transformation, the ability of the system targeting, planning and target implementation. The Web-oriented IIS have their own specifics based on the semantic processing of the distributed and heterogeneous information [7].

IIS is a complex system that can perceive, analyze, convert and save the IO models. The main characteristic feature of IIS is the ability to use knowledge to achieve results in their work. The main types of actions that are performed in solving of the intelligent problems are the conversion of relations between objects; creation of new objects; search of object that satisfies the specified requirements; calculation of parameter values of object; inference of object properties; search of the specified elements in the object model etc.

An importance of the estimation and comparison of the IIS intelligence level deals with the fact that now a lot of tasks that need in new knowledge that they find in process of their work. An example of this type of task is smantic search that try by the descriptions of some properties of IO structure elements to extract knowledge from the Web IR.

Improving the efficiency of modern IRS is directly related with their intelligence level. Researchers use such terms as "smart search", "intelligent search", "knowledge-oriented search", "semantic search" etc. However,

the term "intelligence" is considered only in the intuitive sense and usually is not defined formally. This leads to problems in estimation and comparison of the intelligence level of different IRS and complicates their analysis.

Intelligent information systems (IS) can operate autonomously and intellectualized ones – with the direct or indirect participation of the user. Intelligence level of IS depends of it`s ability for self-learning, ability to goal attainment and the domain breadth. But it is quite difficult to distinguish between these types of IS, so it is more correct to speak not about the intelligent and intellectualized IS but about the intelligence level of IS: any IS has more then zero level of intelligence (IS has some knowledge laid by the IS developer about domain and about the methods of information processing), and the upper limit of the IS intelligence level does not exist.

Intellectualization of IS is the process of it's intelligence level increasing. We propose the following definition [8]: suppose two IS – A and B – are capable to find solutions $R_A(d)$ and $R_B(d)$ respectively in situation d from particular data domain D. The system B is an *intellectualization* of A, $B = \text{Int}(A)$, if: 1) system B always proposes a solution in situations where system A gets it; 2) $\exists d` \in D, \exists R_B(d`), \neg \exists R_A(d`)$ – A can't find solution for situation d` but B can do it.

Now many publications deal with the quantitative evaluation and comparison of the intelligence level of various IS which differs by the set of evaluation parameters. For example, in [9] the intelligence level is determined by the presence of the following properties: s1 – autonomy; s2 – ability to interact with other objects; s3 – ability to take in information about the world; s4 – ability to apply abstractions; s5 – ability to use domain knowledge; s6 – adaptive behavior; s7 – ability to learn; s8 – tolerance for errors; s9 – the ability to communicate by natural language.

This approach allows to compare the IS intelligence level from different domains (in some cases) – if the system A has all the properties of the system B but system B has properties not found in system A, then I (B) – the intelligence level of B is above the intelligence level of A: $\exists S(A) \subseteq S, S(B) \subseteq S, S(A) \subseteq S(B) \Rightarrow I(B) \geq I(A)$.

However, in such an approach many IS are not comparable: $\exists S(A) \subseteq S, S(B) \subseteq S, S(A) \not\subset S(B), S(B) \not\subset S(A) \Rightarrow I(B)$ and $I(A)$ are not comparable. For example, it is difficult to compare the intelligence level of systems able to natural-language communication but not the ability to learn (such as various chat simulators) and systems of inductive acquisition of new knowledge that don`t support natural language dialogue with user. Furthermore, all of these parameters may be inherent in different IS in a greater or lesser degree, and quantification of some of them (e.g., the ability to incorporate domain knowledge) is extremely difficult.

In [10] intelligence level of IS is a function of six properties $I(A) = \sum_{i=1}^{6} s_i * w_i$ where w_i is the weight of the i-th property, and s_i is the value of the i-th properties: s1 – use of the world model in generation of own IS plans of action; s2 – generation of alternatives in planning of IS activities; s3 – ability to change the plan during its implementation if the action leads to undesirable for IS consequences; s4 – use of own IS experience to expand and correct the domain model; s5 – natural language communication with IS user; s6 – admissibility of the plan time for problem solving. Unfortunately, the user IP can independently estimate only the last two parameters, so he has to rely on the approvals of the IS developers which are not always objective and correct.

5. Criteria of Intelligent Level Comparison of Search Engines

On base on the above analysis of the publications we propose to use the criterion for evaluation of the IS intelligence level with such parameters: T – type of atomic elements; X – number of atomic elements; L – number of connections between the elements; C – number of commands in the normalized algorithm; P – proportion between the number of successful experiments and the total number of them; K – class of problems solves by IS. This criterion can be applied to various IRS either with semantic search or without it.

The complexity of the data structure that IS can handle is characterized by the type of atomic elements T, their number X and the number of links between L. Complexity of data transformation methods is characterized by the number of commands in the normalized algorithm C (normalized algorithm is any formalization of the algorithm concept such as Markov normal algorithms or the Turing machine).

The probability of the declared by IS goal attainment is estimated as the ratio of successful experiments with their total number P. This parameter user can estimate fairly accurately. For example, for IRS this parameter is determined by the search pertinence – what part of the proposed by the IRS query results the user considers as satisfactory answer to his request.

For this criterion the following condition is executed: any intellectualization of IS increases it`s intelligence level (or at least not decreases). If IS B is a result of intellectualization of IS A and $\forall A, B, B = \text{Int}(A) \Rightarrow I(A) \le I(B)$ then $f(T(A), X(A), L(A), C(A), P(A), K(A)) \lessapprox f(T(B), X(B), L(B), C(B), P(B), K(B))$.

This criterion applies parameters whose values can be evaluate quantitatively not only by IS developers but also by IS users – in terms of their individual strengths and goals.

IS parameters can be estimated independently by transforming intelligence criterion as follows:

$$I(A) = f(T, X, L, C, P, K) = g(T, X, L) * h(C) * P * \varphi(K)$$

We use $g(T, X, L) = g_1(T) * g_2(X, L)$ for simplification of evaluation where the functions g1 (T) and g2 (X, L) are designated by the table.

Thus, $I(A) = g_1(T) * g_2(X, L) * h(C) * P * \varphi(K)$ while h (C) and $\varphi(K)$ are also designated by the table. These estimates are given by IS users on base of their own experience and use of information technologies in different areas and require further refinement.

T takes into account what IO are used in IS as the inputs and outputs, how complex is their structure and how complex methods are used for their analysis and generation. For these purposes user can apply various IO taxonomies. If IS process IO of different types then estimation of it`s intelligence level has to considered evaluation of the most complex IO: $g_1(T) = g_1(IO_1, \dots, IO_n) = \max_{i=1}^{n} g_1(IO_i)$.

But this approach is not satisfactory for situations when we want to compare two IS one of which is the intelligent development of the other one. For example, if both IS can process some complex IO X (such as structured metadata about IR) but IS A uses in this processing only simple additional data Y (for example, user request) and IS B uses instead of Y more complex additional data Z (for example, domain taxonomy) and structure of Z is more complex then Y but simpler then X. In this case the intelligence estimations of A and B but we understand intuitively that IS B is more intelligent then A. Therefore there is advisable to use a criterion that takes into account the total score of the IO complexity:

$$g_1(T) = g_1(IO_1, \dots, IO_n) = \sum_{i=1}^{n} g_1(IO_i).$$

This criterion provides more useful results, but another problem arises: the intelligence estimation is affected equally by the input and output data. However it is clear that the generation of data with complex structure is more complicated than the processing of existing data with the same complex structure. For example, generation of ontology by the natural language texts is more difficult task than use of an existing ontology for semantic markup of the natural language text but the value of the intelligence criterion is the same for both IS.

Therefore, it is advisable to use two separate criteria for the input data and output data. The output of IS are those IO that are created or modified as a result of IS work and input of IS are all other data influenced on the formation of the output data.

$$g_1(T) = a_{input} * g_{input}(T) + b_{output} * g_{output}(T) =$$
$$= a_{input} * g_{input}(IO_1, \dots, IO_{n_{input}}) + b_{output} * g_{output}(IO_1, \dots, IO_{n_{output}}) =$$
$$= a_{input} * \sum_{i=1}^{n_{input}} g_1(IO_i) + b_{output} * \sum_{i=1}^{n_{output}} g_1(IO_i)$$

Where $\bigcup_{i=1}^{n_{input}} IO_i \cup \bigcup_{i=1}^{n_{output}} IO_i = T$ and the coefficients a_{input} and b_{output} evaluate an importance of the input and output data depending on the domain specifics (usually $a_{input} < b_{output}$).

Earlier smart IS commonly used relatively simple types of atomic elements g1(T) which correspond to data types in programming languages (such as integers and real numbers, characters and strings of characters, tables, file and directory names) or data types related with traditional models of knowledge representation (frames, semantic networks, graph structures, etc.). Now the modern applications oriented on the Web are characterized by a more complex structure of input and output IO that are associated with certain objects of real or virtual world (software agents, Web-services, semantically marked up resources, elements of Web of Things, ontological description, etc.).

For   example, intelligent metasearch systems A, B and C handle search results derived from common external IRS. First IRS A take into account user personal properties and the history of interaction with the user for individualized ordering of search results.

Second IRS B orders search results with the help of search domain ontology, and third IRS C proposes to user not only an ordered list of links to documents but performs semantic markup of discovered documents by domain ontology concepts. Input of A contains a request (string of symbols), a set of links to documents received from external IRS and a set of natural language documents (previous search results), and it`s output is a reordered set of links to documents. For IRS B input contains this query (string of symbols), a set of links to documents received from external IRS, and domain ontology, and the output contains the reordered set of links to documents.

For the third IRS C input data contains a query (string of symbols), a set of links to documents received from external IRS, domain ontology and lexical domain ontology and the output of C contains reordered set of links to the documents and these semantically marked up documents. Third IPS is able to use (input) and generate (output) more complex IO. This fact confirms the intuitive assumption that IRS A is the least intelligent, and the third IRS C is the most intelligent.

It should be observed that this approach deals only with formal evaluation of IS intelligence level and it's possibilities but actual results of IS work depend on relevance of models and algorithms of information processing to subject domain and on it's correct software realization.

For example, if SSS uses effective methods for finding and processing of domain ontologies but it`s requests are redirected to external IRS with a small index base and weak search engine then obtained results don`t satisfy the user despite high intelligence coefficient of SSS.

Evaluation of atomic elements complexity g2 (X, L) depends on their number X and the number of connections between them L: various functions can be used for different domains – for example, $g_2(X,L) = X * L$ or $g_2(X,L) = X^L$.

Unfortunately, the user can't accurately estimate these parameters, but he can give them an approximate qualitative assessment by the scale intervals for the parameters: "small", "medium", "big" etc. These ratings can be processed on base of fuzzy-set theory, i.e. multiplication of fuzzy values is defined as the product of their relative weights and their union – as their sum.

In assessing of the intelligence level in addition to the formal parameters user can take into account the complexity of problem solved by IS.

If IS B is in the intelligent superstructure over IS A then the intelligence level of B from the user viewpoint is defined as the sum of intelligence levels of systems A and B – user doesn't know what parts of work are executed by A and by B.

For example, if A is an information retrieval system that find the results relevant to user request, and B reorders these results in consideration of the personal features of the user, then the user has no way to assess separately the properties of these systems, and can count $I(A \cup B)$ by the above formulas .

The analysis of a variety of IS, in particular, SSS, shows that the assessment of the intelligence level of IS by the proposed criteria corresponds to the intuitively estimated intelligence level of considered IS.

Thus, we can assume that the intelligence level of IS is largely determined by three parameters: 1) the ability to use the knowledge with a complex structure; 2) the availability of advanced methods of such knowledge processing; 3) the ability to acquire new knowledge from the open information environment. Therefore, the development of applied SSS in direction of their intellectualization is connected with the development of sub-systems of knowledge management, learning tools and algorithms for the analysis of knowledge represented by complex ontological structure.

6. Ontological Model of Semantic Search

The analysis of publications shows that the most of applications related to semantic search process knowledge implicitly to user and this knowledge is not personified [11]. User hasn't possibilities for independent influence on the choice of knowledge (about search domain, user features, requires IO etc.) that he considers relevant to the search. This leads to the following problems: on the one hand, users don't understand the way of result obtaining, and it reduces the credibility of the search engines; and on the other hand, in many cases these SSS are looking for information that is not pertinent to the user needs. Other problems of knowledge processing in the semantic search deal with retrieval of pertinent domain knowledge bases and their dynamical updating.

Use of ontologies as a knowledge base for SSS helps partially in solving of these problems. Now a lot of ontologies are available through the Web [12]. Lot of them are based on standards and tools of the Semantic Web. Ontological knowledge can be processed without any concordance with their developers [13]. But the use of ontologies in semantic search, in spite of the advantages, such as an explicit representation of semantics and a strict mathematical basis (descriptive logic), has some significant shortcomings which include the processing and inference complexity of ontologies. Therefore, the usefulness of ontologies and their particular cases (taxonomies, thesauri etc.) in representation of knowledge about different elements of semantic search is the subject of a separate study.

It should be noted another important aspect of semantic search – the SSS developers, as a rule, are moving by improving of the traditional search engines and don't take into account the experience of creating and implementation of intelligent applications (beginning with expert systems to the creation of multi-agent systems): if system functioning is similar to the human intellectual activity then users require an explanation of ways to obtain the results and clear visual model of system behavior and use of knowledge (such as BDI-modeling of intelligent agents by the mechanism of beliefs, desires, goals and intentions) [14]. The users of SSS also need in explicit and clear explanation of the ways of affecting of his knowledge and preferences on the search result.

Ontological approach for the SSS modeling provides interoperable use and reuse of accumulated knowledge – knowledge can be imported into this model of other intelligent IS (e.g., semantic Wiki-resource) and exported to other systems which work with the same users (for example, users information can be transferred from SSS to the personalized system of distance learning or skill evaluation system).

The most important classes of SSS ontological model MSP are *user* – a person who needs the information, and *information object* – a structured image of the information that the user wishes to receive from SSS [15].

Semantic search model displays user information that contains such features:

- personal user preferences, his knowledge and ability to perceive information;
- information needs of users, his particular problems and the structure and semantics of the information object that he need to solve these problems;
- experience of user interaction with the search engine;
- user opinion about search experience of other SSS users and his decisions about collaborative research feasibility .

User model is compared with the models of IRs that describe their content and semantics. IR model includes IR formal properties, semantic markup (by domain ontology concepts) of IR content, metadata about IR and it`s evaluations by different users.

In addition, ontological model of interaction between users and IR contains such classes as *domain ontology* which describes the sphere of the user information needs of the; *lexical domain ontology* allows to mark up natural language texts by domain ontology concepts; *user task* – description of the task that is solved with the help of search results; *task thesaurus* – ontology concepts that by their weight characterize current user task; *request* – message explicitly entered by for representation of his current information needs (usually a set of keywords that characterize one of their information needs of the user associated with a particular task with the help of a thesaurus); *theme* – the set of requests related to the same group of information needs of one or different users allows to combine semantically related queries; *query result* – links on IR and their evaluations; *information resource* – relative autonomous element of information environment of SSS; *recommendation* – information provided by the user proactively, as a result of the actions of other users participating in collaborative search; *recommended element* – a reference to IR that is a part of the recommendation; *information environment* – the collection of all available IR, their properties (including their assessment of users) and links between them; and a *group of users* – the subset of SSS users grouped by some conditions for collaborative search.

In order to formalize the description of the model elements and to ensure its automatic processing, it is advisable to use the ontological representation of knowledge about them. Thus, the information needs of users are described by building of ontological model which classes are corresponded with the above elements, and relations are links between these elements (fig.1).
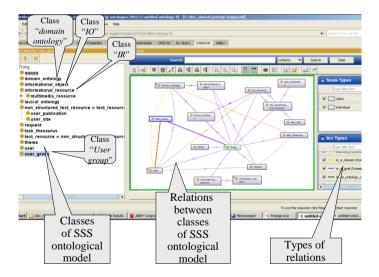

Fig. 1. Ontological model of SSS

In addition, this ontological model contains two auxiliary classes that describe the knowledge and interests of the user – "competence" and "atomic competence". The set of individuals of atomic competencies can be imported from different domain ontologies deal with educational institutions, employment and research organizations [13].

This approach is oriented on relatively narrow subset of the search engines, namely those which are intended to support research activities and users with stable, complex structured knowledge, skills and information needs in a fairly narrow domain. The set of atomic competencies formalizes not all knowledge accumulated by user in the learning and work process but only a small subset of them that is actively used in practice. For example, if researcher is specialed in the field of software engineering then a set of competencies for this discipline will be formed for him (as well as parts of the competencies of other disciplines that are used directly – for example, the theory of algorithms and system analysis). The set of atomic competencies can be enriched by the semantic processing of the scientific publications of user that are marked up by domain ontology concepts.

The use of a class "competence" and it's properties provides more formal description of the properties of the class "user" that allows more simple comparison with domain ontologies and IR properties. It is expected to use a system of weights that assess user proficiency in each competency to distinguish between actually used competencies and ever studied ones. Search of specialists of some fixed level all the weights that are below the this threshold are replaced by zero – it reduces the computational complexity of the task,

To facilitate the formation of a set of competencies of each user, it is advisable to use a hierarchy of competencies where competencies of specialty or area of expertise are determined at the highest level, and the competencies of particular disciplines or areas of activity are determined at the lower levels. Atomic competencies are situated at the bottom of taxonomy.

A significant part of atomic competencies of user can be imported automatically from the set of competencies of his specialty (or specialties) confirmed by official diplomas and certificates, but with the use of weights, allowing them to take into account the current state of the user. For example, if user during the first year at university studied mathematical analysis but then never applied it then the competencies relevant to this discipline receive a non-zero but low weights. However, this user differs from those who don，t acquainted with this discipline at all.

Some user competencies can be imported from the social network by recording of his areas of interest.

Keep in mind that the majority of experts with sufficient experience receive and develop a significant part of their competence in process of their professional activities, not only by education. So sometimes it is advisable to import the expert competencies from the descriptions of their positions (although for many specialties related to scientific and research work such descriptions are so general that they are not suitable for analysis), and take into account their experience in this field.

Ontological model of SSSS displays connection between:

- users tasks and domain of search – as the relations between the task thesaurus and domain ontology;
- IR and domain – as the relations between IR thesaurus and domain ontology;
- requests and tasks – as the relations between the request and task thesaurus;
- different users from the same group.

Let's consider these questions in details by the example of SSS REPS designed to perform multiple information requests in fixed domains.

REPS is a intellectualization of SSS MAIPS enlarged by collaborative and personalized search possibilities and operations on ontologies and thesauri. REPS is based on represented above ontological model. The system is oriented on users with continual and structured information interests that require a regular supply of relevant information from the Web (functionally REPS is aimed at the implementation of complex reusable queries in a fairly narrow areas associated with professional or scientific interests of users; requests of users can be repeated from session to session or be changed but deal with the same domain of user expertise).

Domain ontologies formalizes the sphere of users interests and tasks thesauri built on base of these ontologies by the task description   specify queries relating to current user tasks.

7. Sources of Information about Elements of the Ontological Model

Each of the elements of the information model REPS is appropriate ontology $O_{REPS} = \langle T_{REPS}, R_{REPS}, F_{REPS} \rangle$ class or an individual of class, and information about them is stored as a description of instances of these classes and their properties and relations with instances of other classes.

For the purposes of information retrieval a special case of the ontology is used. $O_{REPS} = \langle T_{REPS}, R_{REPS}, \varnothing \rangle$. For information about instances of classes from the set T_{REPS} are accumulated in the process of interaction between REPS and it's users who try to satisfy their information needs.

Let us consider information sources of O_{REPS} and methods of acquisition of this information into the ontology. Even before the beginning of interaction with the users this ontology contains certain knowledge necessary to perform the search at the semantic level and during it`s work this knowledge can be replenished).

Sources of information about **domain ontologies.** REPS uses for knowledge representatiov ontologies of two types – internal and external. Internal ontologies are created directly by REPS developers and can be replenished by new individuals (not by classes) in the process of interaction with REPS users. The main feature of such ontologies is the structure fixed by developers that helps in prediction of the algorithm convergence and calculation time. In addition, prior knowledge about relation between ontology classes and instances can be processed and analyzed by specialized services (e.g., to enhance the domain knowledge concerning the field of user interest the service of the semantic analysis of selected IR can be used, and specialized services rapid tests can help to determine the psycho-physiological characteristics of the user) .

External ontologies allow to integrate into REPS the dynamic and distributed knowledge accessible by the Web. The search of such ontologies can be performed in the various ontological repositories, or they can be formed in the process of the user interaction with other intelligent applications. In general, we have no information about the complexity of the ontology structure, about descriptive logics on which they are based, completeness of their knowledge, and it prevents to predict the work time of algorithms for deep analysis and processing of these ontologies. Therefore REPS handles external ontologies by specially simplified algorithms that use only the most general properties of ontologies (e.g., processing only of "Class-subclass" relations of concepts). that's enough for the majority of information search tasks but search of complex sets of information objects – for example, for discovery and composition of the semantic Web-services, entities of the Web of things or organization of multiagent systems – we need to use more complex structural relationships from ontologies. For these situations REPS includes internal ontology of complex information objects.

The individuals of the class "domain ontology" contain links to external ontology (as opposed to internal REPSontologies information on the structure and complexity of such ontologies is not available) and a description of their subject domain. This information is entered into the ontology manually by system developers and is based on direct analysis of the contents of these ontologies or is exported from external repositories which REPS developers have recognized as trusted.

If REPS users are interested in the domains which are not represented by the existing REPS ontologies then the users can submit a request to the developers on the replenishment of the REPS ontologies by relevant to their need ontologies. There is important that user can not directly add to the list of ontologies without coordination with the developers because usually the user can not estimate the quality and computational complexity of domain ontology (for example, to determine the kind of appropriate descriptive logic), and therefore he can't not predict an efficiency of ontology processing by REPS.

Sources of information about lexical domain ontologies. Lexical domain ontologies are developed automatically for all domain ontologies of REPS. Source of information about lexical domain ontologies is a specialized unit of the lexical analysis which is semi-automatically connects the domain ontology terms with relevant to them fragments of natural language texts. Each instance of class "lexical ontology" contains the

concept of domain ontology and the set wordforms – natural text fragments lexically correspondent with this concept.

In its simplest form (for example, for those languages that REPS is not supported by special methods of lexical analysis) such lexical matching is a set of pairs <ontology concept, text string> where "text string" is equal to "ontology concept". Such pairs can be formed manually and semi-automatic. Furthermore, existing ontology vocabulary may be modified by the user with tailored to the task. Then they can be used for semantic markup of natural language IR from user viewpoint.

Sources of information about thesauri. Task thesaurus is constructed directly by the user in order to display the specifics of the task which causes these information needs.

User can build a thesaurus on base the selected ontology domain by noting important for task concepts and determining this importance by their weight. Users can also manually edit any previously created thesaurus by adding or deleting some terms. In addition, REPS realizes such set-theoretic operations on thesauri as union, intersection and complement.

One more way of thesaurus construction is based on analyzing of the natural language task description with the help of lexical ontology. We can consider this operation as a projection of the task to the domain: this service finds in the task description those fragments that correspond to the terms of the domain ontology and adds these concepts to the thesaurus with weight below zero.

Sources of information about requests. Each instance of class "Request" contains a set of keywords that the user provides to SSS, semantic context of this request formalized through the task thesaurus, reference on ontology that characterize the subject domain of search and user information that reflects his personal specifics in information consumption. This information is entered by user. REPS as a system aimed at satisfying of permanent information needs of users provides storing and multiple execution of queries.

Sources of information about themes. Themes are created by users to combine semantically related queries. One topic may include the needs of different users which are based on different ontologies and thesauri. The theme can be created by the user explicitly by enumeration of requests. Another way of theme forming is an automatic grouping of requests by some formal conditions – for example, referenced on the same domain ontology or with overlapping thesaurus. Set-theoretic operations on themes (union, intersection and complement) simplify significantly their creation and editing.

Sources of information about query results. When REPS receives user request it redirects the set of request keywords to the external IPS (user can choose one of the proposed IPS, for example, Google (www.google.com) for search in the biggest index base or Google Acedemy (https://scholar.google.com.ua) for search of scientific publications. Then REPS receives from this IPS the list of links to IR relevant to this request (the representation of this information depends on individual specifics of selected IPS) and reorders this list on base of domain ontology, task thesaurus, personal user data and, in certain cases, some other information (for example, an experience of interaction with other REPS users, evaluations of proposed IR).

Sources of information about REPS users. User information provides the basis for the personification of semantic search and creation of recommendations [16]. Analysis of user knowledge makes possible to retrieve information much more efficiently. Partially this information is entered directly by users (during the registration and later by express tests) or acquired from user behavior in work with REPS (data about domain ontologies, task thesauri, evaluations of proposed IR etc.). User can propose additional information indirectly by assignment of natural language IR that characterize user – for example, user personal site, Wiki pages, publications, user data from social networks.

Sources of information about the groups of REPS users. REPS is intended for a personalized and collaborative information retrieval, so the system must have the ability to group users respectively to semantics of their information needs. All kinds of grouping we can divide on two types – directly by user or automatically by REPS on base of user criteria.

The easiest criterion of grouping is an integration of people who use the same domain ontology as a basis for information requests. This means that users are interested in such a knowledge domain. However, this classification is very rough due to the fact that one ontology may describe a broad subject area but user's

interests relate to it's narrow subsets. It should also be borne in mind that a single query can use terms from different ontologies, as well as terms that user adds himself. More accurate criterion of grouping use the task thesauri and measures of their intersection (with an allowance for weights of terms). In order to create a system of proactive recommendations REPS ontology contains some additional classes designed for collaborative accumulation of knowledge: group of ontologies; group query terms and group of intersected thesauri. These classes are used to locate users who have such information needs.

User can select the set of other important criteria for grouping (for example, users with the same profession, place of employment, residence, sociopsychological type etc.) or enumerate users or request that are interesting to him.

8. Use of Ontological Model of Semantic Search in Applied Information Systems

Proposed above model of semantic search can be integrated into solving of various applied tasks [17]. We consider it on example of the problem of competence matching. This problem is a component of such tasks as finding of appropriate contractors by the employer; comparison of skills of professionals with different specialties (in particular, defined by the standards of different countries); choice of educational institution relevant to university/college entrant needs; estimation of the possibility of student transfer from one institution to another (what disciplines from a previously studied can be counted, etc.

Model of search domain describes the basic concepts and the relationships between them, as well as the structure of information object which is the search result.

User describes his information need by indicating of the domain ontology class A which describes required IO human, institution, specialty, etc. and conditions imposed on it by the set of instances of class "atomic competence" associated with selected IO by various object relationships from domain ontology.

Thus, the request execution comes to comparison of the values of class "atomic competence" instances that are related with instances of different domain classes – for example, mapping of property "involves competence" of class instance of "expert" which refers to the class of "atomic competence" with the property "provide competence" of class instance "specialty".

This mapping clearly indicates the semantics of information needs, and so we can retrieval of IO with differentiating of the various relations between the required IO and the set of competencies. For example, the same instance of the class "person" has some different subsets of atomic competencies that are associated with object property "know", "are certified", "can teach", "has practical experience". This allows to more accurately meet the information needs of users by retrieval of IO corresponded to his real requirements. There is advisable to use a set of standard atomic instances of each class for comparison of specialties, skills and competencies of people and organizations from various countries with different classifications and standards [17].

An entity A of class X is atomic if any other instance of this class X is not a subset of it: $\neg \exists B \in X, B \subseteq A$. Atomization is a process of building of new entities by the set of the non-atomic ones X the result of which is a set of atomic entities Y.

$$X = \{A_i\}, i = \overline{1, n}, \ Y = \{C_j\}, i = \overline{1, m}, \ \forall j, k = \overline{1, m} \text{ if } k \neq j \text{ then } C_k \not\subset C_j,$$

If two competencies from X are overlapped $\exists A_s \in X, A_t \in X, s \neq t, A_s \cup A_t \neq \varnothing$ then we construct tree new potentially atomic competences

$$A_{s`}, A_{t`}, A_{s\_t} = A_s \cap A_t, A_s = A_{s`} \cup A_{s\_t}, A_t = A_{t`} \cup A_{s\_t} A_{s`} \cup A_{t`} \cup A_{s\_t} = A_s \cup A_t$$

and repeat this operation recursively until all elements of the competence set becomes atomic.

Therefore we add special class "atomic competence" that is a subclass of "competence", to the ontological model (fig.2). Every instance of "competence" connects by object property "include_competence" at least one element of class "atomic competence".
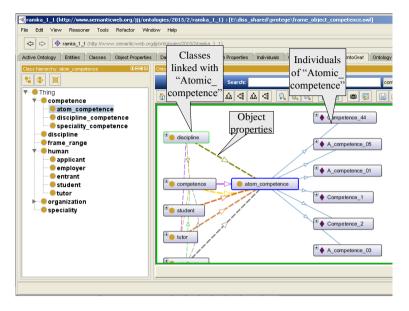

Fig. 2. Visualization of object properties of class "Atomic_competence" of the REPS ontological model

User can describe his information need by indicating the class of the domain ontology A (empoyer, discipline etc.) and impose conditions on it as a set of instances of class "atomic competence".

Let`s consider one of possible spheres of practical use of this approach. Eurointegration processes in Ukraine, balancing of national interests on the education and the labor markets, improving of the quality of training, on the one hand, and technological, economic and demographic processes, globalization, increasing labor mobility – with another, require the development and implementation of the National Qualifications Framework (NQF). NQF should promote transparency in establishing the relationship between diplomas and education certificates, expand access to acquiring of qualifications. This project should become a catalyst for modernizing of the national education system. Implementation of NQF includes development of tools for national and regional qualification policy, of institutional procedures, and quality assurance processes. It provides assessment and awarding of qualifications, recognition of such learning outcomes as knowledge, skills and competencies that link the education system with the labor market. The successful NQF operation will contribute to the development both of vertical mobility of citizens in the area of professional education and horizontal mobility from one professional trajectory to another one through the mechanism of official recognition of partial qualifications.

The national qualification system of Ukraine deals with NQF coordinated with the European Qualifications Framework (EQF). Their integration provides greater understanding and better comparison of national qualifications and implementing of the "education through life" paradigm. Development of intelligent information systems for cognitive support of national qualification framework based on ontological approach is seen relevant and timely scientific practical task.

This system can facilitate implementation of European standards and principles of quality assurance with the requirements of the labor market to competencies to professionals; ensure the harmonization of legislation in the field of education and industrial relations both in Ukraine and abroad; promote national and international recognition of qualifications acquired in Ukraine. The system would be useful for various organizations and

people: the executive authorities, institutions implementing state policy in education, employment and industrial relations spheres, educational institutions, employers etc. for identification, matching, recognition, planning and development of qualifications.

Unfortunately, the process of formation of the set of atomic competencies requires considerable intelligent efforts of experts and can be automated only in part. In addition, replenishment and actualization of the competence knowledge base needs in analyzing of the large amounts of data that can be retrieved from the Web resources:

- various national and international standards and guidelines related to the description of the structure and level of qualification framework;
- official sites and organizational ontology of educational institution containing authoritative information about their specialties and lerned disciplines;
- external evaluations of various educational institutions and organizations that provide the comparison of the quality of the competence study;
- personified information about people derived from social networks, Wiki-resources and various natural language documents that are available through the Web (scientific publications, technical reports, teaching materials, etc.).

9. Conclusions

Use of the ontological model of interaction of the semantic search subjects and objects is a base of integration of the semantic search component with different applications: this model explicitly defines all features of it's components and their relations but allows to expand it by task-specific concepts and properties (by references on external domain ontologies of by transformation of this ontology). The development of methods and tools for automation of this process is a task for further work.

Other important aspect of use of this ontological model deals with search personification: ontological model helps to user in understanding of search engine possibilities and in explaining of search results (by causal relationship of used knowledge, search condition defined by user and properties of information objects). But ontological analysis in general is too complex for unskilled users, and it caused the necessity in development of means of visualization and adaptation of ontological knowledge to user needs.

References

[1] Gladun A, Rogushina J, Valencia-Garcá R, Martínez-Béjar R Semantics-driven modelling of user preferences for information retrieval in the biomedical domain. Informatics for health and social care. V.38, N.2 2013: P.150-170.
[2] Minsky M Progress Report on Artificial Intelligence. Seymour Papert ec 11, 1971. – http://web.media.mit.edu/~minsky/papers/PR1971.html.
[3] Russell S, Norvig P Artificial intelligence: a modern approach. Upper Saddle River, New Jersey: Prentice Hall, 2003.
[4] Nilsson N Artificial Intelligence: A New Synthesis. Morgan Kaufmann. 1998.
[5] Luger GF Stubblefield W Artificial intelligence: structures and strategies for complex problem solving. Pearson education. 2005.
[6] Negnevitsky, Michael. Artificial intelligence: a guide to intelligent systems. Pearson Education, 2005.
[7] Nayak R, Ichalkaranje N, Jain LC. eds. Evolution of the Web in Artificial Intelligence Environments, Vol. 130. Springer, 2008.– 277 p.
[8] Lyubich AA, Pleskach VL, Rogushina JVAbout selection of criteria for evaluation of information system intelligence. UsiM, № 1, 2005:3-7. (In Russian).

[9]   Gavrilova TA, Horoshevsky VF Knowledge bases of intelligent systems. St. Petersburg: Piter. – 2001. – 382 p. (In Russian).

[10] Kuzmenko GE, Litvinov VA   Pragmatic approach to estimation of intelligence level on intelligent systems. Mathematical machines and systems. № 1, 2003:3-9. (In Ukrainian).

[11] Amerland D. Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and Amplifies Your Online Presence. Que Publishing, 2013. – 230 p.

[12] Davies J, Fensel D, van Harmelen F Towards the Semantic Web: Ontology-driven knowledge management. John Wiley & Sons Ltd,, England, 2002. – 288 p.

[13] Staab S, Studer, R. (Eds.) Handbook on ontologies. Springer Science & Business Media. 2013.

[14] Rao AS, Georgeff MP Modeling rational agents within a BDI-architecture. In R. Pikes and E. Sandewall, eds.. Proc. of Knowledge Representation and Reasoning (KR&R-91), Morgan Kaufmann Publishers: San Mateo, CA, 1991:473-484.

[15] Rogushina J, Gladun A Ontology-based competency analyses in new research domains. Journal of Computing and Information Technology. V.20, N. 4, 2012:277-293.

[16] Ricci F, Rokach L, Shapira B, Kantor P Recommender Systems Handbook. Springer, 2011. – 842 p.

[17] Gladun A, Rogushina J, Schreurs J, Salem Abdel-Badeeh Ontology-based knowledge recognition in service-oriented virtual research environments. Proc. of The 7th International Conference on Information Technology ICIT-2015, Al Zaytoonah University of Jordan, Amman, Jordan, 2015:148-155.

Authors' Profiles

Dr.Julia Rogushina (1967) is a currently working as a senior researcher at the Institute of Software Systems, National Academy of Sciences of Ukraine. She has presented papers in many national and international conferences and published articles in many international journals. She has written books for Computer Science in the field of ontological analysis and semantic search including monographы "Agent technologies", "Knowledge-oriented means of the Web semantic search", "Ontology-based knowledge management for distant learning". Her research interests include intelligent information retrieval, theory of software agents behavior, inductive knowledge acquisition, ontological analysis, the Semantic Web technologies.