

# Means of the Semantic Search Personification on base of Ontological Approach

Rogushina J.

*Institute of Software Systems of National Academy of Sciences of Ukraine, Glushkov, 44, Kiev, Ukraine*

---

## Abstract

The main trends of information retrieval deal with its personification and semantization are analyzed. Sources of knowledge about main subjects and objects of the search process are considered. Ontological model of interaction between the Web information resources and information consumers is proposed as a base of the search personification. Methods of development, improvement and usage of this model are defined. User characteristics are supplemented with sociopsychophysiological properties and ontologically personalized readability criteria.

Software realization of semantic search on base of this ontological approach is described.

**Index Terms:** Semantic search, ontological model, personification of search, readability, information object.

© 2016 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science

---

## 1. Introduction

Now a lot of complex and comprehensive systems of information retrieval are widely used either for global or for local search of different information resources. Information retrieval systems (IRS) index the great number of *information resources* (IR), and maybe all possibilities of algorithms for more relevant matching of user requests and these IR have been realized. Therefore the search efficiency improving depends on semantization and personification of retrieval process. Semantic search has to use knowledge about IR and IRS users – their needs, interests, tasks and possibilities. Knowledge about individual user allows IRS to take into account his/her personal preferences and characteristics of specific problem for which user tries to find some information.

*Semantic search* is the process of information retrieval that satisfies user's information demands in process of problem solving where knowledge about various subjects and objects of this search procedure and methods

\* Corresponding author – Rogushina Julia.  
E-mail address: [ladamandraka2010@gmail.com](mailto:ladamandraka2010@gmail.com)

of knowledge analysis are used [1].

In this work we consider an *information object* (IO) as a model of some 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 [2]. 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 information space as ontologies, software agents, Web-services, information resources, metadata, databases, etc. are classified as IO too.

A lot of IO has 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.

In semantic search the relevance of a document and query is defined semantically (by the content proximity). The meanings of knowledge that are used in semantic search, facilities of their representation and processing depend either on the specifics of developed *information retrieval system* (IRS) or on the concept chosen by IRS developers, but in general the search results are a function of the user's query, the IRS index and knowledge contained in Knowledge Base (KB) of IRS:

$$I_s = \{i_j, j = \overline{1, n}\} = f(z, DB_{IRS}, KB_{IRS})$$

In case of the Web semantic search the Web can contain not only the retrieved information object (IO) but also external knowledge base used for the search semantization. Therefore, the creation of such systems should take into account that these external KBs can change their content, structure and accessibility independently of the IRS developers. The results of the Web search are the function that depends also on the content of external KB:

$$I_{web\_s} = \{i_j, j = \overline{1, n}\} = f(z, DB_{IRS}, KB_{IRS}, \{KB_{web\_k}, k = \overline{1, m}\})$$

Note that today many IRS (e.g., Google) tend to accumulate and use the experience of interaction with a specific user. But often the user information needs are limited in time (for example, the user accumulates information for selection of new phone, buys it and no longer needs the data about phones but IRS continues to offer him this information) or not connected with him (for example, query is executed for other user). In addition, quite often user does not want to make public some part of his information needs – for example, queries related to vacation or health he does not want to mix with work requests. Therefore users prefer to have opportunity to include the query to one of specified profiles (“work”, “hobby” etc.) or not store it for further processing.

*Semantic search system* (SSS) is IRS that semantically enriches user requests by knowledge acquired from such *information resources* (IR) as following:

- *Personal profile* of this IRS user containing information about individual features and advantages of the user in information consuming, personal registration data (login information: user ID; password for the SSP, date of registration etc.), questioning results (for example, physiological properties of the user that affect its ability to perceive information), competence level in different areas, history of previous requests, the model of the interests domain etc. [3];
- Applied semantic *information about domain* of interest, explicitly or implicitly provided by the user – e.g., ontology of relevant domain, thesaurus of task, links to external knowledge base (KB) [4];
- *SSS experience* and history of interaction with user – user queries to IRS (or queries of subgroups of similar users);

- Information *imported from external knowledge sources* obtainable through the Web (for example, Wikipedia, dictionaries, personal Web-Sites).

In order to develop the effective means of personification for intelligent information systems (IIS) we need to determine what user information can improve the functioning of these IIS; the sources for extraction of this information; construction techniques for conversion of suitable data in automatically processed knowledge; knowledge representation form that facilitates their reuse; and the structure of the user information with account of solved problem domain. To address these issues the ontological model of information interaction of user with IIS that is focused on knowledge acquisition from external ontologies, thesauri and distributed Web resources is proposed.

## 2. Related Works

Web search is a key technology of the Web which is essentially based on a combination of textual key-word search with an importance ranking of the documents depending on the link structure of the Web. In [5] semantic search on the Web is considered as a form of search where meaning and structure are extracted from both the user's Web search queries and different forms of Web content, and exploited during the Web search process. Different approaches to semantic search on the Web can be separated by query types:

- Approaches on base of structured query languages [6, 7];
- Approaches where user don't know query languages (keyword-based approaches [8, 9] and natural-language-based approaches [10, 11].

A lot of semantic search systems use ontologies. Ontological knowledge and annotations that are underlying our semantic search on the Web can be classified according to their contents, they may describe general knowledge (such as the knowledge encoded in Wikipedia) or they may describe some specific knowledge for vertical ontology-based search on the Web.

In [12] search engines that focus on information retrieval from the Web and criteria of their classification (architecture, coupling, transparency, user context, query modification, ontology structure and ontology technology) are considered. A large number of promising approaches to semantic document retrieval are analyzed to show an importance of use of the search personification means that are understandable to user.

A broad range of approaches to semantic document retrieval has been developed in the context of the Semantic Web. Ontological analysis is a powerful instrument for semantic search. But many problems are not solved now and we need in new researches deal with more pertinent satisfaction of the information requirements of people. The important question is a choice of the set of parameters for semantic matching of user needs with IR properties.

## 3. Features of Personalized Semantic Search

It should be noted that further a special case of the personalized semantic search where the user agrees to expend some efforts for improvement of search efficiently is considered. User can do it by a number of reasons, and in each case IRS should take into account the specific situation.

Consider the most typical situations:

- User has stable long-term information needs in the areas where he is an expert and where the domain structure and structure of IO are well-known to him (in particular, researchers, analysts and technical experts);
- Decision that user is going to accept on base of the search results is very important, but it is difficult to user to evaluate the quality of the results without assistance because he did not have sufficient

knowledge about this domain. For example, this situation can deal with some crucial uncommon for user decision with the risk of money loss or harm to health. In this case user can analyze an experience of other users in similar situations but for effective analysis user has to describe the solvable problem and his own parameters in details.

- It is difficult to formalize user needs and preferences although user can easily rate different IO. This situation is typical in search of artwork in literature and music or choosing the persons for convenient communication – situations where ratings are defined by subjective factors. In this case, effective user search can be based on generalization of user relevant IO ratings and the search engine try to find generalized properties of "good" IO.

The user can have interests in various fields and solve various unrelated one to the other problem, but his own properties in consuming of information remain rather stable and depend on the physiological characteristics of the person and therefore the IRS can use this information in solving of new problems.

To ensure the effective access of various IIS to Web resources it is necessary to develop appropriate means for semantic search for these resources. Learning methods on base of inductive and traductive inference increase the pertinence of this *search and recommending system* (SRS) results – they help in acquisition of personal user knowledge. But these methods are effective only if they use adequate models of users and information resources that describe their interaction. Ontological approach to modeling of semantic search ensures interoperability of knowledge and possibility of its exports from the Web.

Recently a lot of research projects in information technologies (IT) are directly associated with the development of intellectualization and personalization of various applications. *Intelligent information systems* (IIS) support knowledge-based adaptive interface with the user that is able to consider both the context of interaction and user-specific features. Another important aspect of semantic design technologies of intelligent systems is the personification of content that IIS offers to different users. It is requires the means for acquisition and processing of knowledge either about users or specifics for domain of user task. Development of personification means for IIS needs to determine structure and content of user data, it's sources and methods of acquisition, form of representation.

The ontological model of user interaction with information IIS focused on bringing knowledge from external ontologies, thesauri and distributed resources of the Web can be used in solving of these tasks.

IR personification allows to:

- Provide information in a form that is better interpreted by the particular user (choice of visual and structural data presentation, determination of presentation speed, the use of additional features etc.);
- Select content required to particular user (selection understandable but interesting to user information – using the personalized and domain-specific criteria readability, providing new information resources for the user based on collaborative search, taking into account the knowledge about area of user interests, his experience and solved task descriptions, etc.);
- Dialog with the user in the optimal for him and his task manner (intelligent adaptive interface, choose the IIS dialogue type and level of interface detailing) – for example, if the user makes a lot of mistakes during the data input than changing the interface to more "calming" allows him to work more efficiently;
- Proactively offer recommendations to user – the information that he clearly did not ask but that information can be useful in solving of his the problem or can change his status to more efficient by viewpoint of the work or psychological comfort (for example, propose link to dictionary if user makes many spelling mistakes).

#### 4. Sources of Personalized Information about User

Let us consider IIS sources of personalized information about its users.

First of all, this is the information that the user himself clearly advertises to IIS – during the registration of later in the process of interaction with the system.

On the one hand, this is the easiest for IIS way to obtain information because the user makes the main part of the efforts. In this case, developers need only to plan what information can be useful in IIS adaptation for user needs and provide a mechanism for its input – the relevant questionnaires, tests, etc. The disadvantages of this method deal with its complicity for the user and the insufficient likelihood of information proposed by user (for example, the user can inflate his qualifications or misunderstand the question).

More objective information about user can be obtained in the process of testing. For example, special tests can determine the user's physiological characteristics associated with the ability to perceive information (e.g., extroverts and introverts differently range information in various forms of presentation), speed of perception, the ability to understand complex information objects characterized by different readability criteria) and erudition. But in this case user has to waste a lot of time to provide to IIS this information.

This is useful only if the user plans the long-term interaction with IIS or if it is important to obtain from IIS outcome that takes into account his personal characteristics. Such information as opposed to dynamic knowledge about IR and user tasks is relatively stable – user data is practically unchanged (it can only grow by user learning), but personal preferences are almost identical in applying to different types of IR. For example, physiological properties such user benefits in the form of information submission are permanent either for the analysis of the scientific article or for fiction reading. This allows SSS to collect and integrate such information from various IR.

## 5. Ontological Model of Semantic Search

To more accurately formulate the theory of intelligent information retrieval the semantic model of their basic components, their properties and relations between them should be strictly defined (fig.1). This model describes the IRS user, information resources, requests and other components important to retrieval. The most important of them are information object that satisfy user information need and user. Other classes are used for semantic matchmaking of user request and acquisition of knowledge about IO from retrieved IR with account of user specifics and tasks.

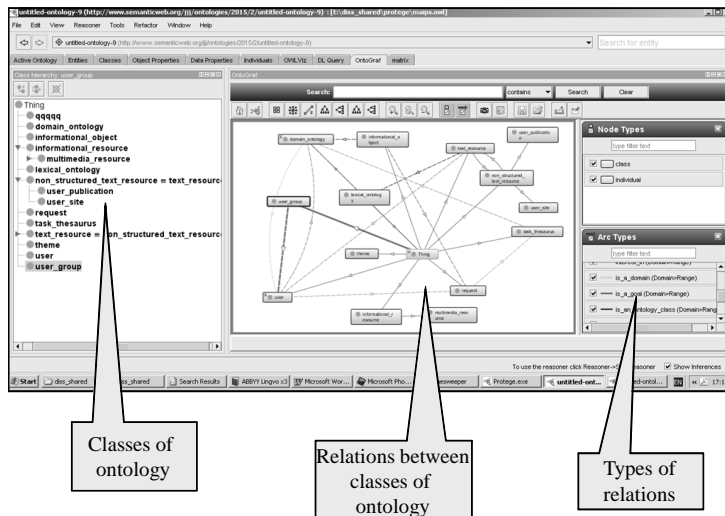


Fig.1. Ontological model of semantic search

Ontological model of semantic search includes these components as following ontology classes:

- *Domain ontology* describing the area of the user information needs  $O_{\text{domain}_i} = \langle T_{\text{domain}_i}, R_{\text{domain}_i}, F_{\text{domain}_i} \rangle, i = \overline{1, n}$ ;
- *Lexical ontology* of domain containing information about natural language fragments corresponded with domain ontology terms  $O_{\text{lex}_i} = \langle T_{\text{lex}_i}, R_{\text{lex}_i}, f \rangle, i = \overline{1, n}$  that  $\forall t \in T_{\text{domain}_i}$  – ontology that is used to establish links between elements of natural language documents and terms of ontology;
- *Task thesaurus* – the set of pairs where the first element is the term of ontology which characterizes a specific user task  $o$ , and the second – the weight (positive or negative) of this term for this problem  $Th_{i_j} = \left\langle \left\{ th_{k_{ij}} \in T_{\text{domain}_i}, v_{k_{ij}} \right\}, k = \overline{1, s_{ij}}, j = \overline{1, m_i} \right\rangle$ ; request – the set of keywords describing the user information needs associated with a specific task by using of the task thesaurus  $z = \langle \{k_q\}, Th_{i_j} \rangle, q = \overline{1, u}$ ;
- *Theme* – the set of queries related to semantically similar information needs allowing combining semantically related needs of different users, based on different ontologies and thesauri  $z = \langle \{k_q\}, Th_{i_j} \rangle, q = \overline{1, u}$ ;
- *Query result* – the set of pairs where the first element is IR reference and the second – user rating of this IR  $rez = f(z, u) = \langle \{id_{ir}, rating_{ir}\} \rangle$ ;
- *User* – a class that has more complex structure and includes the following attributes which can be classified into several groups:

Group 1. Registration information: - User ID; - Password for the IRS; - Registration date;

Group 2. IRS experience of interaction with user: - ontologies that formalize domain of user interests; - thesauri that describe user tasks; - earlier user requests and their results;

Group 3. Social user characteristics imported from external sources (optional information): - User IDs in social networks allowing update dynamically information about him; - User ratings in social networks; - Address of the user in Wikipedia and other resources; - User site address; - The scope of user competencies (keywords imported from social networks); - A reference to the user publication;

Group 4. User self-certification: - User sphere of competence (a list of keywords entered by the user directly); - Competence level for every keyword;

Group 5. biographical user data (optional information that allows the IRS to form a group of users with similar characteristics): residence; age; profession, education, etc.

- *User group* – class  $gr = \langle id_{gr}, \{u_i\} \rangle, i = \overline{1, n}$  with such properties as group ID and the user list combined into this group by certain reasons (the group may be formed explicitly by user selection or automatically based on compliance with any conditions such as a group of users with similar biographical data or similar requests);
- *Information resource* – information about IR previously found by IRS and their ratings  $\langle U_{url}, \{z_i, m_i, q_i\}, i = \overline{1, n} \rangle$  that include the resource identifier, requests by which it was discovered, user rating and it's personal level of readability for this user;
- *User information need* which user tries to satisfy by request to IRS;
- *IR* – document which IRS can access;
- *IO* – information object contained in one or more IR with a certain structure defined by user;

- *Information environment* – collection of all available IR, their properties (including their user ratings) and relations between them;
- *Domain* – some part of information environment separated by some content properties;
- *Recommendation* – information proposed to user by IRS proactively, as a result of analysis of personal user information and of collaborative search experience – interaction with other users of selected group.

## 6. Personification of IIS Interface

Now many studies are oriented on problems of adaptation of information systems to users. One of the most common methods of solving this problem deals with user model, and increasing of model complexity provides higher degree of adaptation but it causes an increasing of the model creation cost. In particular, such approaches are widely used in developing of search and recommending systems.

Strategy of adaptive interface building can be represented as follows: each user receives some personal configuration of system and its services that is optimal for him and is derived by IIS from the set of user answers on interface questions. If user is already known to system then the subsystem of user modeling analyzes his last search session and updates the user model. Usage of ontology for user modeling for IIS with adaptive interface allows considering a set of domain concepts and relationships between them that occur in user interaction with IIS. The problem of user model constructing is an identification of values of the individual user characteristics that provide IIS ability to adapt to the needs, opportunities and objectives of this user. This user model contains classes for such concepts as "information system", "interface requirements", "user", "methods of user interaction with the system", "methods of user model forming" etc.

Characteristics of these concepts are represented as data and object properties of these ontology classes. Data properties describe the values of corresponding characteristics and object properties define the relations between individuals of this class and individuals of other classes. For example, the term "user" has following characteristics: demographic characteristics; user knowledge about IIS; user experience of interaction with IIS; user interface benefits; user goals and plans in problem solving; self-interest in the outcome; cognitive features; user specific properties (psychological, physical restrictions etc.).

Often adaptive intellectual interface of IIS is considered as a means of their personification. Now the lot of conflicting approaches to adaptive human-machine interface exist. This is explained by the IIS complexity: on the one hand, this complexity can cause sudden and unpredictable system behavior, on the other hand, owing to its complexity IIS can change their behavior for adaptation to changing conditions.

Adaptive IIS interface allows the system without human assistance, as a result of analysis of solved problems in current situation, adjust representation of user information and user actions that correspond with his current goals and opportunities. The current situation and the solved problem are the factors generated by controlled object and environment of activities.

Interface adaptability provides filtering and ordering of content offered to user – depending on the user preferences, context and external factors.

## 7. Physiological Characteristics of SSP Users

Physiological properties of the user that affect his ability to perceive information are important factors of IIS personification. Most of present IIS don't use these properties that's why we consider them in detail. Now a lot of psychological tests are used for clearly classification of personality (e.g. Eysenck test IQ, MMPI) [13]. These classifications can be used for characterization (in terms of the representation form rather than actual content) of IR that are suitable for user of different classes [14].

For example, common classification of personalities by the rate of excitation and inhibition that distinguishes four types – choleric, sanguine, phlegmatic and melancholic, or division of people into extroverts and introverts like. Sociopsychophysiological study is provided by a number of computer systems that are used design of individual adaptation sociopsychophysiological portrait of human in different spheres of life and consider the

problem of creating comfortable for user individual adaptation and information space.

Individually-typological properties of human are strength, mobility, balance of nervous processes, the type of response, emotional background, speed characteristics of analyzer systems, features of the higher nervous functions (perception, attention, memory, thinking, intelligence); behavior in stress, the type of optimal efficiency. They can greatly determine the user's choice in the perception of information – and, accordingly, in rating of the proposed IR and in choosing the type of searched IO.

For example, positive features of IR for extrovert are the presence of illustrations, the use of short simple sentences, eclectic design, the use of bright colors to highlight important elements, etc., and positive IR features for introvert are complicated text structure, compound sentences, pastel color palette. Therefore, we can assume that similar by content IR with different design are ranged in different ways by extroverts and introverts. Use of these physiological user parameters in inductive inference provides more accurate generalize rules for personalized search.

SSS can use semantic and physiological characteristics of the user that in terms of problem can be divided into the following groups:

- Permanent (those that do not change over time and the transition to other problems) – for example, classification on choleric, sanguine, phlegmatic and melancholic, division of people on extroverts and introverts etc.
- Rather stable that can be changed over time – for example, by learning, training, health status changes (speed of perception, ability to remember information etc.);
- Dynamic, dependent on the current mood state, work motivation and so on.
- On the other hand, user characteristics can be classified on semantic, psychological and physiological.
- Semantic characteristics are associated with:
  - Intelligence level (generally unchanging characteristics with the possibility of development or degradation);
  - General level of erudition and development that provides terminological and conceptual basis for further education (generally this level only increases during human's life);
  - Availability of domain knowledge (in most cases can only increase, but to dynamically developing domains it is possible that the user knowledge becomes outdated and not suitable for effective use);
  - Skills to obtain new knowledge (previous parameters depend of this feature);
  - Ability to inference (how complex logical structure of this person is able to understand and build yourself): in general, this ability increases with training but with a negative external environment falls (example – manipulated TV viewers).

SSS can use different types of tests whose reliability and scientific validity must be certified by experts-physiologists. The first two types of characteristics – semantic and psychological – can be evaluate by rather complex tests or be exported from other IIS. But physiological information can be received from quick and simple tests or from tools that analyze current activity of the user (for example, mouse click speed) and on the basis of this information fix his conditions.

The availability of psychological knowledge reduces a cumbersome task of user clustering to more compact classification problem – due to the fact that knowledge about the user classes and their possible links comes from the experts, and not acquired from the data collected in process of user interaction with IIS.

Therefore, personalized SSS needs in knowledge about:

- Personalized features of SSS users, assessment of their dynamism and possible values (e.g. users can be divided by "communication orientation" on introverts and extroverts, and this division is quite stable and does not depend on what the user current activities and other conditions);
- Acquisition means of sociopsychophysiological knowledge about users (tests, methods of their interpreting and tools of their integrating with the problems of information retrieval);



- Connection of user properties with personalized SSS behavior (i.e., SSS adapts its behavior for different types of users) – for example, extroverts prefer text with illustrations and introverts – with tables and structuring;
- Rules for determining of relations between types of users and suitable IR (by explicit classification of resources or by ratings of users from various classes); e.g. 80% of extroverts like resource A and 15% introverts like resource A if A is relevant to their needs.

## **8. Social Networks as a Source of SSS Personifications**

One more important source of personal information about users is social Web. User information can be obtained from a variety of social networks, Wiki-resources, blogs etc. A significant amount of information about users can be acquired from the user profiles that are created for them in other IIS because the basic properties of the individual user (character, temperament, reaction rate) in solving of very different problems are unchanged (or change very slowly over time). Therefore reuse of knowledge from other systems is acceptable. In addition, the analysis of applications which user operates and of user requests to the search engines can also enrich SSS user profile.

Some problems can be caused by the fact that different IIS process similar user characteristics described in different terms. To ensure interoperability of personal data ontological approach can be used: ontologies provide integrating of information based on the matching of different IIS ontologies.

## **9. SSS Experience as a Source of User Information**

An important source of personalized information about user is an experience of his interaction with IIS. This data implicitly contains various user preferences and requirements specified for this SSS. That is why the knowledge obtained from such information is the most useful for improving the efficiency of SSS. If user regularly rejects all SSS recommendations with some common characteristics than the identification of these characteristics provides SSS no longer offer him such recommendations. It saves user time and computational resources of SSS. For example, if user in online purchases never prefers red and pink goods then after the analysis SSS stops offer him objects of these colors, even if they clearly meet other criteria such as size and price.

Efficient obtaining of such personified rules demands, on the one hand, the relevant methods of SSS learning, and on the other hand – successful structuring of processed information objects: for example, if the goods description doesn't contain option "color" then even the most advanced learning techniques can not build the rules that distinguish the group of pink objects.

Therefore, it seems appropriate to include to IIS subsystem for extraction of rules and patterns from experience. Personalization knowledge acquisition can be carried out by various methods Data Mining. Different methods of induction and trductive inference can extract knowledge from data, for example, ID3, ACLS, CART and their different versions focused on data processing with different properties – quantitative, qualitative, dynamically changing etc. [15].

## **10. Ontologies in the Personalized Search**

User interaction IIS elements in the most general form can be divided into the following groups:

- IIS user interface;
- Proposed to user content;
- Repository where IIS accumulates its experience of interaction with the user;
- IIS model of information interaction with the user that provides structured representation of knowledge

about key elements of IIS required for satisfaction of user needs.

All these components can use ontological knowledge representation for personalization – as for internal IIS knowledge that are laid in it in under the development, and for the external knowledge that IIS imports from the open environment or acquires from it's own experience of interaction with users.

Targeting on specific user and adaptation to his personal capabilities and needs can largely improve the IIS efficiency, so it is advisable to use the user information model at all of these components and improve it by learning. Intelligent adaptive interface is only one component of IIS personification. It relates primarily to the choice of methods of information presentation for user. Another important component of personification is the selection of interesting and useful for user IR and the choice of means and methods of these IR processing. External dynamically updated knowledge in modern applications – especially in Web-oriented – is an obvious necessity of today.

Ontologies are widely used for these purposes. But their use has both advantages and disadvantages. Ontologies should be used in IIS with a structure that changes or in cases where the developer doesn't know in advance not only the structure of knowledge but also the properties of this structure, and therefore by a priori information about them can not select some more simple means of knowledge representation,

Benefits that ontologies provide are well known and need no further justification:

- Interoperability and reusability in various applications;
- Theoretical basis in the form of descriptive logics that provides guarantees of deductibility and solvability;
- A large number of previously created ontologies;
- Numerous languages, standards and software for processing of ontologies.

Disadvantages of ontologies are less obvious but we must consider them in creating of real applications:

- In general ontologies are too complex in handling and require a lot of time and large computational resources for check of deductibility of some fact on the ontology;
- Considerable part of real ontologies (as opposed to "toy" examples) are too cumbersome for their processing (changes or expansion) by user who doesn't develop it (even if user has sufficient knowledge at this domain);
- Scalability problem – if the number of terms in the ontology and relations between them increases then the processing time increases too so that no longer meets the needs of real problems.

Furthermore, it should be noted that most studies related to the ontological knowledge representation their full expressive possibilities just are not used, but because their processing is still available, it unnecessarily increased the complexity of the calculations. At the same time is difficult to predict what kind of domain knowledge will be essential for particular ontology-based IIS creating. Therefore, it seems appropriate to use a two-level ontological model of personalized knowledge.

The *first* level is domain model where domain knowledge are represented by domain ontology  $O_{\text{domain}} = \langle X_{\text{domain}}, R_{\text{domain}}, F_{\text{domain}} \rangle$ , where  $X$  is the set of concepts,  $R$  – the set of relationships between concepts, and  $F$  – interpretation functions of concepts from  $X$  and relations from  $R$ .  $O$  doesn't imposed restrictions beyond descriptive logics.

The *second* level is a task model. It uses only a small subset of the domain model concepts and relations. Task model is a projection of the user problem on domain ontology. For example, some relations of domain ontology are discarded or replaced by some other like "there is some connection between terms" or terms from fixed subset and associated with them are left, only hierarchical links are left etc. A special case of this task model is ontological thesaurus that is used for the semantic search.

This two-level approach forms much simpler in handling ontology, which can already be more clearly attributed to some particular descriptive logic.

## 11. Personalized SSS Architecture

In the most general form architecture of SSS that provides a personalized interaction with user in the Web information space describes the links between user information needs [16] and those information objects and information resources that SSS offers to user for solving of this problem (fig.2) .

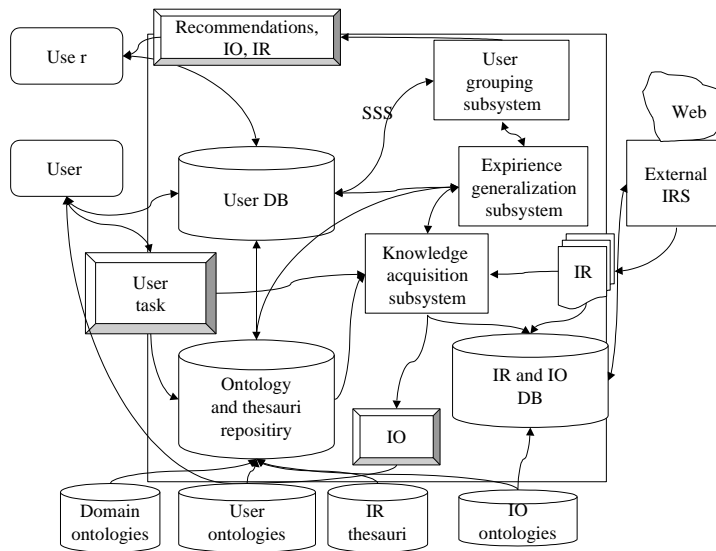


Fig.2. Generalized Architecture of Semantic Search System

SSS collects information about users in its database, and then on base of its analysis specific rules intended to account personal characteristics and needs of the user are formed.

An important aspect of the proposed above architecture is the use of external knowledge sources – the repository of ontologies and thesauri, and other IIS that forms personalized knowledge about users in interoperable form. This approach is used in the development knowledge-oriented search and recommending system MAIPS [17].

This approach to personalification of semantic technologies of IIS design provides more complete account of the individual user characteristics , and the use of ontology allows knowledge reuse.

## 12. Individual readability criteria in SSS

One of important factors of search personalization is a degree of clarity of the received information to the user, its readability [18]. The readability is a comprehensibility measure of information contained in the text that depends on how easy to understand the text meaning [19]. Text readability also depends on factors not related to text language, such as font type, size and color. It is necessary to distinguish between formal readability of text  $R_{form}(I)$  that is a function only of the IR I parameters, and individual readability of text  $R_{ind}(I, u)$  that depends both on characteristics of IR I and on the properties of the reader u.

The readability can be seen as a combination of text and reader performance characteristics. In general, it

depends on the of the text properties, such as print quality (in the electronic version of IR – from font legibility, the choice of text and background colors); availability illustrations; vocabulary; conceptual difficulties; syntax; structural organization of IR; and reader properties, such as reader abilities and his interest in the material presented in the IR.

Reader abilities include his educational level, expertise in subject domain, IR language proficiency and so on. For example, the text translated into the local reader's language is perceived him quite easily then the original one that is read with considerable loss of contents. Today IRS allow users to specify the desired language of IR but don't allow them to specify their own level of competence in the chosen language.

The reader interest is an important factor, but, unfortunately, it is poorly formalized. Internal motivation of reader and the domain of his interests are very important for information retrieval, so we need in their formalization for selection of interesting for specific user IR. External user motivation can be defined in more simple way, in particular, by indication of the IR utility for task solving, performing of specified work, passing tests or learning courses. For example, IR which title contains the word "manual" is more preferable to user (*ceteris paribus*) for learning then IR which title contains the word "summary" or "abstract".

Vocabulary of text is considered the best indicator of text readability. The average length of words and sentences (in letters or symbols) are statistical factors that are often used in various formulas that are designed to readability assess. These settings are easy to quantify and are suitable for automatic evaluation. Unfortunately, most existing IRS that propose to user a variety of other formal settings for IR founded by his request IR don't derive any information about these important properties. We consider using the evaluation of readability for the personalization of IRS.

### 13. Methods of Estimation of Natural Language Text Readability

There are a number of methods for estimation of the natural language text complexity. These methods examine the length of sentences and words in them, but don't consider the sentence structure and word order. The formulas used for calculation vary for different languages (for example, in English words are shorter on average than in Ukrainian or Russian – by using of endings). Counting of the syllable number of in different languages is carried out in different ways and based on the pronunciation of words. For example, in English (as

Table 1. Parameters used in Readability Formulas

	b	f	k	l	L	p	P	s	w	x
Gunning Fog index			+	+				+	+	
Flesch Formula		+	+			+		+	+	
Flesch-Kincaid formula						+			+	
Fry readability Schedule		+	+			+		+	+	
Coleman-Liau Index			+				+	+	+	+
Raygor readability evaluation		+	+			+		+	+	
Powers-Sumner-Kearl formula		+	+			+		+	+	
McLaulin formula "SMOG"			+	+	+					
FORCAST formula	+			+						

opposed to Ukrainian and Russian) the number of syllables may not coincide with the number of vowels in this word. For example, the word "surface" has two syllables and the word "surfaces" – three. Abbreviations commonly evaluated by the rule "the number of letters coincides with the number of syllables" and standard abbreviations (km, kg, etc.) are valued at one syllable. The headings and subheadings usually consist of not complete sentences, and that's why they are ignored. Because the procedure for formula processing is ambiguous, they usually also are ignored too.

Where

- b – number of monosyllabic words in text;

- f – total number of syllables in a text;
- k – total number of words in the text;
- l – average number of "long" words in the text;
- L – number of "long" words in the text;
- p – average word length (in syllables);
- P – average word length (in characters).
- s – number of sentences in the text;
- w – average length of sentence;
- x – number of characters in the text;

Table 1 describes parameters used in the most common methods for determining the text readability. Some of these parameters (b, f, l, L, p) depend of natural language, other (k, P, s, w, x) are independent of language but all formulas use different coefficients for different natural languages. The constants used in these formulas for estimation of the English text complexity in most cases do not coincide with those that used for the estimation of the Ukrainian and Russian text. However these coefficients are not important for ranging texts in the same language by their readability. These methods are useful for understanding of principles of the readability level determining.

*Gunning Fog index.* One of the most popular methods for assessing readability of text information is fog index designed in 1952 by the American scientist R.Gunning [20]. It allows determining the minimum age of the reader who can understand the text. This index is used for evaluation of text oriented on wide audience, and uses the average statistical data for estimation of educational level and intelligence of readers. Fog index measures the difficulty of reading on base of the average length of sentences and percent of words consisting of three or more syllables. The fog index is higher for the more difficult texts. To assess the selected text at least two random text fragment that contain approximately 100 words are selected:

$F_{\text{Gunning}} = \sum_{i=1}^s w_i / s + 0.4 * \sum_{i=1}^s l_i / s = w + 0.4 * l$ . This formula is used to understand whether text clear to particular audience. For example, for clear to majority English texts the fog index is below 12.

*Flesch Formula.* The formula developed R.Flesch [21] allows to set the readability of English text and an approximate education level necessary to understand it. It calculates not only the number of words in a sentence but also the number of syllables in each word. Test FRES (Flesch Reading Ease Score) has widespread after the adoption of legislative regulations for a number of US states that requires that the text of the insurance contract has be understandable to persons with secondary education. FRES processes the text with size approximately 100 words. Abbreviations, symbols and words written with a hyphen are processed as separate words. The level of readability is estimated by the formula:

$$F_{\text{Flesch}} = 206,835 - (k/s * 1,015 + f/k * 84,6) = 206,835 - (w * 1,015 + p * 84,6)$$

*Flesch-Kincaid formula.* Flesch-Kincaid formula converts Flesch quantitative index to the level of education necessary for text understanding. It is used by teachers, librarians etc. to choose recommended textbooks and tutorials:  $F_{\text{Flesch-Kincaid\_grade}} = w * 0,39 + p * 11,8 - 15,59$ .

*Fry readability Schedule.* This method processes, as in previous criteria, the text with size about 100 words. A graph of dependence between the number of sentences and syllables in the processed text is built [22]. The point on the chart that corresponds to values of w and p determines the Fry level of readability.

*Coleman-Liau Index.* This index is designed for evaluation of the text readability. As opposed to other readability evaluation (except ARI), it is based on the average number of characters in the words instead of the average number of syllables in the words [23]. Coleman-Liau Index is calculated using the formula:

$$F_{\text{Coleman-Liau}} = 5,89 * x/k + 30 * s/k - 15,8$$

Raygor readability evaluation. Raygor readability graph [24] is similar to Fry's readability graph. Estimation parameters are the average length of sentence and the average number of characters in words. For the estimated English text these characteristics are calculated, and at the intersection of corresponding lines on the graph turns age of reader who can understand this text.

Powers-Sumner-Kearl formula. Powers-Sumner-Kearl formula [25] is used only for texts intended for children. As in most other such assessments, it analyzes the text size of about 100 words and determines the age and educational level of the reader, which can understand the proposed text:  $F_{\text{PowersSumner-Kearl\_grade}} = w * 0,0778 + f * 0,0455 - 2,2029$ .

McLaulin formula "SMOG". These formulas analyze text with 30 sentences [26]. SMOG counts the number of words with three or more syllables  $F_{\text{SMOG\_grade}} = \sqrt{L/s} * \sqrt{1} + 3$ . McLaulin formula usually gives better grades compared to similar tests because McLaulin test is oriented on prediction of absolute (complete) text understanding.

FORCAST formula. This formula was developed for the assessment of technical documentation clarity for the US Army [27] and therefore it is targeted only on adult readers. The analyzed text may contain not complete sentences  $F_{\text{FORCAST\_grade}} = 20 - k * 0,0667 * b$ .

#### **14. Comparison of Text Readability Criteria**

The above examples show the diversity of approaches to assessing of the readability level. Despite the fact that in the above formulas used different parameters (average word length in syllables or symbols, the number of polysyllabic or monosyllabic words) and various coefficients to calculate age and educational level of the reader who can understand the analyzed text, but assessments received by them are similar in general, and their differences are usually caused by author's style features.

Analysis of research work shows that most readers prefer texts slightly above their level. Tests can determine the age limit of understanding that the average reader of this age "on the verge" of understanding of information contained in the text. Most estimates are based on the fact that the text is considered to be clear to the reader, if this reader is able to answer correctly on 50% of questions by the text, i.e. he understands half information in the text. It should be noted that these tests are designed to assess some written text but not for recommendations for their writing. In addition, they do not account for order of words which also influences largely on the text clarity.

Moreover, all these tests give the average estimates and don't take into account that one person may have a different level of education in various fields (even pupil of secondary school under the standard program may have different marks on various subjects). Therefore, if we try to use such assessments for the Web information retrieval it seems appropriate to apply different evaluations for the same user for different domains.

#### **15. Semantization of Readability Criteria and their Use in SSS**

On the one hand, the competent reader is not uninteresting in too primitive text. On the other hand, the unskilled readers (e.g., child) don't want to analyze the text that he can not understand freely. But the same information can be presented by more difficult or simpler language. For example, we can compare the two definitions: "Ontology is a formal specification of shared conceptualization that occurs in some context of subject domain. It is a structure of hierarchically or polihierarchically related semantic categories, allowing self-description of used their categories by related topics" and "Ontology is a formal description of domain. It defines the general terms and relations between them". Obviously the second definition is much clearer because it hasn't special terms and uses shorter words and sentences.

But the text readability rating is much more useful if it is estimated not "in general" but from viewpoint of particular reader: in addition to the text characteristics it should take into account the characteristics of the reader – his experience, skills and motivation. For example, if the reader is looking for reference books from

his professional activities, then their complexity must correspond with professional level of reader or even be higher, and while reader needs in information help in some non-core for him field than some easier more easily written document satisfies this need.

Therefore we need in some means suitable for automated processing to describe the areas where user has a more profound knowledge than others do.

It is advisable to utilize a particular case ontology – thesaurus – for modeling of the user knowledge about search domain. As a rule thesaurus  $T$  is defined as a dictionary containing the lexical units with an explicit indication of semantic connections between them. In this case, thesaurus defines the user vocabulary and those terms which user can operate free.

Maximum amount of semantic information the user gets if his thesaurus matches with IR meaning – if the information is clear to user and contains some knowledge that lacks in user knowledge base. If SSS offers to user some IR that doesn't contain known to user terms than user can't acquire any information from this IR. If user already knows all IR content (terms and relationships between them) than no changes in his knowledge would be generated by this IR, and thus the amount of semantic information of such IR for this user is also equal to zero.

For filtering from the external IRS results those IR that are pertinent to information needs of users we have first to create a thesaurus of interesting for user domain – a set of terms familiar to users IRS, then we can compare it with the IR thesauri. We assume that the thesaurus of domain is a set of known to user terms. These terms are contained in the IR found earlier on user requests and recognized him as related to that domain.

Domain thesaurus allows to calculate the individual readability for each of the found natural language IR. In this analysis the personalization of readability is provided by the following procedure: every long word from those IR are considered understandable to this user without additional intellectual effort if this word is contained in the user thesaurus and that's why all these word are taken into account in the formulas as simple monosyllabic words.

For processing in readability formulas all words of thesaurus are replaced by simple one-syllable three-letter word that doesn't exist in any language:

If word  $a \in IR_j, j = \overline{1, n}$  and  $a \in T_{user}$  then replace (a, "tat")

For analysis of a large number of IR we use a simplified algorithm for constructing of IR thesauri: at first we build the complete list of words used in IR, and then discard stop words from it (stop words are contained in a specially designed user list).

This algorithm is used for IR that have not any metadescriptions. If IR has metadescriptions (in RDF or OWL formats) than thesaurus terms and relations between them are acquired from this metadescription and are added to IR thesaurus built on IR content.

Such criteria of individual readability we propose to use for personification of the semantic search of the Web resources in different ways:

- User himself sets the desired for him 1) text readability level – formal text readability  $R_{form}(U)$  – for a particular query, and SSS then filters the results of this query by proximity of their readability values to this level; 2) text readability level in this domain – individual text readability  $R_{ind}(U, Th)$  – for a particular query, and SSS then filters the results of this query by proximity of their readability values to this level;
- User himself sets the desired for him interval of formal text readability level –  $R_{form\_min}(U)$  and  $R_{form\_max}(U)$  - for a particular query, and SSS then filters the results of this query from this interval of their readability values;
- User himself sets the desired for him in this domain interval of individual text readability  $R_{ind\_min}(U, Th)$

- and  $R_{\text{ind\_max}}(U, Th)$ - for a particular query, and SSS then filters the results of this query by proximity of their readability values to this level;
- Formal text readability level corresponding to user education in a specific field is determined automatically by analyzing of the domain relevant texts that user marks as understandable and interesting enough for him  $R_{\text{form}}(U) = \sum_{j=1}^d R_{\text{form}}(I_j) / d \quad j = \overline{1, d}$ ;
- Individual text readability level corresponding to user education in a specific field is determined automatically by analyzing of the domain relevant texts that user marks as understandable and interesting enough for him. Task thesaurus is used for separation of the "long" words known to user from all other long words that complicate the understanding of the text. This thesaurus is generated by relevant IR and domain ontologies as stated above: all IR words from thesaurus are processed by readability criteria as short ones independently of their actual length  $R_{\text{ind}}(U) = \sum_{j=1}^d R_{\text{ind}}(I_j) / d \quad j = \overline{1, d}$ .

All these variants of text readability can be used for the entire IR and or only for it's annotation that is offered to user in response to a search engine query. In the first case the result is much more reliable but it requires the downloading of all found IR.

We consider a lot of readability criteria that analyze different parameters. There is an open question that of them is more appropriate for search task in different domains. Therefore we proposed the following solution – the user selects several relevant IR that he groups in identical sets by readability, and they readability criteria are calculated. The criterion that gives the closest estimates for the same type of IR and differentiates IR from different sets is the best for this user. This approach caused by differences in the text perception of different users: some people have difficulties with long sentences (bad short-term memory but high scholarship) and others – with complex terms (good short-term memory with low scholarship).

Thus, the search of the "best for user" IR includes several stages:

1. User forms  $X$  – the set of relevant to request domain IR with appropriate to him level of readability;
2. User forms request – the set of keywords  $K$  that is sends to external information retrieval system (IRS);
3. This IRS returns  $A$  – initial set of IR relevant to request  $K$  (we use external IRS as a retrieval preprocessor and all search problems that can be solved without semantics are not considered in this work). If this set is not empty then go to the next step;
4. SSS forms user thesaurus  $T$  by IR from  $A_1$  (selected by user set of IR relevant to task) and by domain ontology  $O$  (selected by user from ontology repository);
5. SSS calculate readability ratings for each IR from  $X$  and  $A$ ,  $a_k \in A, k = \overline{1, m}$  where such parameters are took into account:

- Number of user thesaurus terms  $T$  contained in the text (with the help of lexical ontology or other linguistic KB like WordNet) and number of related with them terms in the ontology  $O$ ;
  - Average length of the sentences and average number of syllables in the words of IR from  $X$  (without terms from  $T$ );
  - Average length of sentences and average number of syllables in a sentence of IR from  $A$  (without terms from  $T$ );
6. SSS calculates readability values of IR from  $A$  and  $X$ ;
  7. SSS finds IR from  $A$  nearest by these characteristics to IR from  $H$ .

If user prefers to analyze only the annotations of IR returned by external IRS then it is necessary to take into



account a number of annotation features. Annotations of IR usually are much shorter than one hundred of words, and often contain non-complete sentences. Thus, most of the above criteria can not be applied for their analysis. If the annotations don't contain complete sentences, it is impossible to estimate the average length of sentence but for simplicity of calculation we can assume that all analyzed fragment is one sentence. In addition, annotations can contain information in different languages, and the coefficients for calculating the readability level of for these languages differ. Also, different natural languages have different rules for splitting words into syllables. Therefore in this analysis we propose to take into account the average length of words expressed in symbols only. The abstract often contains authors' names with initials (one or two letters with a point after them). If we take initials into consideration at separate words then the average length of words for such fragment will be significantly lower than for the same fragment without initials. Therefore, we propose to reject initials in analysis of abstracts.

We propose to use the following personalized readability criterion for annotations in English that is constructed as a modification of the Coleman-Liau index and allows individual user level of proficiency in particular domain:  $F_{\text{annot}} = 5,89 * \frac{x}{k} + 30 * k - 15,8$ , where  $x$  – number of characters in the character string  $A_{\text{mod}}$  received from annotation  $A$  after the rejection of initials and replacing of thesaurus words by the "tat";  $k$  – total number of words in  $A_{\text{mod}}$ .

Thus, combining user readability criteria with formalization of his personal skills (by task thesauri) in SSS provides differentiated offering of IR with various complexity that match with real information needs and capabilities of the user.

## 16. Realization of Ontological Model of Personified Semantic Search in MAIPS

Each of the items of information model of recommending retrieval system “MAIPS” (MAIPS is an advanced by collaborative facilities and adaptive user interface SSS) is a class of ontology  $O_{\text{MAIPS}} = \langle T_{\text{MAIPS}}, R_{\text{MAIPS}}, \emptyset \rangle$ , and information about their values is stores as instances of these classes.

Information about these instances of classes from the set  $T_{\text{MAIPS}}$  is accumulated during the process of interaction with MAIPS users who try to satisfy their information needs.

Information about users is the basis for personalization of semantic search and creation of recommendations. Preparation and analysis of information about users can make information retrieval more efficient. Information model of user represents such features as:

- Personal user preferences, his knowledge and ability to perceive information;
- User information needs, specifics of the current task and of information objects that are the result of retrieval procedure ;
- Experience of user interaction with MAIPS;
- Information describing the user beliefs about utility of experience of other MAIPS users and feasibility of collaborative search [28].

A significant part of this information user inputs himself during the registration: more detailed user answers provide more pertinent search results. For example, ontology individual of class “user” can contain information about place of residence, education, work status, age, language skills and so on. Knowledge on areas of user interest and areas of user competencies can be imported from external sources proposed by user (social networks, Wiki-resources, personal pages, and external IR of natural ontologies) or found in the web by specialized services (e.g., service of semantic analysis of selected IR can be used to replenish of the knowledge of areas of interest, and specialized services for rapid testing can provide the physiological characteristics of the user). Some information about user competencies can be imported from other

intelligent ontology-based applications [29].

Knowledge about user is replenished during his interaction with SSS. Some user information can be entered to his profile automatically: every time if user selects some domain ontology, processes (creates modifies) the task thesaurus then thesaurus or executes the created earlier query then information about it is recorded to his personal data. In addition, the user can himself make some changes in his personal profile for more correct representation of his information needs: for example, he can specify other user whose search activity is interesting for him on base of ontological model of the semantic search [30].

As we consider above there is important for search personalization to evaluate and process in retrieval process certain physiological user properties that have an influence on his ability to perceive information. For example, the widespread classification that divides individuals by speed of excitement and inhibition provides classification of people on four personality types – choleric, sanguine, phlegmatic and melancholic. Another well-known classification divides of people on extroverts and introverts. Sociopsychophysiological study provided a number of the computer systems used to produce individual adaptation sociopsychophysiological portrait of people in different areas of life activities and consider the problem of creating comfortable for user individual adaptive information space. There are many psychological tests to clearly classify personality (e.g. Eysenck test, MMPI).

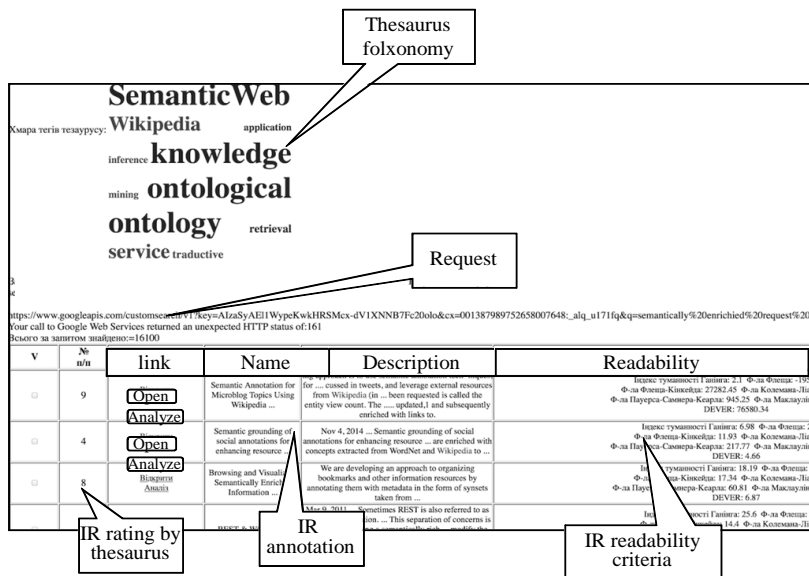


Fig.3. Search results of MAIPS with readability indexes of IR annotations

In most of other SSS these properties are not used therefore we consider hear in more details how they are used in MAIPS. MAIPS user can directly enter the following information (if he knows it) or execute some rapid tests. These user properties and based on them user classifications are used for traductive acquisition of mutual properties of IR (IR information representation and IR content) that satisfy some categories of users. In addition, the MAIPS user can assess what IR is quite understandable and useful to him by readability index (Fig.3). To do this, he can either explicitly specify it (e.g., documents oriented on people with secondary education), or specify the IR that meet and do not meet his requirements, and then the system will automatically calculate his personified readability index for this domain by previously used user ontologies and thesauri.

## 17. Summary

Modern intelligent applications provide the adaptive user interface and personalize the content. For semantic search systems these possibilities are associated with the accumulation and generalization of personal user information, his properties, interests and ability to information processing.

Personification of search mechanism allows to provide the user such information resources and objects that are convenient to his perception, that match exactly with his the current information needs and can proactively enhance effectiveness of his work.

An important aspect of personalization deals with user physiological features associated with his ability to information perception. These features can be obtained directly from the user (by tests and questionnaires), be acquired by analysis of his actions or be imported from other intelligent applications.

Personification of semantic search also provides individual assessment not only of the presentation of information resources but also of their content. Use these personalized (by task thesauri) readability criteria of natural language texts to characterize the ability of individual users to the perception of information allows proposing to user more understandable and usable information.

## References

- [1] Lawrence, S. Context in the Web Search. – <http://citeseer.nj.nec.com/lawrence00context.html>.
- [2] Gladun, A., Rogushina, J., Valencia-García, R., Martínez-Béjar, R. (2013) Semantics-driven modelling of user preferences for information retrieval in the biomedical domain. *Informatics for health and social care*, 2, 150-170.
- [3] Gladun, A., and Rogushina, J. (2007) Formalization of Search Context on Base of Ontologies and Multilinguistic Thesauruses. *Computing*, v.6, issue 3, 16-22.
- [4] Gladun, A., Rogushina, J. (2009) Use of Semantic Web technologies in design of informational retrieval systems. *Building and Environment*, Nova Scientific Publishing: 89-103.
- [5] Fazzingaa B., Lukasiewicz T. (2010) Semantic search on the Web. *Semantic Web – Interoperability, Usability, Applicability*: 1-7.
- [6] Corby O., Dieng-Kuntz R., Faron-Zucker C. (2004) Querying the Semantic Web with Corese search engine. *Proc. ECAI-2004*, IOS Press: 705-709.
- [7] Finin T. W., Ding L., Pan R., Joshi A., Kolari P., Java A., Peng Y. (2005) Swoogle: Searching for knowledge on the Semantic Web. *Proc. AAAI-2005*: 1682-1683.
- [8] Lei Y., Uren V. S., Motta E. (2006) SemSearch: A search engine for the Semantic Web. *Proc. EKAW-2006*: 238-245.
- [9] Tran T., Cimiano P., Rudolph S., Studer R. (2007) Ontology- based interpretation of keywords for semantic search. *Proc. ISWC/ASWC-2007*:523-536.
- [10] Damljanovic D., Agatonovic M., Cunningham H. (2010) Natural language interface to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction. *Proc. ESWC-2010*, Part I: 106-120.
- [11] Fernandez M., Lopez V., Sabou M., Uren V. S., Vallet D., Motta E., Castells P. (2008) Semantic search meets the Web. *Proc. ICSC-2008*: 253-260.
- [12] Mangol C. (2007) A survey and classification of semantic search approaches. *International Journal of Metadata, Semantics and Ontologies*, N 2(1):P.23-34.
- [13] Types of psychological tests for psychologists, mental health professionals and educators. – <http://www.assessmentpsychology.com/tests.htm>
- [14] Murphy, K. R., Davidshofer, C. O. (1998) Psychological testing. Principles, and Applications, Englewood Cliffs. – <http://tocs.ub.uni-mainz.de/pdfs/058927786.pdf>.

- [15] Quinlan, J.R. (1979) Discovery rules from large collections of examples: a case study. *Expert Systems in the Microelectronic Age*: 87-102.
- [16] Gladun, A., Rogushina, J. (2012) Use of Semantic Web Technologies and Multilinguistic Thesauri for Knowledge-Based Access to Biomedical Resources. *International Journal of Intelligent Systems and Applications*, №1: 11-20. – <http://www.mecs-press.org/ijisa/ijisa-v4-n1/IJISA-V4-N1-2.pdf>.
- [17] Rogushina, J., Gladun, A. (2006) Ontological Approach to Domain Knowledge Representation for Informational Retrieval in Multiagent Systems. *Information Theories & Applications*, V.13, N.4: 354-362.
- [18] Readability. – <https://en.wikipedia.org/wiki/Readability>.
- [19] Bormuth, J. R. (1967) Development of Readability Analysis. – <http://files.eric.ed.gov/fulltext/ED029166.pdf>.
- [20] Miles, T. H. (1990) The fog index: a practical readability scale. *Critical Thinking and Writing for Science and Technology*: 280-284. – <http://www.as.wvu.edu/~tmiles/fog.html>.
- [21] Flesch Reading Ease Readability Formula. – <http://oleandersolutions.com/fleschreadingease.html>.
- [22] Long, A. Calculating Reading Level. Tameri Guide for Writers. – [www.tameri.com/edit/levels.html](http://www.tameri.com/edit/levels.html).
- [23] Coleman-Liau Index [http://en.wikipedia.org/wiki/Coleman-Liau\\_Index](http://en.wikipedia.org/wiki/Coleman-Liau_Index).
- [24] Raygor Readability Estimate – [http://en.wikipedia.org/wiki/Raygor\\_Estimate\\_Graph](http://en.wikipedia.org/wiki/Raygor_Estimate_Graph).
- [25] Powers, R. D., Sumner, W. A., Kearsley, B. E. (1958) A recalculation of four adult readability formulas. *Journal of Educational Psychology*, 49.2: P.99-105.
- [26] McLaughlin, H. (1969) SMOG grading – a new readability formula. *Journal of Reading*, N 22: 639-646.
- [27] Sticht, T. G. (1973) Research towards the design, development and evaluation of a job-functional literacy training program for the US Army, *Literacy Discussion*, N 4: 339-369.
- [28] Ricci, F., Rokach, L., Shapira, B., Kantor P. *Recommender Systems Handbook*. – Springer, 2011. – 842 p.
- [29] Rogushina, J., Gladun, A. (2012) Ontology-based competency analyses in new research domains *Journal of Computing and Information Technology*, 20 (4): 277-291
- [30] Rogushina J. (2016) Use of the Ontological Model for Personification of the Semantic Search. *International Journal of Mathematical Sciences and Computing(IJMSC)*, Vol. 2, N 1: 1-15.

### Authors' Profiles



**Dr. Julia Rogushina** received her PhD degree in computer sciences in 1994 in the Institute of Cybernetics, Kiev, Ukraine. She is currently working as a senior researcher at the Institute of Software Systems, National Academy of Sciences of Ukraine. Her scientific interests deal with knowledge management, ontological analysis and semantic search, theory of software agents behavior, inductive knowledge acquisition, ontological analysis, the Semantic Web technologies. She is an author of “Agent Technologies”, “Semantic Technologies: Principles and Practices”, “Semantic Search for Web”.

**How to cite this paper:** Rogushina J., "Means of the Semantic Search Personification on base of Ontological Approach", *International Journal of Mathematical Sciences and Computing(IJMSC)*, Vol.2, No.3, pp.1-20, 2016.DOI: 10.5815/ijmsc.2016.03.01