

THE ONTOLOGY SUPPORTED INTELLIGENT SYSTEM FOR EXPERIMENT SEARCH IN THE SCIENTIFIC RESEARCH CENTER

Vladimir Cvjetković¹, Marija Đokić¹, Branko Arsić¹ and Milena Ćurčić²

¹Department of Mathematics and Informatics, ²Department of Biology and Ecology,
Faculty of science, University of Kragujevac,
Radoja Domanovića 12, 34000 Kragujevac, Republic of Serbia
E-mail: vladimir@kg.ac.rs

(Received January 24, 2014)

ABSTRACT. Ontologies and corresponding knowledge bases can be quite successfully used for many tasks that rely on domain knowledge and semantic structures, which should be available for machine processing and sharing. Using SPARQL queries for retrieval of required elements from ontologies and knowledge bases, can significantly simplify modeling of arbitrary structures of concepts and data, and implementation of required functionalities. This paper describes developed ontology for support of Research Centre for testing of active substances that conducts scientific experiments. According to created ontology corresponding knowledge base was made and populated with real experimental data. Developed ontology and knowledge base are directly used for an intelligent system of experiment search which is based on many criteria from ontology. Proposed system gets the desired search result, which is actually an experiment in the form of a written report. Presented solution and implementation are very flexible and adaptable, and can be used as kind of a template by similar information system dealing with biological or similar complex system.

Key words: Ontology, knowledge base, bio-experiments, active substances, SPARQL.

INTRODUCTION

The main role of information system is to support the operation of some real system which is mainly some enterprise or organization. Real system that is supported in this paper is the Research Center (RC) [3] for testing of active substances. Active substances are candidates for medicaments that are tested in laboratory, prior to being approved or not, for medical treatments. The RC is also the leader of the large Project [2] financed by the Ministry that consists of many institutions, departments, equipment and staff working on the project.

The subject of various analysis that are carried out at the RC includes monitoring of *in vitro* effects of active substances in the cell lines of different origin, primarily cancer cell lines and primary cells isolated from different tissues. Tests include cytotoxic active substances in human cancer cell lines, while monitoring includes the type of cell death, the mechanisms of apoptosis, migration and angiogenesis and prooxidant-antioxidant mechanisms which underlie the regulation of these processes. Tests are based on protocols such as MTT cytotoxicity test, AO/EtBr staining of cells for examination of the type of cell death, Western blot technique for examining proteins, Multiplex PCR, etc.

The specific organization of the RC and its main field of work resulted in unique and complex experiment structure that was defined as hierarchical structure, coded in plain language on a paper by the expert, hard to be remembered, explained and adequately presented when needed. Experiment structure displays specific and complex relationships among various terms and concepts from the RC work area. Concepts are generally divided into levels, which are further branched at the sublevels. Some sublevels are common to different upper levels. Kinds of these relationships were also the subject of research, determined according to how RC functions. The structure is expected to further expand in the future, so it requires flexible modeling and representation that can be easily updated.

In order for this experiment structure to be used or referenced by the staff, the system for presenting and search of experiments was planned. First possible solution was relational database. Relational databases are designed for organizing information that is easily categorized by common characteristics, and described by simple string or number data [11]. But, our complex experiment structure, with concepts on various levels which are related in specific way, cannot be categorized in this manner. Also, our structure is flexible and expandable, and increased storage and utilization of massive amounts of complex data have a tendency for implementation of more complex database schemas [11]. Therefore, the authors concluded that complex hierarchical structure of the experiment is not suitable for modeling and presenting in relational database due to the nature and relationships of terms and concepts at various levels.

Authors turned to technologies developed for Semantic Web [1], as an adequate choice. Using of ontology [4], as the core component of Semantic Web, for modeling, representing and visualization of an arbitrary semantic structure of experiment, justified the expectations. Semantic Web and ontology allow representation of complex semantic structure and enable data to be found, shared and combined in an easy way.

In order to support laboratory staff to quickly reference and use complex experiment structure, PIBAS (Preclinical Investigation of Bio Active Substances) ontology for modeling complex experimental structure of the RC was proposed in this paper. The knowledge base which contains experiments from the real world was organized and classified according to ontology.

The main goal of this paper is creation of the system for search of experiments which is based on developed ontology and knowledge base. The proposed system allows laboratory staff to search for experiments in a simple and easy way, to obtain information about the active substances, model systems, methods and protocols used in the experiments and to achieve conclusions for directing the future research. Proposed system gets the desired search result, which is actually the result of an experiment in the form of a written report. For retrieval of experimental results from ontology, the SPARQL query language [5] based software on web server is used in the background.

This paper is organized in the following way: The second section describes the experiment structure according to RC specification. The third section describes a motivation for ontological modeling of experiment structure. The fourth section gives an overview of the literature in existing field of work. Fifth section describes ontology design principles, visualization of ontology, and ontological knowledge base populated with real experiments data. The sixth section describes architecture and functioning principles of proposed system search. One part of this section is dedicated to dynamic forming of SPARQL queries used for search. Seventh section discusses results and benefits of proposed system. Conclusion contains short survey of paper key points and directions for future work.

EXPERIMENT STRUCTURE

Various complex tests and experiments are performed at RC. Each experiment is characterized by active substances that are examined, model systems used for testing and

protocols which are standard experimental procedures applied in the experiment. In general, well designed experiment performed at RC enables detection of effects that are of biological and scientific importance.

Experiments performed at RC include monitoring of *in vitro* effects of active substances in the cell lines of various origin (primarily cancer cell lines) and primary cells isolated from different tissues. This field of work results in complex experiment structure which is presented in Fig. 1. The structure is coded by expert in plain language, and poses multilevel hierarchical structure.

<p>EXPERIMENTS</p> <p>A) TYPE OF EXPERIMENTS</p> <p>A. MODEL SYSTEMS</p> <p>I. ANIMALS</p> <p>1. RATS</p> <p>2. FISH</p> <p>II. CELL LINES</p> <p>1. PRIMARY</p> <p>2. CANCER</p> <p>III. PATIENTS</p> <p>IV.</p> <p>B. METABOLIC SYSTEMS</p> <p>I. ENERGY METABOLISM</p> <p>II. OXIDATIVE STRESS</p> <p>1. REACTIVE SPECIES</p> <p>2. OXIDATIVE DAMAGE</p> <p>3. ANTIOXIDATIVE SYSTEM</p> <p>III. HEMATOLOGY</p> <p>IV. ENDOCRINE SYSTEM</p> <p>V. IMMUNE SYSTEM</p> <p>C. TREATMENT</p> <p>I. <i>IN VIVO</i></p> <p>II. <i>IN VITRO</i></p> <p>D.</p> <p>B) THE AIM</p> <p>C) TYPE OF TREATMENTS</p> <p>A. ACTIVE SUBSTANCES</p> <p>B. DRUGS</p> <p>C. TREATMENT</p> <p>I. <i>IN VIVO</i></p> <p>II. <i>IN VITRO</i></p> <p>D. DOSES</p> <p>I. ACUTE</p> <p>II. CHRONIC</p> <p>E. PATHOLOGY</p> <p>D) METHODS</p> <p>A. MATERIAL</p> <p>B. MODEL SYSTEMS</p> <p>I. ANIMALS</p> <p>1. RATS</p> <p>2. FISH</p> <p>II. CELL LINES</p> <p>1. PRIMARY</p> <p>2. CANCER</p> <p>III. PATIENTS</p>	<p>IV.</p> <p>C. ANALYTICAL METHODS</p> <p>I. HEMATOLOGICAL PARAMETERS</p> <p>II. OXIDATIVE/ANTIOXIDATIVE METABOLISM</p> <p>1. REACTIVE OXYGEN SPECIES</p> <p>A. SUPEROXIDE ANION RADICAL</p> <p>B. HYDROGEN PEROXIDE</p> <p>2. REACTIVE NITROGEN SPECIES</p> <p>A. NITRIC OXIDE</p> <p>B. PEROXYNITRITE</p> <p>C. NITROXYL ANION</p> <p>3. OXIDATIVE DAMAGE PARAMETERS</p> <p>A. METHEMOGLOBIN</p> <p>B. HEINZ BODIES</p> <p>C. LIPID PEROXIDES</p> <p>4. NON-ENZYMATIC ANTIOXIDATIVE COMPONENTS</p> <p>A. REDUCED GLUTATHIONE</p> <p>B. OXIDIZED GLUTATHIONE</p> <p>C. VITAMIN C</p> <p>D. VITAMIN E</p> <p>5. ANTIOXIDATIVE ENZYMES</p> <p>A. SUPEROXIDE DISMUTASE</p> <p>B. CATALASE</p> <p>C. GLUTATHIONE PEROXIDASE</p> <p>D. GLUTATHIONE REDUCTASE</p> <p>E. GLUTATHIONE S TRANSFERASE</p> <p>6.</p> <p>III. ENERGY METABOLISM</p> <p>1. OXIDATIVE PHOSPHORYLATION PARAMETERS</p> <p>2. GLYCOLYSIS PARAMETERS</p> <p>IV.</p> <p>D. CELL CULTURE ASSAYS</p> <p>I. CELL ASSAYS</p> <p>II. MIT ASSAY</p> <p>III. IMMUNOFLUORESCENCE</p> <p>E. PROTEIN ANALYSIS ASSAYS</p> <p>I. PROTEIN EXTRACTION FROM TISSUES</p> <p>II. PROTEIN EXTRACTION FROM RBCs</p> <p>III. PROTEIN EXTRACTION FROM MITOCHONDRIA</p> <p>IV. SDS PAGE</p> <p>F. DNA AND RNA ANALYSIS ASSAYS</p> <p>I. RNA EXTRACTION FROM TISSUES</p> <p>II. RT-PCR</p> <p>E) RESULTS</p> <p>F) RESULTS AND COMMENTS</p>
---	---

Figure 1. Hierarchical multilevel structure of the Experiment.

The first level contains terms that generally correspond to concepts of each biological experiment. The explanations of these terms are given below:

- The type of experiment designates assigning subjects to conditions by the experimenter.
- *The aims* of research generate measurable data that can be tested, and contribute to gradual accumulation of human knowledge.
- Type of treatment specifies *in vivo* or *in vitro* application of active substances /drugs defined doses on experimental model systems.
- Methods are techniques for phenomena investigation, new knowledge acquiring, or correcting and integrating previous knowledge. It is based on gathering empirical and

measurable evidence subject to specific principles of reasoning and consisting in systematic observation, measurement, experiment, formulation, testing and modification of hypotheses.

- Result is the final consequence of a sequence of actions or events expressed qualitatively or quantitatively.

Further, these levels or terms are divided into sublevels or sub terms that are expected to be further expanded in future. Explanations of terms currently in the second levels are given in Table 1.

Table 1. Explanations of some experiment concepts.

Level	Sub level	Explanation
Type of experiments	Model systems	Biological systems (animals, cell lines, patients) that are extensively studied with the expectation that discoveries will provide insight into specific biological phenomena.
	Metabolic systems	The set of chemical reactions that happen in the cells of living organisms. These processes allow organisms to grow and reproduce, maintain their structures, and respond to their environments.
	Treatments	Treatment is <i>in vivo</i> or <i>in vitro</i> application of defined doses of active substances/drugs on experimental model systems.
Type of treatment	Active substance	Chemical or plant substance that affects the physiology, the function of the body of a human or animal.
	Drugs	A drug, broadly speaking, is any substance that, when absorbed into the body of a living organism, alters normal bodily function.
	Treatments	Treatment is <i>in vivo</i> or <i>in vitro</i> application of defined doses of active substances/drugs on experimentally model systems.
	Doses	Administration of tested active substances/drugs in experimentally defined amounts.
	Pathology	Pathology is the precise study and diagnosis of disease. Pathology addresses four components of disease: cause/etiology, mechanisms of development (pathogenesis), structural alterations of cells (morphologic changes), and the consequences of changes (clinical manifestations).
Methods	Material	Tools or apparatus for the performance of a given task. Also, chemicals used for experiment.
	Model systems	Biological systems (animals, cell lines, patients) that are extensively studied with the expectation that discoveries will provide insight into specific biological phenomena.
	Analytical Method	Techniques used to draw statistical inferences including multiple regression, path analysis, discriminate analysis and logistic analysis.
	Cell Culture Assay	Assay employed in cell culturing.
	Protein Analysis Assay	Assay employed in protein analysis.
	DNA and RNA Analysis Assay	Protocol employed in molecular biology methods.

Some terms in Table 1, such as “Model system”, appear more than once as sub terms of various terms – “Type of experiments” and “Methods”, as specified in Fig. 1.

MOTIVATION

Experiment structure printed on paper and shown in Fig. 1 displays specific relationships among various terms and concepts in accordance with experiments performed at the RC. Planning and designing experiments for active substance testing were done according to this experiment structure, which is often updated and extended, in order to reflect newly discovered relations among active substances before start of new experiment or results publishing. Having the experiment structure defined only on paper would favor further production of paper documents related to performed experiments, which would not be an adequate foundation for development of intelligent automated IT system.

The experiment structure should be presented in a simple and understandable way, allowing for easy modifications and changes in order to be used and referenced by the staff and other users. Simple and easy search for various information related to experiments is required, which can be used to obtain conclusions for directing the future research. To achieve that, the intelligent system for search of experiments which will enable users to get all the necessary information should be developed.

The idea of using a relational database to represent the complex structure of the experiment proved to be unsuitable due to the nature and relationships of terms and concepts on various levels. Relational database is not suitable for modelling of knowledge structure with classes, subclasses and various logical constraints expressing the characteristics of the real world system, while knowledge representation is straightforward and natural with semantic web languages, as they were developed exactly for that purpose.

The Semantic Web is a new generation of Web, which allows data to be both, human and machine readable and processable. The main purpose of the Semantic Web is driving the evolution of the current Web by enabling users to find, share, and combine information in a new way. [12]. Using of semantic Web technologies enable construction of a system for required custom search of experiments, that provide users with an easy and intuitive way of specifying various search criteria, for obtaining the relevant results. The complex experiment structure can be easily represented in the form of ontology. Ontology describes the semantic structure of some knowledge domain in terms of classes, properties, relationships among concepts and various logical constraints (union, restriction, enumeration, etc.). Using of SPARQL query language with developed ontology and ontology knowledge base would enable efficient retrieval and analysis of required important experiment information.

RELATED WORKS

With growing trend of biological data amount, efficient retrieval and analysis of this data becomes more and more important. Ontologies can represent the abstract knowledge required for data integration and analysis [17]. As the ontology engineering is a relatively new research field the well developed theory and technology for ontology construction still does not exist. Accordingly, many solutions have been proposed. For example, SOLDATOVA and KING [15] defined rules for bio-ontology development. They proposed some principles: explicitly list the principles of an ontology design, its constraints, along with definitions and axioms; keep separately domain-dependent and domain-independent knowledge, to provide efficient sharing and knowledge reuse; build ontologies so that they are purpose-independent. Given principles are discussed in detail when determining the construction of our ontology.

For any ontology to be of public value, it has to be widely disseminated and accepted by various experts. In this regard, the important standard is OBO Foundry (Open Biomedical

Ontologies) [13]. This ontology library contains interoperable reference ontologies in the biomedical domain and provides a set of principles for ontology development. The Basic Formal Ontology gives the top-level classes under which OBO Foundry ontologies should be built, while the Relation Ontology (RO) [14] provides the relations that should be used. Ontology for Biomedical Investigations (OBI) [9] is an integrated ontology for description of investigations in the field of biology and medicine. Mentioned ontologies are important examples of presenting relationships among concepts, which significantly influenced development of our ontology.

Very important part of natural sciences, especially biology, is to increase knowledge through experimental work. One of the ways to formalize knowledge is to define an explicit ontology. But as the experimental work often depends on particular circumstances, the general purpose ontology for scientific experiments currently does not exist.

The MGED Ontology (MO) [7] was designed to formalize the descriptors required by minimum information about a microarray experiment (MIAME) standard. This ontology aims to provide a conceptual structure for microarray experiment descriptions and annotation. A number of ontological developments related to MO also exist.

The HUPO PSI General Proteomics Standards and Mass Spectrometry working groups are building an ontology that will support proteomic experiments [19]. The metabolomics standards initiative (MSI) ontology working group is seeking to facilitate the consistent annotation of metabolomics experiments by developing ontology to help the scientific community to understand, interpret and integrate experiments.

Reviewing and comparing the concepts of mentioned ontologies, authors came to a conclusion that the structure of our ontology is quite specific, with somewhat more precise terminology. The common concepts between our ontology structure and given ontologies are rare, so that mentioned ontologies cannot serve as a solution or as a concept template.

Authors of paper [16] represent the most general elements of a common ontology (EXPO) for scientific experiments. They formalized generic knowledge about scientific experimental design, methodology and results representation. As all the sciences follow similar experimental principles, use related instruments and materials, execute and analyze experiments in similar ways, such a common ontology is feasible and desirable. The main aim of EXPO ontology is to abstract out the fundamental concepts in formalizing experiments that are domain independent. Generally, this ontology offers many benefits: makes scientific knowledge more explicit, contributes to error detection, promotes the interchange and reliability of experimental methods and conclusions, and removes redundancies in domain-specific ontologies. EXPO ontology contains some concepts that are similar with terms of our structure. However, these concepts are generally present in almost any experiment, such as the results, aim, and type of experiment. Other concepts of our structure are quite specific, in accordance to experiments at RC.

Although mentioned ontologies have important contributions to formalization of experiments in areas of biology, they are inappropriate as a template for a general ontology of experiments, as they are primarily oriented to specialized biomedical domains.

Bio ontologies today have a wide variety of uses. The most important is the representation of knowledge in a computer comprehensible way. To this end, many software solutions have been proposed such as BioPortal [8]. It is a virtual library of ontologies on the Semantic Web and a tool set enabling the community to access, discuss and improve ontologies. It contains over 50 ontologies from the biological and medical domains. Besides Web interface enabling researchers in cyberspace to locate these knowledge resources, BioPortal provides a suite of Web services, including ontology categorization, term search, graphical ontology visualization, and ontology version histories.

TOMLINSON *et al.* [20] developed web application for use by laboratory scientists, consisting of a browser-based interface and server-side components which provide an intuitive platform for capturing and sharing experimental metadata. Information recorded includes details about the biological samples, procedures, protocols, and experimental technologies, all of which

can be easily annotated using the appropriate ontologies. In that way they can capture, describe and share details about their experiments.

DUMONTIER *et al.* [21] made software, which aims to provide intuitive tools for life scientists for representing, integrating, managing and querying of heterogeneous and distributed biological knowledge. Ontologies in this application are described using OWL DL, a sublanguage of the Web Ontology Language (OWL) [6]. Features of the application include semantic query composition and validation using DL reasoner, graphical representation of the query, mapping of DL queries to SPARQL query language and retrieval of pre computed inferences from an RDF triple store.

All mentioned software solutions directly emphasized the great importance of Semantic Web. Semantic technologies allow representing, using, managing and sharing of biological knowledge. Intelligent system for search of experiments, which is suggested in this paper, is based on the Semantic Web technologies. Developed ontology for experiment and knowledge base, populated with real experiment data, is an important part of system. By specifying various search criteria in combination with SPARQL, the efficient retrieval and analysis of information are provided. In this way system performs the selection of results that are in form of written report. The advantages of our system are easier searches, adequate representation of the results and reduced time needed to get the relevant information. Intuitive and easy for use, system allows users to obtain in a simple way all required information for further research and publication of results.

Considering the above principles and proposed solutions, and taking into account certain defined relationships in our structure, specific rules for construction of ontology for RC experiments structure were developed. In the next section, the principles and ontology development are described.

EXPERIMENT ONTOLOGY DESIGN

Ontology is a formal explicit specification of concepts and relationships among them in some domain of interest from the real world, based on previous expert knowledge and understanding with clear idea of the ontology purpose and use. Main components of ontology are classes and properties. Class represents the set of individuals. Properties can be object or data type, and both link individuals from domain to individuals from range. Data type properties permit the instance to have data values of simple data types (string, integer, float ...).

Having thoroughly observed and analyzed the semantic relationships among experiment concepts from Fig. 1, a simple pattern for creation of corresponding ontology followed: every concept corresponds to a class or property of some other class or can be both, a property which is a class. The importance of modeling the experiment semantic structure is that it defines the structure of experimental data and results obtained at the RC.

General development principles that are adopted for the PIBAS ontology design can be characterized as:

- Top down, starting with the most general concepts first.
- Hierarchical concepts with subclass relationships organized in taxonomy like structure.
- Incremental with adding of new concepts to existing structure, rather than rearranging the existing structure to "create place" for new concepts.
- Iterative when advanced insight into ontology design might require changes in previous structure.
- Direct correspondence of real world concepts with ontology classes.
- Direct correspondence of properties and subclasses with real world concepts.
- Direct correspondence of real world entities with ontology individuals.
- Expression of real world constraints with available ontology restrictions.

Ontology was created with Protégé platform independent environment for creating and editing ontologies in various RDF syntaxes [10]. Process of ontology creation according to mentioned design principles also required an active role of expert staff from RC laboratories. Ontology class hierarchy is defined by a “subclass” relationship of ontology concepts and all classes being direct or indirect subclass of class *Thing*. For example, the concept “model system” from structure is represented as *ModelSystem* class that has appropriate subclasses: *Animal*, *CellCulture* and *Patient*. It is also possible to assert that two classes are disjoint for each other. For example, class *Chemical* is disjoint with the class *Plant*, which means that individuals of one class cannot be member of other disjoint class. Fig. 2 shows the class hierarchy for PIBAS ontology in Protégé editor.

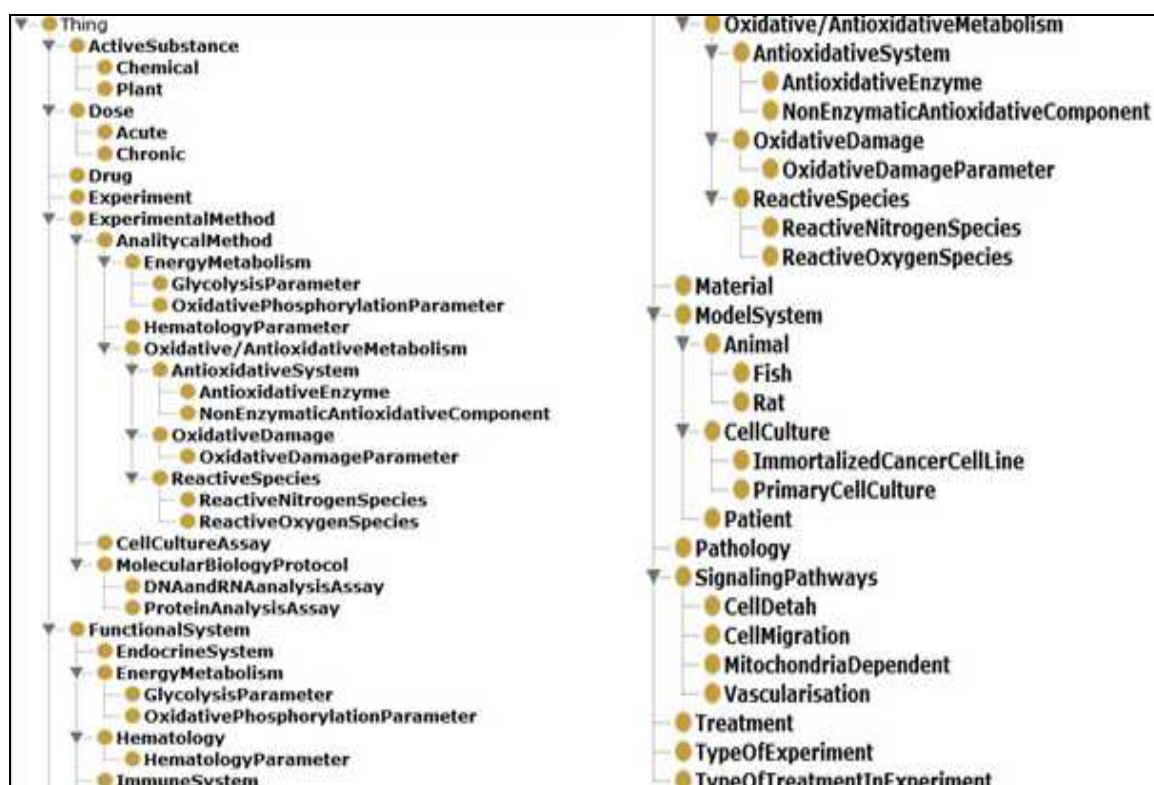


Figure 2. Class hierarchy for Experiment ontology in Protégé.

Example of definition for class Experiment in the RDF/XML syntax is given in Listing 1.

```
<!--http://cpctas-lcmb.pmf.kg.ac.rs/2012/3/PIBAS#Experiment-->
<owl:Class rdf:about="&PIBAS;Experiment">
  <rdfs:comment>
    An experiment is a methodical trial and error procedure carried out with
    the goal of verifying, falsifying, or establishing the validity of a
    hypothesis. Experiments vary greatly in their goal and scale, but always
    rely on repeatable procedure and logical analysis of the results. An
    experiment is a method of testing - with the goal of explaining - the
    nature of reality.
  </rdfs:comment>
</owl:Class>
```

Listing 1. Class definition for the Experiment.

Property describes a link of class instance with some value that can be simple value in case of data property or an instance of other class in case of object property. For example, object

property *modelSystem* links instances of classes *TypeOfExperiment* and *ExperimentalMethod* with instances of class *ModelSystem*, according to relationships in hierarchy. On the other hand, for example, the data type property *theAimOfExmperiment* was used to connect class *Experiment* with aim of real experiment in form of string. Fig. 3 shows object and data type properties of PIBAS ontology in Protégé editor.

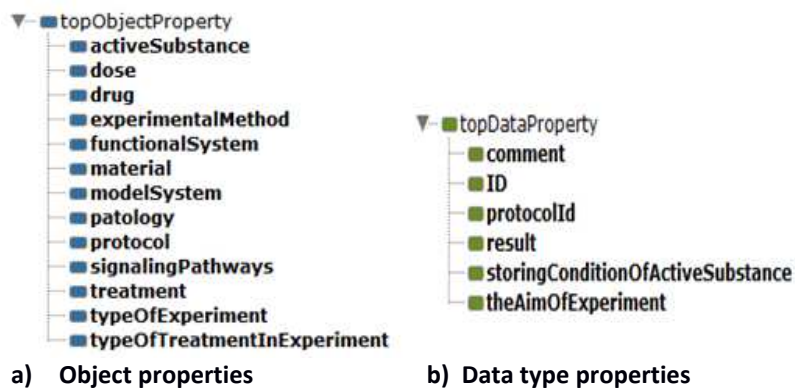


Figure 3. PIBAS ontology properties.

Ontology visualization

When using ontology for modeling and representing the semantic structures on the web, it is desirable to have some kind of adequate visualization of that semantic structure accessible to web users. Visualization is very important as it increases the user's benefits obtained by using the ontology.

Graphical representation of PIBAS ontology in Fig. 4 is available at:

<http://cpctas-lcmb.pmf.kg.ac.rs/jit/>.

The blue segments in Fig. 4 present the concepts / classes that are properties of the concepts on the left, while the yellow segments present the concepts which are subclasses of the item on the left. All segments can further develop into branches when selected by the user, until leafs are reached. For ontology visualization, the InfoVis Toolkit library was used.

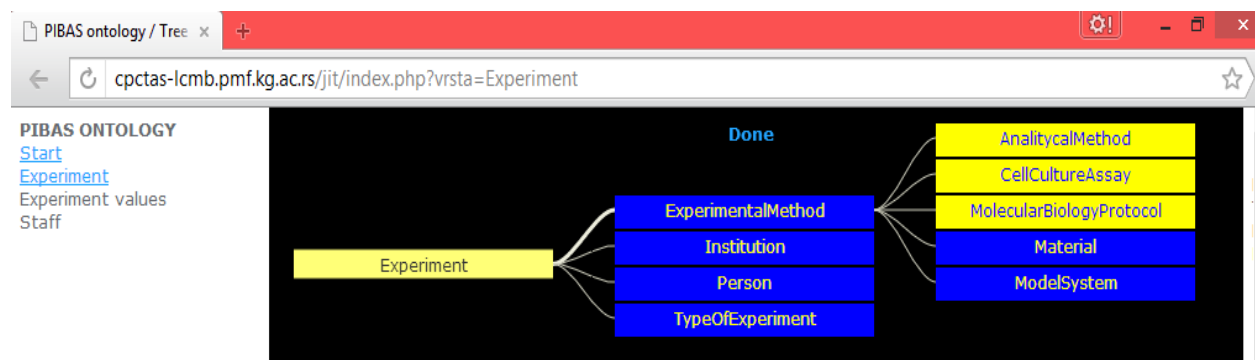


Figure 4. Graphical presentation of Experiment hierarchical structure on web page.

Ontology supported knowledge base

Terms “Knowledge base” and “Ontology” are similar as they both relate to some kind of real world knowledge. Ontology represents knowledge about concepts and their relationships, not dealing with class instances. The ontology is knowledge about the structure, while the knowledge base also contains the class instances from the real world, organized and classified according to ontology [18]. Class instances representing the objects from the real world can be in the same file with ontology, which can be convenient. But, in some cases, it may be desirable to

separate the ontology that describes the semantic structure, from many independent groups of instances. In other words, more precisely, to divide the knowledge base to file with ontology, and to files containing data - the instance groups, each group from one particular experiment. Each group of instances is independent from all other groups, and despite the Protégé being an excellent ontology software tool, the knowledge base may become quite “messy” and hard to work with, due to presence of very large number of instances. Different active substances, protocols and model systems, which are used in experiments, are represented in different RDF/XML files and they are related with real experiments using data type property *ID*. For example, RDF/XML file with protocols contains 35 instances, which are used in RC for experiments (Fig. 5). One protocol instance *UM08* belongs to *ProtenAnalysisAssay* class and has name *ProteinExtractionFromTissue*.

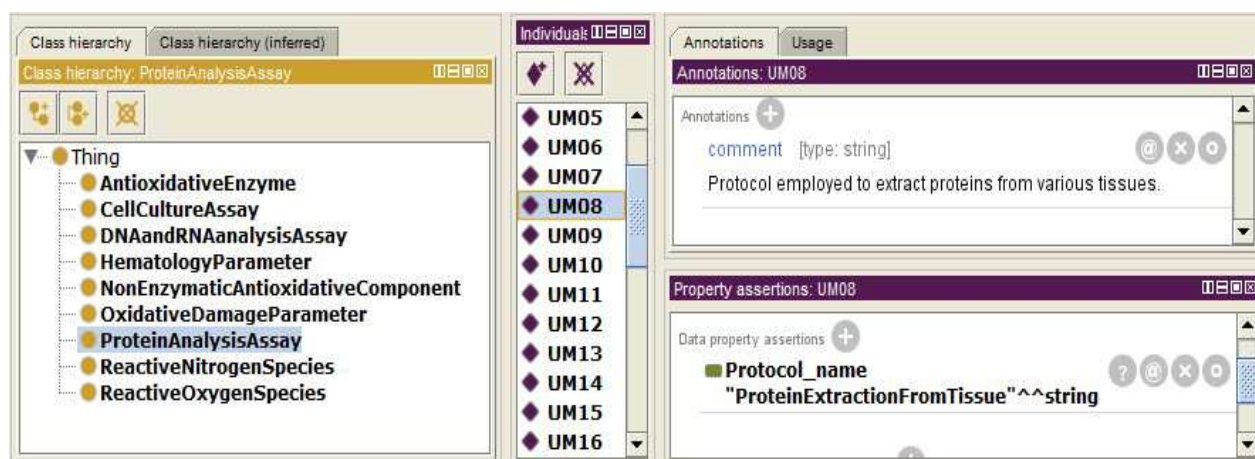


Figure 5. Separate file with protocols.

PIBAS ontology describes the experiment structure, but the Research Centre performed many real experiments according to described structure, which resulted in many data sets, the each data set for one experiment. In that case, it is possible and desirable to separate the each data set relating to one experiment into separate ontology file. That file has only the instances with values from experiments, without the structure itself, which is in separate ontology file, but without data. No matter how many experiments are performed, data for the each experiment are placed in separate RDF/XML file, which is related to PIBAS ontology by the common namespaces, classes and class properties which can be object and data type.

The example of ontology for Experiment data set is given in Listing 2. This RDF/XML syntax represent experiment called *Antiproliferative and proapoptotic activities of methanolic extracts from different Teucrium species on HCT-116 cell*. For this real experiment certain active substances, protocols and model system are connected, which is represented by the use of appropriate object properties and instances.

```
<owl:NamedIndividual
rdf:about="&PIBAS;Antiproliferative_and_proapoptotic_activities_of_methanolic_extra
cts_from_diferent_Teucrium_species_on_HCT-116_cell_line">
  <rdf:type rdf:resource="&PIBAS;Experiment"/>
  <PIBAS:ID rdf:datatype="&xsd:int">104</PIBAS:ID>
  <PIBAS:theAimOfExperiment>
    The aim of this study was to determinate the antiproliferative effects
    of
    different Teucrium species and determinate type of cell death on HCT-116
    cell line, human colon cancer.
  </PIBAS:theAimOfExperiment>
  <PIBAS:storingConditionOfActiveSubstance>'at freezer';
```

```

</PIBAS:storingConditionOfActiveSubstance>
<PIBAS:modelSystem rdf:resource="&PIBAS;MS07"/>
<PIBAS:activeSubstance rdf:resource="&PIBAS;AS34"/>
<PIBAS:activeSubstance rdf:resource="&PIBAS;AS35"/>
<PIBAS:activeSubstance rdf:resource="&PIBAS;AS36"/>
<PIBAS:activeSubstance rdf:resource="&PIBAS;AS37"/>
<PIBAS:activeSubstance rdf:resource="&PIBAS;AS38"/>
<PIBAS:activeSubstance rdf:resource="&PIBAS;AS39"/>
<PIBAS:activeSubstance rdf:resource="&PIBAS;AS40"/>
<PIBAS:activeSubstance rdf:resource="&PIBAS;AS41"/>
<PIBAS:experimentalMethod rdf:resource="&PIBAS;UM01"/>
<PIBAS:experimentalMethod rdf:resource="&PIBAS;UM02"/>
<PIBAS:experimentalMethod rdf:resource="&PIBAS;UM03"/>
<PIBAS:experimentalMethod rdf:resource="&PIBAS;UM04"/>
<PIBAS:experimentalMethod rdf:resource="&PIBAS;UM05"/>
<PIBAS:experimentalMethod rdf:resource="&PIBAS;UM21"/>
</owl:NamedIndividual>

```

Listing 2. Ontology for Experiment data set.

SYSTEM FOR EXPERIMENT SEARCH

Architecture for proposed system of experiment search is presented in Fig. 6.

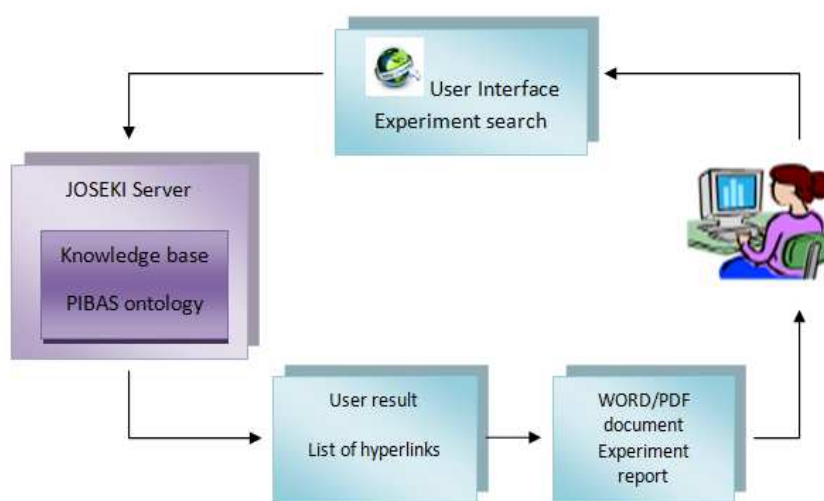


Figure 6. Architecture of system for experiment search.

Generally, architecture is divided into three layers:

- User Interface layer
- JOSEKI layer
- Query processing layer

First layer of architecture refers to the web page for search of the experiment, which is shown in Fig. 7. This web page contains many various search conditions in the upper part, which are grouped according to similarity. Conditions are logically classified in groups such as “Institution/personal data”, “Required analysis”, “Dates”, “Analysis method”, “Active substance” and “Model system”. In the lower part is the table with the ordered list of experiments that satisfy current search conditions.

Experiment search

Institution / personal data

Institution	User	User representative

Required analysis

User required analysis

Laboratory staff

Researcher	Responsible researcher	Manager

Oct 28 2012

Calendar date value Clear date Experiment search

Date from Date to

Analysis dates

Date	<input checked="" type="radio"/> Reception date	<input type="radio"/> Analysis date	<input type="radio"/> Final results date
From			
To			

Analysis method

Analysis method type	Analysis method

Active substance

Active substance type	Active substance
Chemical <input type="checkbox"/> Plant <input type="checkbox"/>	

Model system

Model system type	Model system

No	Experiment title
1	Effects of palladium complexes and cisplatin on proliferation HCT-116 cell line
2	Antiproliferative and proapoptotic activities of methanolic extracts of Taxus baccata L. on HCT-116 and MDA-MB-231 cell lines

Figure 7. Experiment search web page.

The search principles are very simple, if no constraints are specified by search conditions, then the complete experiment list is obtained in the table below, as shown in Fig. 8. Experiments for which there is a report in the form of a word document are displayed as hyperlinks targeting the document with report, while experiments without such documents are displayed as ordinary text, like experiments 15 and 25.

Specifying one or more constraints or search conditions, results in obtaining the appropriately filtered list of experiments that satisfies the specified constraints. Logical dependencies of search conditions are implemented as connected or coupled drop down lists with search condition items from knowledge base. Changing of some more general or independent search condition, results in automatic appropriate constraining of drop down item lists of dependent search conditions.

Selection of “Institution” that requests some experimental analysis from the RC, automatically constrains the elements of drop down lists for “User” and “User representative”, which are selected institution staff members, and also the list of “Required analysis” only to items ordered by the selected institution.

Laboratory staff, the “Researcher”, “Responsible researcher” and “Manager” is independent of any other selections. Although some dependencies might exist (be deduced) for mentioned laboratory staff in knowledge base, no dependencies are implemented in web page, in order not to make the user interface unnecessarily complex – with too little benefit for increased complexity.

No	Experiment title
1	Effects of palladium complexes and cisplatin on proliferation HCT-116 cell line
2	Antiproliferative and proapoptotic activities of methanolic extracts of leaves and seed cones from Taxus baccata L. on HCT-116 and MDA-MB-231 cell lines
3	Flowcytometric Evaluation of Mechanism of Palladium Complexes-Induced Cell Death on HCT-116 AND MDA-MB-231 Cell Lines
4	Cytotoxic activities of lichens methanol extracts
5	Antioxidant properties of various lichen species on HCT-116 cell line
6	Antioxidant properties of various lichen species on HCT-116 cell line 2
7	Antioxidative properties of Teucrium plant species on HCT-116 cell line
8	Antiproliferative and proapoptotic activities of methanolic extracts from diferent Teucrium species on HCT-116 cell line
9	Cytotoxic activities of five lichens methanol extracts
10	Structure activity relationships of 3 substituted 5 5 diphenylhydantoinas as potential antiproliferative agents
11	Cytotoxic activities of platinum complexes
12	Effects of apox ligand and palladium complex on proliferation of HCT-116 cell line
13	Antiproliferative activity of methanolic extract from Sedum acre L. on HCT-116 cell line
14	Cytotoxic activities of mushroom methanol extracts
15	Antioxidant properties of methanolic extracts of leaves and fruits from Ligustrum vulgare L. on HCT-116 cell line
16	Antioxidant properties of methanolic extracts of leaves and fruits from Ligustrum vulgare L. on HCT-116 cell line
17	Antiproliferative activity of acetone, ethylacetate and methanolic extracts from Cornus mas L. on HCT-116 cell line
18	Antiproliferative and proapoptotic activity of Ligustrum vulgare L. extracts in cotreatment with Pd(apox) complex
19	Inhibition of HCT-116 cell proliferation and induction of apoptosis by methanolic extracts of leaves and fruits from Ligustrum vulgare L. in dependance of time exposure
20	Antiproliferative effects of mushroom methanol extracts in cotreatment with platinum complex [pt(dpe-(s,s)-eddp)cl4] on HCT-116 cell line
21	Antiproliferative effects of mushroom methanol extracts in cotreatment with [pt(dpe-(s,s)-eddp)cl4] on HTC-116 cell line
22	Antiproliferative effects of mushroom methanol in cotreatment with platinum complexes on HCT-116 cell line
23	Synergistic effects of mushroom methanol extracts and cisplatin on the type of cell death of HCT-116 cell line
24	Synergistic effects of mushroom methanol extracts and platinum complex [pt(dpe-(s,s)-eddp)cl4] on the type of cell death of HCT-116 cell line
25	Cytotoxic activities of ethylacetate extract of cornus mas

Figure 8. Unfiltered lists of experiments.

Various analysis dates can be specified, both dates “from” and dates “to” for “Reception date”, “Analysis date”, and “Final results date”. Specification of any or all of dates constraints the experiment list accordingly. “Analysis method type” dropdown list items select a corresponding group of analysis methods, and filter the “Analysis method” dropdown list items just to methods from the selected group. Fig. 9 a) shows the complete items list of “Analysis method type” dropdown list, on the left, and for “Analysis method” dropdown list on the right. Fig. 9 b) shows filtered items list in “Analysis method” dropdown list for selected item “Cell culture assay” in “Analysis method type” dropdown list.

Figure 9 consists of two screenshots of a software interface for selecting analysis methods. Screenshot (a) shows a dropdown menu for 'Analysis method type' with options: Cell culture assay, DNA & RNA analysis assay, Protein analysis assay, Reactive nitrogen species, and Reactive oxygen species. Below it, there are checkboxes for 'Active substance type' (Chemical, Plant) and a 'Model system type' dropdown. Screenshot (b) shows the 'Analysis method type' dropdown set to 'Cell culture assay', which has filtered the 'Analysis method' dropdown to only include cell culture-related methods: UM01 Cell Freezing and de-Freezing, UM02 Cell Passaging, UM03 Cell Counting and Viability Cells, UM04 Cell Growth Curve, UM05 MTT ASSAY, UM06 Immunofluorescence, and UM21 AO-EB Double staining.

a) List of all “Analysis method type”

b) Constrained lists of “Analysis method” items for selected “Analysis method type”

Figure 9. Analysis method restrictions.

“Model system type” dropdown list items select a corresponding group of model systems, and filter the “Model system” dropdown list items just to model system from the selected group.

Fig. 10 shows filtered items list in “Model system” drop down list for selected item “Cell Culture – Immortalized Cancer Cell Line” in “Model system type” dropdown list.

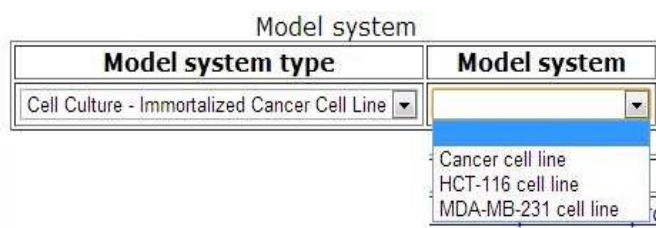
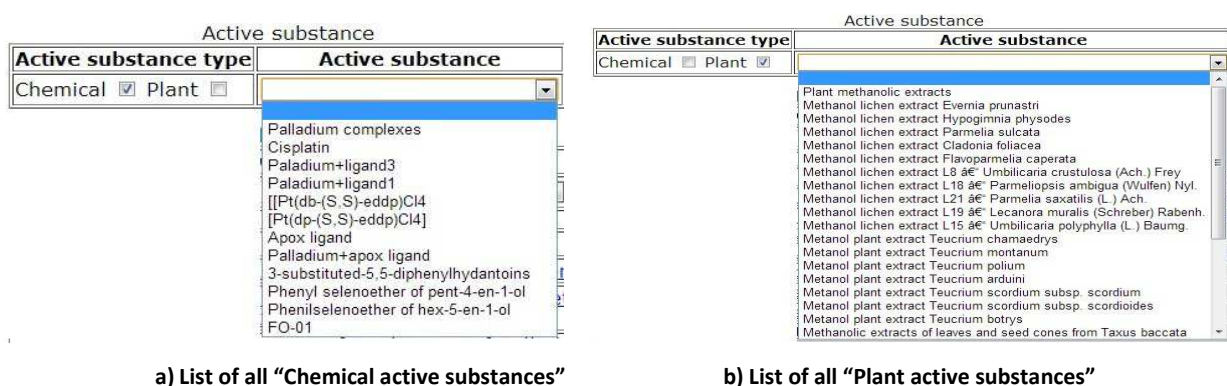


Figure 10. Model system restriction.

An “Active substance type” dropdown list has following options: when “Chemical” is selected only chemicals extracts are given (Fig. 11 a), when “Plant” is selected only plant extracts are given (Fig. 11 b.), while when both “Chemical” and “Plant” are selected or non selected, all chemical and plant extracts in the “Active substance” list are given.



a) List of all “Chemical active substances”

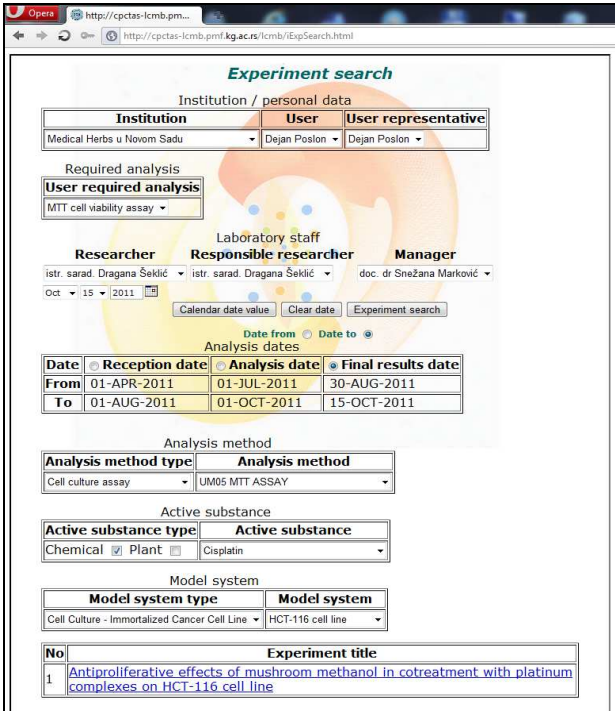
b) List of all “Plant active substances”

Figure 11. Active substance restriction.

The actual search of experiments is performed by defining dynamic SPARQL queries on the server. SPARQL is very powerful ontology query language. Dynamic SPARQL query that was used for retrieving experiment list according to arbitrary selection of search conditions was logically structured with software switches implemented in PHP on the server. Every search condition that can be selected by user, has an equivalent part of appropriate well structured SPARQL query that is controlled by corresponding software switch, which is included in final SPARQL query if the search condition is selected. Selected values of the search conditions are transferred as parameters to SPARQL query. Suggested search procedure for a user is to start with the most important search condition, and to gradually continue with others, as it is easily possible to select the search conditions combination without intersection, and therefore with no results.

Dynamically created SPARQL query is then forwarded to JOSEKI layer of architecture. JOSEKI is SPARQL server which contains ontology knowledge base. When the query is forwarded to the server and executed, the result which is actually filtered list of experiments – experiment titles in the form of hyperlinks, appear on the web page. The hyperlinks point to documents containing reports officially prepared for the each experiment by the staff that performed the experiment. Those reports are official documents for the institution that requested the experiment. The experiment search system enables the new insight and statistics by dynamic grouping of performed experiments according to various search conditions that can be arbitrarily combined.

For example, Fig. 12 a) shows the search result with all search conditions selected, while Fig. 12 b) shows the corresponding SPARQL query.



Experiment search

Institution / personal data

Institution	User	User representative
Medical Herbs u Novom Sadu	Dejan Poslon	Dejan Poslon

Required analysis

User required analysis

MTT cell viability assay

Laboratory staff

Researcher	Responsible researcher	Manager
istr. sarad. Dragana Seklic	istr. sarad. Dragana Seklic	doc. dr. Snezana Markovic

Date from: Oct 15 2011

Calendar date value | Clear date | Experiment search

Date from: Date to

Analysis dates

Date	Reception date	Analysis date	Final results date
From	01-APR-2011	01-JUL-2011	30-AUG-2011
To	01-AUG-2011	01-OCT-2011	15-OCT-2011

Analysis method

Analysis method type	Analysis method
Cell culture assay	UM05 MTT ASSAY

Active substance

Active substance type	Active substance
Chemical <input checked="" type="checkbox"/> Plant <input type="checkbox"/>	Cisplatin

Model system

Model system type	Model system
Cell Culture - Immortalized Cancer Cell Line	HCT-116 cell line

No | Experiment title

1	Antiproliferative effects of mushroom methanol in cotreatment with platinum complexes on HCT-116 cell line
---	--

```

prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix rdfs:<http://www.w3.org/2000/01/rdf-schema#>
prefix owl:<http://www.w3.org/2002/07/owl#>
prefix pibas: <http://cpctas-lcmb.pmf.kg.ac.rs/2012/3/PIBAS#>
prefix xsd:<http://www.w3.org/2001/XMLSchema#>
select distinct ?value ?search
{
  ?exp rdf:type pibas:Experiment;
  pibas:ExpTitle ?search;
  pibas:Web ?value;
  pibas:ReceptionDate ?reccdate;
  pibas:AnalysisDate ?andate;
  pibas:Date ?date;
  pibas:institution pibas:Medical_Herbs_in_Novi_Sad;
  pibas:experimentalMethod pibas:UM05;
  pibas:modelSystem pibas:MS07;
  pibas:activeSubstance pibas:AS11;
  pibas:user pibas:Dejan_Poslon;
  pibas:userRepresentative pibas:Dejan_Poslon;
  pibas:UserRequiredAnalysis "MTT cell viability assay";
  pibas:researcher pibas:Dragana_Seklic;
  pibas:responsibleResearcher pibas:Dragana_Seklic;
  pibas:Chemical ?a;
  pibas:experimentalMethod ?expmet.
  ?expmet pibas:protocolType pibas:CellCultureAssay.
  FILTER (?reccdate > "2011-04-01"^^xsd:date)
  FILTER (?reccdate < "2011-08-01"^^xsd:date)
  FILTER (?andate > "2011-07-01"^^xsd:date)
  FILTER (?andate < "2011-10-01"^^xsd:date)
  FILTER (?date > "2011-08-30"^^xsd:date)
  FILTER (?date < "2011-10-15"^^xsd:date)
}
order by ?date

```

a) User web page with selected conditions and result

b) Appropriate background SPARQL query

Figure 12. Experiment search result for all selected search conditions.

CONCLUSION

The main mission of the RC is to contribute to improvement of existing and development of new knowledge about the effects of different active substances and their potential application to living systems (cells, tissues, organs, organisms) by scientific research and education process. Many various experiments that were carried out at RC resulted in creation of a unique and complex structure of the experimental work performed at RC, which requires adequate representation, in order to assist RC staff and other users.

Ontology, as the core component of Semantic Web, was used as a universal mean for easy modeling of various semantic structures using the Protégé semantic editor, and is suggested as a universal mean for fast and easy representing of required various semantic structures. In accordance with developed PIBAS ontology, knowledge base was designed and populated with real experimental data. Data from developed ontology and knowledge base can be retrieved using specifically developed navigation system of experiment search based on user selected multi criteria for arbitrarily filtering of search results. User selected filtering is dynamically mapped to SPARQL queries which define data to be obtained from ontology knowledge base on the server. Obtained experiment search results are listed on a web page in the form of hyperlinks pointing to documents containing official experimental report created by RC staff.

The experiment search system allows laboratory staff and other users to easily search and identify which active substances, methods, protocols and model systems were used for the experiment. Besides, the clients that requested the experimental tests can be searched for, together with RC staff that worked on a particular experiment. Search results obtained in that way can provide for new insight in performed experimental work thus helping in making decisions for directions of the future work.

The multi criteria search will be improved in the future by introducing of new additional search criteria and connecting with other ontologies which is one of the most important tasks for the future work.

Acknowledgments

This paper is result of the work on the project III 41010 Preclinical tests of bioactive substances financed by the Serbian Ministry of Education and Science.

References:

- [1] BERNERS-LEE, T., HENDLER, J. and LASSILA, O. (May 2001): The Semantic Web. *Scientific American* **284** (5): 29-37.
- [2] Center Project data <http://cpctas-lcmb.pmf.kg.ac.rs/lcmb/pibasEn.htm>.
- [3] CPCTAS-LCMB, Faculty of Science, University of Kragujevac, Serbia <http://cpctas-lcmb.pmf.kg.ac.rs>.
- [4] GRUBER, T. (2001): *What is an Ontology?* Stanford University. <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.
- [5] HARRIS, S., PRUD'HOMMEAUX, E., SEABORNE, A. (2012): Sparql 1.1 Query language. W3C working draft 05 January 2012. *World Wide Web Consortium*, <http://www.w3.org/TR/2012/WD-sparql11-query-20120105/>.
- [6] MCGUINNESS, D.L. and VAN HARMELEN, F. (2004): OWL web ontology language overview. *W3C recommendation*, **10** (2004-03): 10.
- [7] MGED Ontology mged.sourceforge.net/
- [8] NOY, N.F., SHAH, N. H., WHETZEL, P.L., DAI, B., DORF, M., GRIFFITH, N., and MUSEN, M.A. (2009): BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*, **37** (suppl 2): W170-W173.
- [9] Ontology for Biomedical Investigations <http://obi-ontology.org/>
- [10] RDF/XML <http://www.w3.org/TR/rdf-syntax-grammar/>
- [11] RISHE, N., FURHT, B., ADJOUADI, M., BARRETO, A., DAVIS, D., WOLFSON, O. and YESHA, Y. (2011): Semantic wrapper: concise semantic querying of legacy relational databases. In: *Handbook of data intensive computing*, FURHT, B. and ESCALANTE, A. (Eds.), Springer New York pp. 415-444.
- [12] Semantic Web Wikipedia, http://en.wikipedia.org/wiki/Semantic_Web
- [13] SMITH, B., ASHBURNER, M., ROSSE, C., BARD, J., BUG, W., CEUSTERS, W. and LEWIS, S. (2007): The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology* **25** (11): 1251-1255.
- [14] SMITH, B., CEUSTERS, W., KLAGGES, B., KÖHLER, J., KUMAR, A., LOMAX, J. and ROSSE, C. (2005): Relations in biomedical ontologies. *Genome biology* **6** (5): R46.
- [15] SOLDATOVA, L.N. and KING, R.D. (2005). Are the current ontologies in biology good ontologies? *Nature biotechnology*, **23** (9): 1095-1098.
- [16] SOLDATOVA, L. N. and KING, R. D. (2006): An ontology of scientific experiments. *Journal of the Royal Society Interface*, **3** (11): 795-803.
- [17] SOLDATOVA, L. N. and KING, R. D. (2004): Ontology of Experiments for Bioinformatics. *Japanese Society for Bioinformatics GIW 2004 Poster Abstract*: P180.
- [18] STAAB, S., & STUDER, R. (2010): *Handbook on ontologies*. Springer.
- [19] ORCHARD, S., KERSEY, P., HERMJAKOB, H. AND APWEILER, R. (2003): The HUPO Proteomics Standards Initiative meeting: towards common standards for exchanging proteomics data. *Comparative and functional genomics*, **4**(1): 16-19.
- [20] TOMLINSON, C. D., BARTON, G. R., WOODBRIDGE, M. and BUTCHER, S. A. (2013): XperimentR: painless annotation of a biological experiment for the laboratory scientist. *BMC bioinformatics*, **14** (1): 8.
- [21] VILLANUEVA-ROSALES, N. and DUMONTIER, M. (2008): yOWL: An ontology-driven knowledge base for yeast biologists. *Journal of biomedical informatics*, **41** (5): 779-789.