SEMANTIC PUBLISHING AS A TOOL IN SMART, SELF-PROTECTIVE CRITICAL INFRASTRUCTURE

ABSTRACT

Semantic publishing as a semantic representation of publications makes articles machine-readable. It involves linking an article with another articles and external sources. Also, data provided along with the publication can be made machine-readable, which allows visualizing the data in different ways according to a reader's needs. This kind of semantic technology deployed on physics domain can serve for improving interoperability in critical infrastructure as far as it includes such facilities as atomic power stations, atomic plants and local experimental atomic devices, experimental fusion reactors and demonstration fusion reactors among others. Having shared terminology and data represented in different ways, services and units of critical infrastructure may communicate more effectively. At first, it may help staff members of the critical infrastructure facilities in making decisions, but since information becomes machine-readable this approach may exclude the human factor from the issue of the security of the critical infrastructure at all. This paper gives proposition for improvement of the current state of semantic publishing in physics domain, particularly, utilizing SPAR suit of ontologies that is designed for semantic publishing, and open problem ontology that was developed within the work on this paper.

KEYWORDS:

Critical infrastructure, semantic publishing, semantic representation of documents, ontology in physics.

INTRODUCTION

Instrumentation and control systems at nuclear facilities, related to critical infrastructure, are intended for secure control of complex processes, which may cause harm for people and/or the environment. Among such facilities one can specify [Foundations, 1997], foremost, atomic power stations, atomic plants and local experimental atomic devices, experimental fusion reactors and demonstration fusion reactors. Undoubtedly, the mentioned control systems require protection, whilst nowadays the security function can be a part of the system as a self-protective feature of the autonomic system [Hinchey, 2006].

There are two main reasons for the accidents at nuclear facilities: one of them is external attacks, and another one is software failures, the latter are often compounded by human error [Byres-2004]. Thus, the improvement of control systems is an important issue for the industry, which can be solved only with the joint efforts of physicists and computer scientists.

Hence, the purpose of a research, the part of which is the current paper, is to find a more efficient way to communicate for personnel and for units of critical infrastructure by using a formal and shared knowledge of the infrastructure. In other words, the objective is was to improve interoperability between different services and between tools or systems, which will be able to cooperate with each other better, to "understand" each other, and be consequently more efficient, particularly for the issues of cyber security.

At the present time, Searching the Web is inefficient because the searching engine is looking for text or picture matches, but very often generalizations are used as search words, which leads to insufficient searching results. Especially it matters when Web is used for scientific research [Hendler, 2003]. Thus, Artificial Intelligence (AI) presents with a new generation of Web technology, namely the Semantic Web that implements machine-readable descriptions (annotations) of Web resources and machine-readable thesauruses (ontologies). This approach will help to establish links between generalizations and specific concepts, moreover, the links may appear where relationships are currently unsuspected, and in other words, it will make the web-search smart.

This paper includes an overview of the semantic publishing state-of-art and ontologies in physics that exist at the present time. The paper makes a proposition to apply an approach of semantic publishing in documents representation in critical infrastructure.

1. SEMANTIC ANNOTATION AND PUBLISHING OF ARTICLES IN PHYSICS

1.1 Demand for semantic publishing

Scientists write articles and publish them for a number of reasons apart from that their job is evaluated on the basis of a number and quality of the articles: to secure intellectual property, to share results in the community. A sharing implies a wide accessibility to materials; however the current format of scientific journals, even though most of them are nowadays online, does not meet the requirement of an accessibility. This happens because a search on the Web is processed in an ineffective manner since it is still just the Web and web technologies, not Semantic Web and semantic technologies. The most popular format of files is PDF and it is perhaps convenient for readers, but a PDF file is not machine-readable in the current form in the meaning that searching engines are not able to "understand" the sense of the files. Hence, the semantic annotation and publishing of scientific reports will bring a significant improvement in accessibility to scientific results and, thus, will make sharing more effective.

Another point, which is not less important than the aforenamed, is disambiguation in using terms. Even though authors of scientific articles aim to avoid ambiguous terms, still the disambiguation may come from using terms in different sub-communities, moreover, different mother-tongues of people may cause a misunderstanding. Both these problems are addressed by semantic technologies.

The approach of semantic annotation for scientific proposes perhaps arose in 2005 with a YeastHub web-based application [Cheung, 2005], which is, in a nutshell, a biology data warehouse based on RDF technology. Then, the term semantic publishing appeared in 2009 by virtue of David Shotton. In his article [Shotton, 2009b] he described how authors, publishers, and readers would benefit from semantic publishing and what it would cost for them. Nevertheless, his conclusion is univocal: "publishers, who continue to offer traditional 'text' journals, whether these are in print or online, will lose out" [Shotton, 2009b]. Confirmation for his claim might be seen by a fact that one of the major academic publishers, Elsevier, conducted a so called Elsevier Grand Challenge contest (www.elseviergrandchallenge.com), which was intended to find the best solutions in semantic publishing in the domain of life sciences.

In the same article [Shotton, 2009b], David Shotton said that some Scientific, Technical, and Medical Publishers (STM publishers) and software companies (for example, Microsoft) are already bringing to life a number of semantic technologies advanteges, mostly XML markups were utilized though. However, XML is not full power of semantic technologies, because there is still place for an ambiguity, since nests might be interpreted in many ways. Thereafter David Shotton claims that more effort should be done in the developing approaches in the semantic publishing.

1.2 State of art in semantic publishing

In another article [Shotton, 2009a] David Shotton gives the determination of term semantic publishing with follows: "to include anything that enhances the meaning of the published journal article, facilitates its automated discovery, enables its linking to semantically related articles, provides access to data within the article in actionable form, or facilitates integration of data between articles". There is another definition from the main page of David Shotton's web-blog about semantic publishing [Semantic publishing]: "Semantic publishing is the enhancement of scholarly publications by the use of modern web standards to improve interactivity, openness, and usability, including the use of ontologies to encode rich semantics in the form of machine-readable RDF metadata".

In his example described in the same article, he presented a semantically published article in biology. The example includes "semantic markup of textual terms, which links to relevant third-party information resources", which is RDFa and means HTML pages enriched with RDF, machine-readable reference list as separate RDF file, machine-readable self-referencing metadata in separate RDF file. Some useful enrichment has been done manually such as tag cloud and tag tree.

It is worth to say, that in the domain of biology progress in developing ontologies and using semantic technologies, in general, is much higher than in physics. To name one in semantic publishing, there is the Semantic Biochemical Journal experiment, within which Utopia Documents software suit was created [Attwood, 2010].

In simple words, Utopia Documents is a PDF reader, but it makes articles semantically enriched. Utopia Documents connects terms from an article with outside online resources such as Wikipedia, UniProtKB, PDB, etc., it links references to their online versions. Utopia Documents using Bio2RDF project, which has SPARQL and REST interfaces, and DBPedia. It is based on Utopia toolkit [Pettifer, 2009], which has an exemplary architecture (Fig. 1).

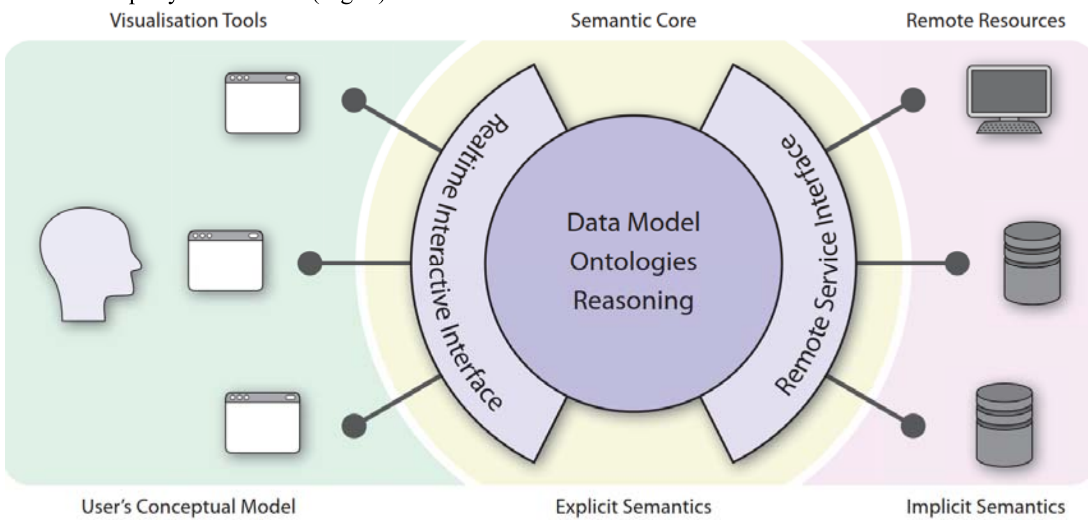

Figure 1. The semantic architecture of Utopia toolkit. (credits: Pettifer, Steve, et al., 2009 Visualising biological data: a semantic approach to tool and database integration. *BMC bioinformatics* 10.Suppl 6: S19). "The Utopia architecture consists broadly of three layers: the visualization tools, which encapsulate the User's Conceptual Model; the core, which encodes and reasons over explicit semantics in its data model; and finally, the remote resources and third-party tools with their implicit semantics."

In addition to applying semantic technologies in semantic publishing as a representing the text itself it often implies access to data used in an article. This is not only about downloadable files with data, but also embedded tools to work with data, for example plotting graphs. The approach of interactive publishing is very attractive and plays a big role in semantic publishing, but it diverts away from the topic of semantic technologies. Visualization tools are desirable, but they are on the second place after the knowledge engineering conception.

Here were mentioned a couple of examples of the semantic publishing efforts in the biology domain, but they are not only.

There is a Semantic Scholar web-service (https://www.semanticscholar.org/) that is a searching engine of scientific articles, which uses semantic technologies approach. Its approach slightly different from semantic publishing: it lies in using machine learning techniques, methods from data mining, natural language processing, and computer vision. The algorithm analyzes an article as a PDF file and extracts keywords, phrases, and important features as figures, tables, and citations so that originally articles do not have any semantic annotation or metadata. However, this searching engine deals only articles in computer sciences.

Semantic publishing and processing of articles is a growing branch, but in physics domain, it is not so developed as in biology or computer sciences, for instance.

2. ONTOLOGY OF PHYSICS

2.1. The need of ontology of physics

The need of ontologies in the domain of physics for semantic publishing was stated in the Introduction, but in addition, there are other applications for them.

The Department of Defense Modeling and Simulation community (USA) is interested in ontology of physics for describing physical-based models in order to provide an interoperability of the simulation modules [Collins, 2004a]. Even though the experts in the field may have a common understanding of the background concepts, when simulation model is being built many assumptions stay hidden even from other experts in the field, thus a comprehensive description of the model is needed.

Another application for the ontology is to create electronic repositories of scientific papers and provide a shared use of them. This is very similar to semantic publishing in a sense of semantic annotation of articles, but the difference is on the level of organizing.

Both applications stated above are expressed by Dr. Josef B. Collins from Naval Research Laboratory in his paper [Collins, 2004b], particularly, he is stated about the communication with the following organizations: American Physical Society (APS), the American Astronomical Society (AAS), the American Institute of Physics (AIP), and the International Union of Pure and Applied Physics (IUPAP), that have expressed interest in electronic document repositories.

2.2 XML approach for ontology of physics

There are number of XML applications that are applicable for physics text annotation:

- Content MathML [Collins, 2005] - Mathematics Markup Language (MathML), namely its part Content MathML allows to annotate complex equation, such as partial differential equations, that are widely utilized for representing physical dynamics.
- PhysML - a markup language for documenting physics knowledge; it includes representation of natural laws, observables, physical systems and experiments.
- OpenMath provides content dictionaries for operations on quantities (namely SI quantities and units [Collins, 2009]), numerical representation of quantities, the concept of standard error, but it does not provide proper treatment of uncertainty. There is a content dictionary for tensor concepts [Collins, 2010], additionally.
- OMDoc (Open Mathematical Documents) [Collins, 2008] - a markup format for annotating mathematical and physical knowledge, which uses Content MathML, Open Math, and PhysML. There is also an RDFa extension of OMDoc [Lange, 2011].

According to W3 consortium XML and its applications are part of the semantic web technologies (https://www.w3.org/RDF/FAQ), but it has different from RDF and OWL area of usage. RDF is more robust against changing of Schemas and Ontologies then XML is versus Schemas, as it is stated by W3C [W3C-FAQ]. Unlike to markup languages, OWL ontologies allow reasoning on them; SPARQL queries are applicable on RDF annotations. The shortcoming of XML is that there is no meaning associated with nested tags and the nesting may be interpreted in many ways. Nevertheless, those XML vocabularies and content dictionaries may serve as a base for future work of developing OWL ontologies.

2.3 OWL/RDF approach for ontology of physics

There are also achievements in this topic in RDF-OWL direction: sedOnto is a **S**ynthetic **E**nvironment **D**ata Representation **Onto**logy [Bhatt, 2004]. The term “synthetic environment” means the depiction of the environment in a digital form (http://www.sedris.org/definition.htm). It is based on SEDRIS Data Representation Model (DRM) and designed to be used within the modeling and simulation domain for representation of data related to Synthetic Environment. SEDRIS, **S**ynthetic **E**nvironment **D**ata **R**epresentation and **I**nterchange **S**pecification, is a technology one of the functions of which is a representation of environmental data, for that it has DRM, which consists of a bunch of object-oriented classes represented in UML. In other words, DRM was mapped into OWL ontology utilizing Protégé development environment (http://protege.stanford.edu).

Another achievement is QUDT ontologies (http://www.qudt.org/), which are set of ontologies of **Q**uantities, **U**nits, **D**imensions and Data **T**ypes, and were developed for the NASA Exploration Initiatives Ontology Models project.

3. SEMANTIC PUBLISHING IN THE DOMAIN OF PHYSICS

3.1 Common features of a content of an article in physics

Any published scientific article has a title, author(s), bibliographic properties, sections, citations, a list of references. Also, any article may be attributing to some opened problem in the field. Usually, but not always, an article has an annotation and list of keywords.

There are the Semantic Publishing and Referencing (SPAR) Ontologies, a suit of OWL 2 ontologies encoded using RDF that covered whole publishing domain [Peroni, 2014]. One of the ontologies from this suit, namely the Citation Typing Ontology (CiTO), was developed by David Shotton et al. during the work on the example of semantic published article [Shotton, 2009a]. The alternative is an application of RDF to describe basic concepts and properties of citations and references: Bibliographic Ontology (http://bibliontology.com/). Scholarly Article Ontology (http://ns.science.ai/) is another possibility to describe a publication with terms defined in RDF Schema.

However, it is more reasonable to use SPAR Ontologies, since they were designed for semantic publishing and referencing purposes and have the biggest scope.

The typical scientific article consists of a narrative part and some kind of data regardless of is it qualitative research or quantitative; most of the scientific articles in natural sciences may include plots, schemes, tables. A narrative part is typically full of terms and concepts.

A scientific article in physics has different data content depending on what origin has paper: theoretical or experimental. A theoretical article typically includes complex and sophisticated equations and numerical methods of solving those equations. Whereas an experimental article includes numbers, units, errors, data performing methods, nevertheless some equations are needed in an experimental article as well, at least fundamental physical laws and, for example, an approximation function.

Representation of mathematical equations is a nontrivial problem for semantic technologies. Some XML applications that address this problem are presented in Section 2.2, but there are not ontologies. Although there are some approaches in ontologies. For example, there is an extension to OWL for defining data ranges in terms of linear (in)equalities with rational coefficients solved over the algebraic reals [Data range extension, 2012]. However, transcendental functions are not permitted as well as non-linear polynomials.

Withal, there is no strong need for the tools for annotation mathematical equations in the case of the annotation of a scientific article: the ontologies of the equation types, fundamental physical laws, and named formulae would be enough.

Besides the content of an article, there is a structure on an article, which also may bring essential information. It concerns to rhetorical components (abstract, introduction, discussion, methods, and conclusion) rather than to structural components (paragraph, section, and chapter).

3.2 The ScienceWISE system

Nevertheless, there is no strong need for annotation mathematical equations, the ontology of physics is needed in every specific subfield. Fortunately, there is a ScienceWISE web-service (http://sciencewise.info/), which is a platform for semantical annotation of the research articles. WISE in the name stands for **W**eb-based **I**nteractive **S**emantic **E**nvironment. The ScienceWISE project was established by the group of physicists from Ecole Polytecnique Federale de Lausanne (EPFL), Switzerland, and CERN in collaboration with computer scientists from EPFL, the University of Fribourg and Vrije Universiteit Amsterdam, Netherlands. The goal of the project is the organizing the scientific knowledge in the domain of physics, however, ScienceWISE has been designed in such a way that it can be used by scientists from any field: nowadays they have Life Sciences, Digital Humanities, and Information Technologies sections under constructions. The project has a connection to ArXiv.org, the CERN Document Server and also Inspire (comprehensive bibliographic database in high-energy).

The basic principal of the project implies scientists create and edit their specific ontologies within the framework of the project by the instrumentality of the ScienceWISE portal. The scientific article may be annotated with the number of concepts that consist in ontologies, alternatively the article from the ArXiv.org archive may be bookmarked automatically, and afterward, a user can edit this annotation. As a result article or author may be searched by one or more concepts. Currently, the ScienceWISE ontology contents more than 60 000 unique entries, each with its own definitions, alternative forms, and semantic relations to other entries [Aberer, 2011; Boyarsky, 2013]. There are several hundreds of active users. The users of the project were and are mostly physicists [Boyarsky, 2015], thus the most described ontology is in physics. Particularly, there are subcategories of astrophysics and cosmology, atomic, molecular and condensed matter physics, nuclear physics, particle physics, plasma physics and a number of others and each of them has thousands of articles related and allocated on the ScienceWISE.

The project has been started using Semantic Web technologies some time ago, performing the "RDFization" of the ontology as they called it [Boyarsky, 2015], and now it is available in RDF format. ScienceWISE ontology has links to DBpedia, Wikipedia and SKOS ontology through the RDF property "sameAs".

Uniqueness and a merit of this project are that they involve specific-domain scientists into the process of ontology developing providing them convenient interface. As a result, ScienceWISE includes the biggest ontology of physics, which is continually growing.

3.3 The proposition of semantic annotation of an article in physics

Taking to the account current state-of-art in semantic publishing and physics the need of semantic processing of scientific articles in physics is evident. The needed result is a possibility of efficient browsing through the scientific articles in physics with semantically rich querying, for example, "*Find all articles which have particular subject terms*". Thus, the demanded end-product is searching engine and representation system (2 in 1) for scientific articles in physics.

The semantic enrichment of the articles is the key task here since semantic querying can be applied only on machine-readable documents. This can be done automatically utilizing Natural Language Processing and Linked Open Data-based entity detection as claimed by Bahar Sateli and René Witte [Sateli, 2015], this means that no manual effort in annotating from publishers or authors is not required. Semantic vocabularies (ontologies) are the last element needed. There are several public ontologies for semantic publishing as it was stated in Section 3.1. and there is a big ontology of physics terms and concepts namely ScienceWISE (Section 3.2.), but it is not public.

The following example of annotating a scientific article was designed based on examples given for SPAR suit of ontologies [Peroni, 2015]. Here the ontology from the SPAR suit was used, namely FaBiO, and the ScienceWISE ontology (a version of 2014 year). FRBR, DC Terms, and PRISM ontologies were used also since they are used in FaBiO. Then the FOAF ontology was used to annotate persons and XML Schema ontology since it contains basic data types.

```
@prefix : <http://www.example-site.net/example_ontology> .
@prefix fabio: <http://purl.org/spar/fabio/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix frbr: <http://purl.org/vocab/frbr/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix prism: <http://prismstandart.org/namespaces/basic/2.0/> .
@prefix application: <http://purl.org/NET/mediatypes/application/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix swconcept: <http://data.sciencewise.info/concept/> .
#
#:param_o-modes a fabio:ResearchPaper ;
      dcterms:creator :vogirka , :avgirka , :vvyarko ;
      frbr:realization :published_item ;
      fabio:hasSubjectTerm swconcept:Plasma ,
           swconcept: Plasma_physics ,
           swconcept: Maxwell's_equations ,
           swconcept: Surface_wave ,
           swconcept: Controlled_fusion,
           swconcept: Nuclear_fusion ;
      prism:keywords “Vlasov-Boltzmann equation” ,
```

```
            “surface electron O-modes” ,
            “SECOM” ,
            “parametric instability” ,
            “electron cyclotron frequency”,
            “electron cyclotron waves”,
            “surface cyclotron wave propagation” .
#
#:published_item a fabio:JournalArticle ;
      dcterms:title “Parametric excitation of surface electron cyclotron
      O-modes by an external alternating electric field” ;
      fabio:hasPublicationYear “2011”^^xsd:gYear ;
      prism:doi “10.1088/0031-8949/84/04/045503” ;
      frbr:embodiment :pdf ;
      frbr:partOf :phys_scripta-84-045503 .
#
#:phys_scripta-84-045509 a fabio:JournalIssue ;
      prism:issueIdentifier “84” ;
      frbr:embodiment :printed_issue;
      frbr:partOf :phys_scripta .
#
#:phys_scripta a fabio:Journal ;
      dcterms:title “Physica Scripta” .
#
#:printed_issue dcterms:publisher :iopscience ;
      frbr:part :printed .
#
#:pdf a fabio:DigitalManifestation ;
      prism:publicationDate “2011-09-20”^^xsd:date ;
      dcterms:format application:pdf ;
      dcterms:publisher :iopscience .
#
#:vogirka a foaf:Person ;
      foaf:givenName “Volodymyr” ;
      foaf:familyName “Girka” .
#
#:avgirka a foaf:Person ;
      foaf:givenName “Anastasiia” ;
      foaf:familyName “Girka” .
#
#:vvyarko a foaf:Person ;
      foaf:givenName “Vitalii” ;
      foaf:familyName “Yarko” .
#
#:iopscience a foaf:Organization ;
      foaf:name “IOP Science” .
```

This fragment of semantic annotation describes the published article: Girka, V. O., A. V. Girka, and V. V. Yarko. "Parametric excitation of surface electron cyclotron O-modes by an external alternating electric field." *Physica Scripta* 84.4 (2011): 045503, available online at stacks.iop.org/PhysScr/84/045503. The annotation contains the following information: this is a research paper written by 3 persons (lines 13, 14) and it is a printed item (line 15); it has a number of subject terms (lines 16-21) and keywords (lines 22-28); it was published as a journal article with corresponding title (lines 30-33); the publication year and DOI are stated then (lines 34, 35); its embodiment is a PDF file (line 36) and it is a part of journal issue #84, journal title is “Physica Scripta” and publisher is IOP Science (lines 37-48); it was published online on 20th of September 2011. There is also information about the date of publication, a format of publication and a publisher (lines 50-53). Authors are described in lines 55-65 and the publisher is described in lines 67-68. The subject terms and keywords are picked up from the ScinceWISE ontology, version of 2014, but the ontology is growing up every day, users can add new terms in it according to their needs in annotating. *Theoretical physics* could be added into ScienceWISE ontology in order to describe this particular article and the term would be used for many of publications. Also, such a concepts as *wave generation*, *modes conversion*, and *plasma instabilities* describe the article but are not in the ScienceWISE ontology.

Ontology of open problems could be the third component in annotating a scientific article in physics in addition to publishing ontology and physics ontology. There is proposition of ontology of open problems below:

```
@prefix : <http://www.example-site.net/OpenProblemsOnto#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://www.example-site.net/OpenProblemsOnto> a owl:Ontology .
##############################################################
# #    Classes
##############################################################

# http://www.example-site.net/OpenProblemsOnto#FieldOfScience

:FieldOfScience a owl:Class .
#
# http://www.example-site.net/OpenProblemsOnto#OpenProblem

:OpenProblem a owl:Class .
#
# http://www.example-site.net/OpenProblemsOnto# SubfieldOfScience

:SubfieldOfScience a owl:Class .

##############################################################
# #    Object Properties
##############################################################
# http://www.example-site.net/OpenProblemsOnto#concernTo

:concernTo a owl:ObjectProperty ;
      rdfs:domain :OpenProblem ;
      rdfs:range :SubfieldOfScience .
#
# http://www.example-site.net/OpenProblemsOnto#containSubfield

:containSubfield a owl:ObjectProperty ;
      owl:inverseOf :isSubfieldOf ;
      rdfs:domain :FieldOfScience ;
      rdfs:range :SubfieldOfScience .
#
# http://www.example-site.net/OpenProblemsOnto#inFieldOf

:inFieldOf a owl:ObjectProperty ;
      rdfs:domain :OpenProblem ;
      rdfs:range :FieldOfScience .
#
# http://www.example-site.net/OpenProblemsOnto#isSubfieldOf
```

```
:isSubfieldOf a owl:ObjectProperty , owl:FunctionalProperty ;
      rdfs:domain :SubfieldOfScience ;
      rdfs:range :FieldOfScience .
```

The aforecited example of open problems' ontology includes three classes: *FieldOfScience*, *SubfieldOfScience,* and *OpenProblem*. Classes are connected by four properties: *isSubfieldOf* property links *FieldOfScience* with *SubfieldOfScienced*, there is a reverse connection, namely *ContainSubfield*, *OpenProblem* class is linked with *FieldOfScience* class by *inFieldOf* property and with *SubfieldOfScience* class by *concernTo* property.

Thus, the example of annotation of the article may be added with mentioning that the article contributes to the open problem of *Plasma Physics* and *Fusion Power*, which concerns to such a subfield of *Physics* as *Nuclear Physics*. The following segment of code illustrates it:

```
#################################################################
# #    Individuals
# #################################################################

#http://www.example-site.net/OpenProblemsOnto#physics

:physics a owl:NamedIndividual , :FieldOfScience ;
      :containSubfield :astronomyAstrophysics ,
      :atomicMolecularPhysics , :biophysics , :condensedMatterPhysics ,
      :cosmology , :generalPhysics , :generalRelativity ,
      :highEnergyAndParticlePhysics , :nuclearPhysics , :opticalPhysics ,
      :quantumGravity , :quantumPhysics .
#
## http://users.jyu.fi/~angirka/OpenProblemsOnto#nuclearPhysics

:nuclearPhysics a owl:NamedIndividual , :SubfieldOfScience ;
      :isSubfieldOf :physics .

http://users.jyu.fi/~angirka/OpenProblemsOnto#plasmaPhysicsAnd
                                                                FusionPower

:plasmaPhysicsAndFusionPower a owl:NamedIndividual , :OpenProblem ;
      :concernTo :nuclearPhysics ;
      :inFieldOf :physics .
```

Thereby, the annotated scientific article could be found by querying the article, which has subject terms such as *controlled fusion*, *nuclear fusion*, *wave generation*, *modes conversion*, *plasma instabilities* and contributes to the *physics* open problem of *plasma physics* and *fusion power*. These phrases are not a part of the text, thus the regular searching engine, which is based on text matching, would fail in searching this article with these keywords. Here is the SPARQL query composed basing on the requirements stated above:

```
PREFIX fabio: <http://purl.org/spar/fabio/>
PREFIX exmp: <http://www.example-site.net/example_ontology>
PREFIX opprph: http://www.example-
                                        site.net/OpenProblemsOnto#physics
SELECT ?paper
WHERE {
      ?paper
           a fabio:ResearchPaper ;
           fabio:hasSubjectTerm
                      swconcept: Controlled_fusion,
                      swconcept: Nuclear_fusion,
                      swconcept: Wave_generation,
```

```
                     swconcept: Modes_conversion,
                     swconcept: Plasma_instabilities ;
          exmp:contributeTo
                     opprph:plasmaPhysicsAndFusionPower .
}
```

4. SUMMARY

Search engines are used by scientists all over the world in everyday work. There are specialized databases that contain different kind of data on specific topics, which were established in order to help people to find what they need. Some of databases have an atypical search engine as, for example, Semantic Scholar web-service. These tools were created because the ordinary searching in the WWW, which is based on text matching, is not efficient in that cases when the searcher does not know what exactly he or she is searching for, despite it is some particular web page or article. Thus, there is a need for the application of semantic technologies. Particularly, Semantic Web would allow users of it to perform a searching in a smart way, in simple words, it would allow to get shorter and more precise lists of searching results.

The lion's share of searching activity performing by the scientists is devoted to browsing for research articles. Hence, semantic publishing would be useful for every scientist. A number of corporations, including Microsoft and Elsevier academic publisher, are already involved in developing approaches and tools for semantic publishing. Semantic publishing and processing of articles is a growing branch but in physics domain it is not so developed as in biology or computer sciences, for instance.

There is a set of public ontologies for the domain of publishing: SPAR suit of ontologies and some alternatives for it can be found. But still, there is no public ontology in physics which could be used for semantic publishing in physics. However, the ScienceWISE project is continuously working on developing such ontology, particularly in the physics domain. The ScienceWISE ontology of physics has a large scope and its feature is that the contributors are physicists. It is the most suitable ontology for semantic publishing in physics nowadays, but it is not public.

Physicists are in great need of the possibility of searching among scientific reports and data in a more efficient way by utilizing semantic technologies. An example of semantic annotation and querying of the article in plasma physics is presented in this paper. The ontology of open problems is proposed as the improvement to the ontology of physics.

5. CONCLUSION

Semantic technology, as a tool to formalize a domain of knowledge, and ontology, as an essential part of it, can provide shared knowledge and data in the domain of physics, including nuclear and plasma physics for critical infrastructure [Cardoso, 2007].

With this paper the solution for the problem of intercommunication between researchers in physics has been proposed, namely, a semantic publishing in physics as an approach for documents representation, which gives possibility for more effective searching and browsing. An improvement for the existing ScienceWISE system is proposed, namely the ontology of open problems. In addition, ScienceWISE system does not offer the possibility of adding to annotation concepts that are not written down in the text, this limitation needs to be removed in the future. Moreover, semantic publishing includes visualization features of the annotated articles that will be useful for browsing for the articles. The current research is a contribution to the building of the smart, self-protective infrastructure, the semantic publishing in physics may serve as a basis for a document flow in the critical infrastructure, which may decrease a human factor in critical infrastructure protection.

REFERENCES

[Aberer, 2011] Aberer, Karl, et al., 2011 ScienceWISE: a Web-based Interactive Semantic Platform for scientific collaboration. *ISWC (Demonstration Track)*.

[Attwood, 2010] Attwood, Teresa K., 2010. Utopia Documents and The Semantic Biochemical Journal experiment. *EMBnet. news* 15.4: pp-3.

[Bhatt, 2004] Bhatt, Mehul et al., 2004. sedOnto: A web enabled ontology for synthetic environment representation based on the SEDRIS specification. *Fall Simulation Interoperability Workshop*.

[Boyarsky, 2013] Boyarsky, A., et al., 2013. From scientific papers to the scientific ontology: dynamical clustering of heterogeneous graphs and ontology crowdsourcing. *Joint Workshop on Large and Heterogeneous Data and Quantitative Formalization in the Semantic Web (LHD+ SemQuant 2012)*.

[Boyarsky, 2015] Boyarsky, Alexey, and Oleg Ruchayskiy, 2015. ScienceWISE: A Web-Based Interactive Semantic Platform for Paper Annotation and Ontology Editing. *The Semantic Web: ESWC 2012 Satellite Events: ESWC 2012 Satellite Events, Heraklion, Crete, Greece, May 27-31, 2012. Revised Selected Papers* 7540: 392.

[Byres-2004] Byres, Eric, and Justin Lowe. "The myths and facts behind cyber security risks for industrial control systems." *Proceedings of the VDE Kongress*. Vol. 116. 2004.

[Cardoso, 2007] Cardoso, J., Hepp, M. and Lytras, M.D. eds., 2007. The semantic web: real-world applications from industry (Vol. 6). *Springer Science & Business Media*.

[Cheung, 2005] Cheung, Kei-Hoi, et al., 2005. YeastHub: a semantic web use case for integrating data in the life sciences domain. *Bioinformatics* 21.suppl 1: i85-i96.

[Collins, 2004a] Collins, Joseph B., and Doug Clark, 2004. Towards an Ontology of Physics. *NAVAL RESEARCH LAB* WASHINGTON DC.

[Collins, 2004b] Collins, Joseph B., 2004. Standardizing an ontology of physics for modeling and simulation. *NAVAL RESEARCH LAB* WASHINGTON DC.

[Collins, 2005] Collins, Joseph B., 2005. Mathematical Representations for Web Services. Conference paper (*Simulation Interoperability Workshop (SIW)*),
Available at: http://www.nrl.navy.mil/itd/imda/collins-mathematical-representations-web-services

[Collins, 2008] Collins, Joseph B., 2008. Using Mathematical and Scientific Markup as an Approach to Model Specification. *Proceedings of the Society for Modeling and Simulation International [SCS]: Grand Challenges in Modeling & Simulation (GCMS 08)*.

[Collins, 2009] Collins, Joseph B., 2009. OpenMath content dictionaries for SI quantities and units. *Intelligent Computer Mathematics*. Springer Berlin Heidelberg. 247-262.

[Collins, 2010] Collins, Joseph B., 2010. An OpenMath content dictionary for tensor concepts. *Intelligent Computer Mathematics*. Springer Berlin Heidelberg. 293-299.

[Data range extension, 2012] Data range extension : Linear Equations. 2012
Available at: https://www.w3.org/2007/OWL/wiki/Data_Range_Extension:_Linear_Equations

[Foundations, 1997] Foundations, C., 1997. Protecting America's Infrastructures: The Report of the President's Commission on Critical Infrastructure Protection.

[Hendler, 2003] Hendler, James, 2003. Science and the Semantic Web. *SCIENCE*: Vol 299, 520-521.
Available at: http://science.sciencemag.org/content/299/5606/520.full

[Hinchey, 2006] Hinchey, M.G. and Sterritt, R., 2006. Self-managing software. *Computer*,*39*(2), pp.107-109.

[Lange, 2011] Lange, Christoph.2011. Enabling Collaboration on Semiformal Mathematical Knowledge by Semantic Web Integration. PhD thesis, *Jacobs University Bremen, number 11 in Studies on the Semantic Web. AKA Verlag and IOS Press, Heidelberg and Amsterdam*, Chapter 5.

[Peroni, 2014] Peroni, S. 2014. The Semantic Publishing and Referencing Ontologies. *In Semantic Web Technologies and Legal Scholarly Publishing*: 121-193. Cham, Switzerland: Springer.

[Peroni, 2015] Peroni, Silvio, 2015: Example of use of FaBiO #1. Describing a bibliographic entity.
Available at: http://www.sparontologies.net/ontologies/fabio

[Pettifer, 2009] Pettifer, Steve, et al., 2009. Visualising biological data: a semantic approach to tool and database integration. *BMC bioinformatics* 10.Suppl 6: S19.

[Sateli, 2015] Sateli, B. and Witte, R., 2015. Semantic representation of scientific literature: bringing claims, contributions and named entities onto the Linked Open Data cloud. *PeerJ Computer Science*, *1*, p.e37.

[Semantic publishing] Shotton, David, and Peroni, Silvio. The Semantic Publishing Blog
Available at: https://semanticpublishing.wordpress.com/ (Accessed: April 2016)

[Shotton, 2009a] Shotton, David, et al., 2009. Adventures in semantic publishing: exemplar semantic enhancements of a research article. *PLoS Comput Biol* 5.4: e1000361.

[Shotton, 2009b] Shotton, David, 2009. Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing* 22.2: 85-94.

[W3C-FAQ] W3C Semantic Web Frequently Asked Questions. [online] How does the Semantic Web relate to XML? Available at: https://www.w3.org/RDF/FAQ