

Semantic Educational Web Portal

© Victor Telnov

National Research Nuclear University MEPhI,
Obninsk, Russia

telnov@bk.ru

Abstract. The paper deals with the pilot project devoted to the application of the knowledge graphs in the educational activities of the universities. The ontology of the curriculum and the training courses, as well as the means of authoring, enrichment and adaptation of the learning objects are considered. The visual navigation on the knowledge graphs is carried out by using the special searching widgets and smart RDF browser. Working with semantic repository and text analytics is performed on the cloud platforms via SPARQL queries and RESTful services. The software architecture in UML-notation are presented, examples of real use of the educational portal are given.

Keywords: semantic web, educational portal, ontology, knowledge graph, triplestore, RDF storage, graph database, cloud computing.

1 Introduction

Students and professors spend a lot of time and efforts finding useful information, instead of having to comprehend and interpret the learning content. It was rightly observed that the traditional web technologies (sometimes referred to as WEB 2.0) do not provide adequate search and navigation in the environment of distributed knowledge at the semantic level.

Naturally the thought came about some personal smart learning agents (software), which could identify relevant information from any accessible data source and provide an information synthesis tailored to personal learning objective. The idea of semantic educational portals that could provide a meaningful integration of educational objects with the adaptation and personalisation of training courses and curricula, appeared almost simultaneously with the advent of the Semantic Web.

During the semantization the data are combined into triplets in accordance with the RDF model and form a graph. If the data are the learning objects, than that form the so-called knowledge graph. It is obvious that the most adequate repository for the knowledge graphs are the graph databases.

The semantic graph database, also referred to as an RDF triplestore, stands out from the other types of graph databases due to the possibility to support ontologies. The semantic graph database is capable to integrate heterogeneous data from many sources and create relationships between datasets. That database focuses on the relationships between entities and is able to infer new knowledge out of existing information. It is a powerful tool to use in relationship analytics and knowledge discovery.

2 Related Work and Novelty

Proceedings of the XIX International Conference “Data Analytics and Management in Data Intensive Domains” (DAMDID/RCDL’2017), Moscow, Russia, October 10–13, 2017

A recent authoritative overview [14] deals with the Graph and RDF databases makes it possible to navigate among modern products and solutions in the field of the Semantic Web, where the leaders are AllegroGraph, ArangoDB, BlazeGraph, Cray, DataStax, Ontotext GraphDB, IBM Graph, MarkLogic, OrientDB, Neo4j, Stardog, Teradata, Aster, Virtuoso.

It looks very promising the cooperative project Ontotext and Impelsys on the joint using of the platforms GraphDB and Dynamic Semantic Publishing for the development of personalized adaptive learning.

The pilot project [7] which is considered in this article is based on the cloud semantic platform and uses network RESTful services. The preferred repositories for learning objects themselves are the remote data storages. The predecessor of this project is the Cloud cabinet of the Educational portal “Department online” [2]. The project under consideration has been implemented in the educational practice of National Research Nuclear University MEPhI, Russia.

RDF browser is the main highlight of the Semantic Educational Web Portal [7], which distinguishes it from most of the known solutions in the field of the Semantic Web. Once being in the desired place of the knowledge graph via the corresponding widget, then you can to perform a visual navigation in this graph, simply walking along its nodes.

There is a possibility to make a visual walk through the knowledge graph as far as you want in any direction, scooping up the data that appears. By focusing on a specific graph node, it is possible to obtain text metadata, media content and hypertext links that are associated with this node. Along with that the nearest neighborhood of the node becomes visible and accessible for navigation.

3 The Ontology

The fundamental technologies of the Semantic Web, the knowledge graphs for example, are based on a set of universal standards, as set down by the World Wide Web Consortium (W3C) international community [17]. From

the point of view of semantic technology, the key standards that apply are the Resource Description Framework (RDF) and OWL (Web Ontology Language).

RDF(S) [18], or triplets, is the format used to store data in knowledge graphs. OWL [19] is based on the Description Logics language which is designed to show the data schema and to represent rich and complex knowledge about hierarchies of things and the relations between things. It is complementary to RDF and allows for formalizing a data schema/ontology in a given domain of knowledge, separately from the data itself.

In the general case an ontology is a formal specification that provides sharable and reusable knowledge representation. An ontology includes descriptions of concepts and properties in a concrete domain of knowledge, relationships between concepts, constraints on how the relationships can be used and occasionally individuals as instances of concepts.

Figure 1 partially shows the ontology [11] that is used in the Semantic Educational Web Portal [7]. In Figure 1, the Training_Course class is intentionally highlighted, because later this class and its individuals will be used as explanatory examples.

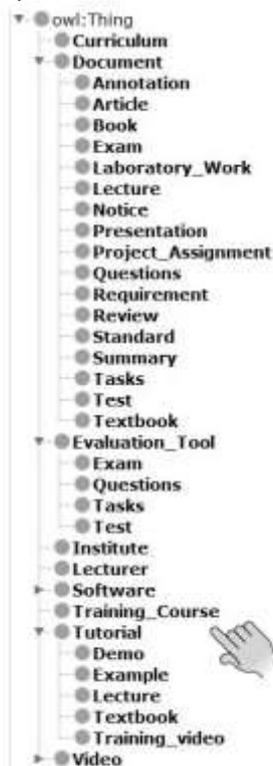


Figure 1 The class hierarchy in the ontology

Often, ontologies are understood as special knowledge repositories that can be read and understood both by people and computers, alienated from the developer and reused. Ontology in the context of information technology is usually a hierarchical system of concepts and terms (structure, model) of a certain subject area. Informally, an ontology is a description of the world view as applied to a particular area of interest. This description consists of terms and rules for the use of

these terms, limiting their meaning within a particular area. At the formal level, an ontology is a hierarchical system consisting of a set of concepts and a set of assertions about these concepts on the basis of which it is possible to describe classes, relations, functions, and individuals (instances of classes).

In the language of Description Logics (DL) [4], a set of assertions of a general kind, or terminology, is called TBox (intensional knowledge). It is TBox that forms an ontology in the proper sense of the word. In Description Logics, sets of assertions of an individual kind – ABox (extensional knowledge) are singled out separately. TBox together with ABox forms a meaningful knowledge base (knowledge graph).

Below in Figure 2 is an example of the relationship between the class and individuals. Here the individual named Semantic_Web belongs to the class named Training_Course. In addition, this individual has a number of relations with individuals of other classes. This can be a relations of different types and directions, as can be seen from the color and direction of the arrows in Figure 2.

The very kinds of relations, like classes, are usually defined in TBox, whereas the facts of the existence of a certain kind of relationship between concrete individuals are intrinsically some RDF-assertion in ABox and each assertion has a triplet appearance.

Below Figure 3 shows a diagram of the relationship between classes from the ontology. This diagram presents only the top-level relationships. Every beam of particular color is a set of relations between individual instances of two classes.

Each individual relation in the ontology (that is in the knowledge graph) inherently is an RDF assertion where the subject is an instance of one class, the object is an instance of another class, and the reference is a predicate in the RDF format.

Depending on the number of relations between instances of two classes, every beam on diagram in Figure 3 can be thicker or thinner and gets a color of the class with a large number of incoming relations. These relations can be in both directions (incoming, outgoing). The number of relations (links) between classes from the ontology is shown in the legend on the diagram in Figure 3.

4 Knowledge Graphs

An ontology enriched with extensional knowledge from a specific subject area is also called the knowledge graph or knowledge base. Extensional knowledge forms the contents of ABox. Practically, knowledge graphs are deployed in the graph database or in a different semantic repository (triplestore or RDF store).

Specifically, the Semantic Educational Web Portal [7] is located on the Ontotext S4 GraphDB cloud platform [11] (physically on the Amazon Web Services – AWS cloud platform [1]).

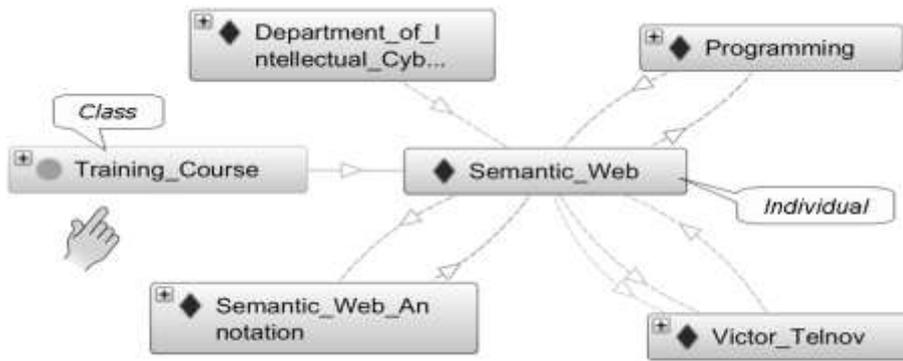


Figure 2 Individual of the class with neighborhood (example)

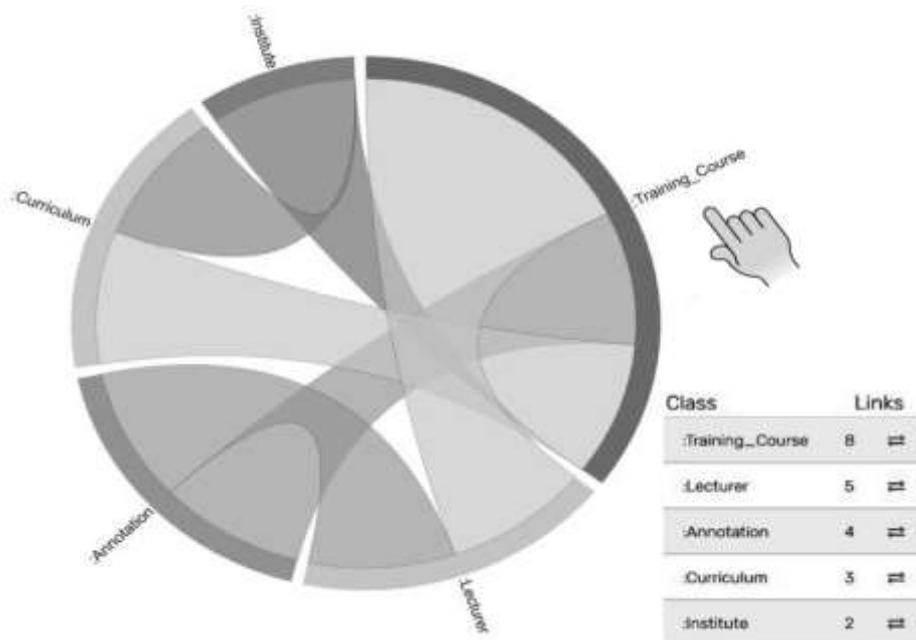


Figure 3 Relationships between classes in the ontology

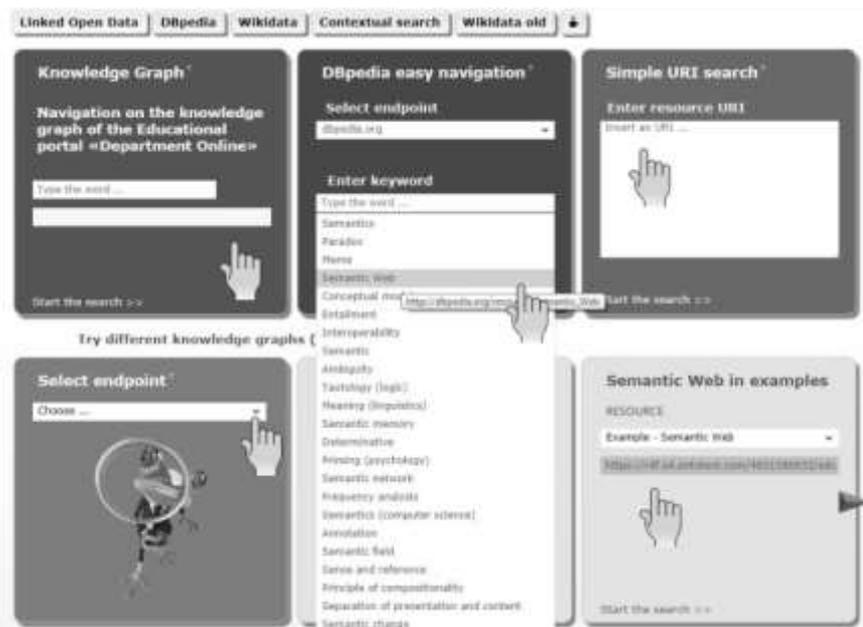


Figure 4 Navigation on the knowledge graphs

The current prototype of the Semantic Educational Web Portal [7] supports the curriculum presented in the Cloud cabinet of the Educational portal «Department online» [2]. Remote work with cloud version Ontotext GraphDB is carried out through the provided REST API. The most common operations are creating, reading, loading, and deleting semantic data. For the practical implementation of network requests HTTP methods are used, such as GET, POST, PUT, DELETE. These network requests contains essentially automatically generated SPARQL queries of the following types.

- SELECT to fetch data from the knowledge graph.
- CONSTRUCT to create a new RDF graph.
- INSERT to add triples to a graph.
- DELETE to remove triples from a graph.

Figure 4 shows the user interface of the Semantic Educational Web Portal, suitable for navigating through the available knowledge graphs. Primary graph of knowledge contains the materials of the master's courses, which are taught at the NRNU MEPhI on the profile «Computer networks and telecommunications». It is

supplemented by the international knowledge bases DBpedia and Wikidata, as well as a number of more specialized knowledge repositories.

Each of mentioned knowledge graphs contains tons of triplets. The widgets shown below in Figure 4 are designed to allow a student or teacher easily get into the right place of the right knowledge graph, where it is likely find the required learning objects.

The principle of working with these widgets is largely similar to how information is searched through popular public search engines (Google, Yandex, etc.). As the user types the letters of the keyword in the input line, the system rolls out a list of relevant concepts from the knowledge graph. User can choose the most suitable concept and dive directly into the desired area of the graph.

Then, more accurate visual navigation on the knowledge graph becomes possible, which is performed in an intuitively clear manner using the RDF browser, as described below.

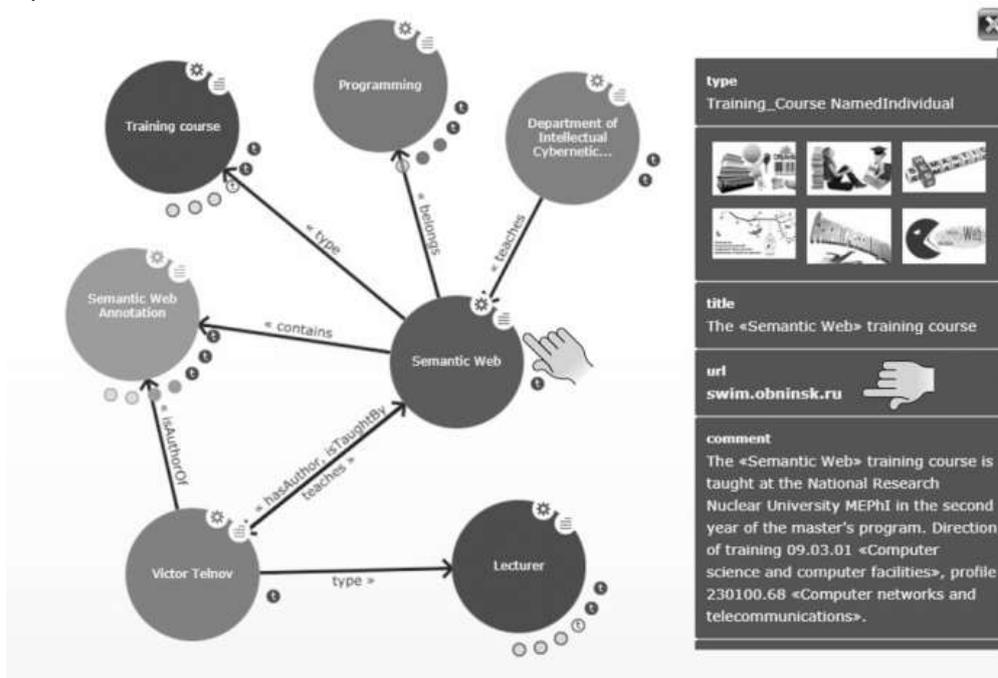


Figure 5 Fragment of the knowledge graph in the RDF browser (example)

5 RDF browser

RDF browser is the main highlight of the Semantic Educational Web Portal, which distinguishes it from most of the known solutions in the field of the Semantic Web. Having got to the right place of the necessary graph of knowledge through the corresponding widget, then you can perform a visual navigation in this graph, simply walking along its nodes.

By focusing on a specific graph node, it is possible to obtain text metadata, media content and hypertext links that are associated with this node. It is very important that

the nearest neighborhood of the node becomes visible and accessible for navigation. This environment includes nodes not only of that graph, through which you originally has come in the semantic web, but also the nodes of all other knowledge graphs of that are supported by the system.

In Figure 5, some elements of the node's neighborhood that correspond to the Semantic_Web individual are displayed, as well as some related metadata. If you focus on the next node that is displayed by the RDF browser, it also becomes available with its neighborhood and metadata.

Thus you can to make a visual walk through the graph of knowledge as long as you like in any direction, scooping up the data that appears. In Figure 5, this is not shown, but in reality, when you hover over different sections of a particular node, pop-up menus, additional information and prompts for various options for continuing navigation through the knowledge graph becomes available.

6 Adaptive Learning Technology

The main challenge of e-learning systems is to provide training courses tailored to different students with different learning rate and knowledge degree. Adaptive learning technologies are based on the fact that each student is unique, learns at varying rates and comes with different levels of knowledge. Traditional methodology of instruction may force the student down a learning path that is either too elementary, resulting in lack of interest or too heavy to grasp the nuances of the course. Adaptive learning, aided by semantic technologies [8], considers learner’s interaction with courses and assessment modules to create personalized learning paths.

The adaptive learning system generally includes the following three subsystems.

1. The subsystem of forming a model of the learner (student model).
2. Learning planning subsystem (instructional model).
3. A subsystem for evaluating training outcomes.

For the student model the most popular means of determining a student’s skill level is the method employed in CAT (computerized adaptive testing). In Semantic Educational Web Portal «Department online» [5] various, not just computerized means for measuring a student’s skill level are used. In fact, the same training course should be built in different ways, depending not only on the level of knowledge and abilities of students, but also on the learning objectives. For example, a training course in programming will look different for students who concentrate in the field of business informatics and in the field of computer networks.

To build the actual instructional model and to fill it with learning objects, the Ontotext S4 Text Analytics RESTful service [13] is actively used. The purpose of text analysis is to create sets of structured data (machine-readable facts) out of heaps of unstructured, heterogeneous documents. Text analytics involves a set of techniques and approaches towards bringing various textual content to a point where it is represented as data and then mined for insights/trends/patterns. Contextual authoring provides lecturers with related texts, images and concepts which enhance the training course, reduces the time and costs of authoring and editing new learning content. Automated content enrichment improves the quality of curriculum and allows for continuous authoring without interruption.

When constructing adapted training courses, they are usually optimized according to two criteria: the effectiveness and adaptability of training. From the mathematical point of view, in the idealized case the

problem can be reduced to finding the shortest path in the knowledge graph. This question has been studied, for example, in [9]. Neo4j Graph Database [10] has built-in means for calculating the shortest paths in the graph.

In real educational practice, the process of constructing a specific training course in the process of formation of a curriculum largely is empirical procedure, based on the experience and knowledge of the lecturer. In order to assess the training outcomes and learning efficiency, the evaluation tools from the Cloud cabinet of the Educational portal “Department online” are used, see [2].

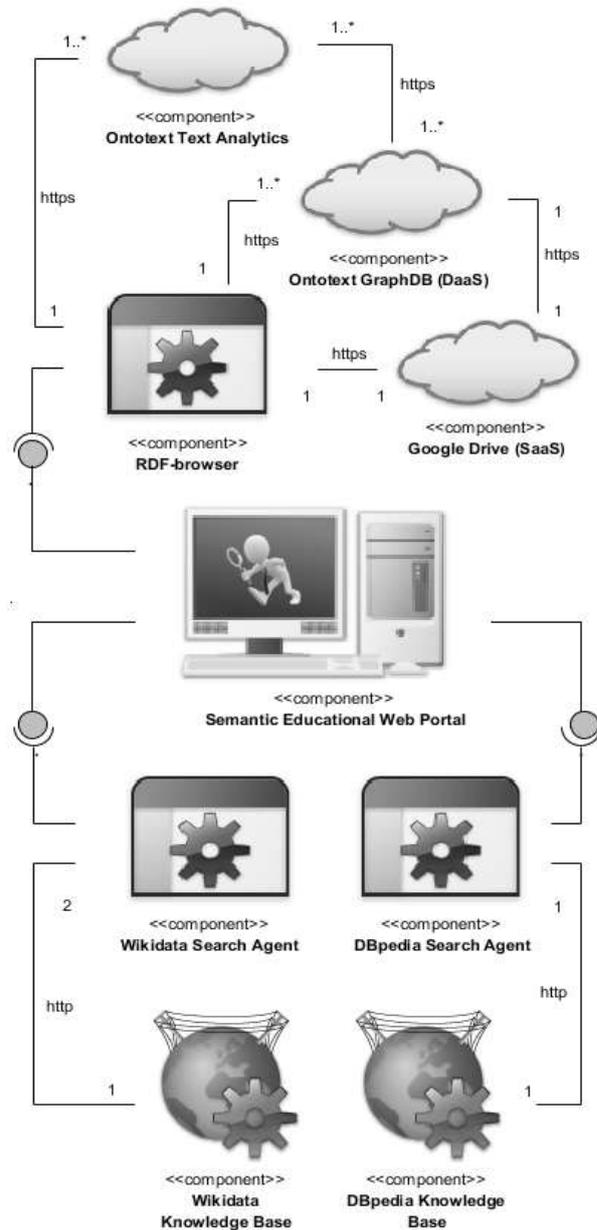


Figure 6 Software architecture

7 Software Architecture

Figure 6 shows the Deployment Diagram for the Semantic Educational Web Portal, performed in accordance with the UML 2 standard [6]. This diagram

can also be considered as an enlarged Component Diagram for this software. As it can be seen from Figure 6, the component named “RDF Browser” does not have its own server code (back-end). It interacts with two cloud RESTful services – Ontotext Text Analytics [13] and Ontotext GraphDB [11], both physically are deployed on the Amazon Web Services (AWS) [1] cloud platform. Cloud service Text Analytics [13] provides tools for semantic annotation and update of educational objects during the creation and adaptation of curricula. Cloud service GraphDB [11] provides the semantic storage for knowledge graphs and is mainly used as a SPARQL endpoint. As a universal repository for educational objects of an arbitrary nature, Google Drive is used. The choice of this particular storage is not principled, in parallel with it, arbitrary remote repositories equipped with data display means, for example such as Microsoft OneDrive or Yandex.Disk can be successfully applied.

The other two components, named “Wikidata Search Agent” and “DBpedia Search Agent” both are advanced SPARQL endpoints to the corresponding international knowledge bases. Both mentioned components are provided with libraries of patterns of search queries, which largely facilitate the work of users, as well as are capable to deliver and show the found content in a variety of formats, including graphics.

8 Discussion

The pilot project presented in this article is aimed not only to provide students and teachers with a flexible knowledge management tool, but also to stimulate them to get acquainted with the world of semantic technologies.

To the middle of 2017 a sufficient toolkit for working with ontologies, knowledge graphs and semantic repositories of triplets, including on cloud platforms, has already been created. There is a great variety of public SPARQL endpoints. The English segment of the World Wide Web is filled with Linked Open Data. This is mainly reference data, bibliographic, media and other information of encyclopedic nature.

Attempts to find the open semantic data in the Russian segment of the World Wide Web infrequently lead to success. We have to agree with the fact, that in Russia there are still little Linked Open Data, suitable for educational activities. The main sources of data for Russian users of the semantic web are still international knowledge bases, including Russian-language content, primarily DBpedia [3] and Wikidata [15]. The prototype of the semantic educational web portal created is intended to partially fill this gap.

9 Concluding Remarks

A well-known skepticism about the fact that semantic educational portals will soon become widespread in the university environment seems fair. The modern realities of higher education are such that the overwhelming number of students and teachers do not suspect the

existence of the Semantic Web and Linked Open Data. They continue to use traditional Content Management Systems (CMS), also known as Learning Management Systems (LMS) or Virtual Learning Environments (VLE), which are built primarily on simple taxonomies and thesauruses.

Students and professors widely practice searching the information on the World Wide Web for keywords, using public search engines for this purpose. Tradition plays a significant role here, as well as the simplicity and high speed of the search query generation, in comparison with the search queries to the Semantic Web.

Despite the growing commercialization of the public search engines, it can be assumed with a great deal of certainty, that they, along with Wikipedia, will remain the most accessible “universal textbooks” for the foreseeable future for that numerous category of students who not always demand the quality and completeness of the training material. An exception to this situation could be students (undergraduates) of universities who specialize in computer science and informatics.

10 Acknowledgements

The work was supported by the NBO “Vladimir Potanin Charity Fund”, project No GK160001360.

References

- [1] Amazon Web Services (AWS) – Cloud Computing Services (2017). <https://aws.amazon.com/>
- [2] Cloud cabinet of the Educational portal «Department online» (2017). <http://cloud.obninsk.ru/>
- [3] DBpedia (2017). <https://ru.wikipedia.org/wiki/DBpedia>
- [4] Description Logics (2017). <http://dl.kr.org/>
- [5] Educational portal “Department online” (2017). <http://ksst.obninsk.ru/>
- [6] ISO 19505 UML Part 2 Superstructure (2012). <https://drive.google.com/file/d/0B0jk0QU2E5q9NVlwMFNIEGxOZVU>
- [7] Knowledge graph of the Educational portal “Department online” (2017). <http://semantic.obninsk.ru/>
- [8] Learning Resource Metadata Initiative (2017). <http://lrmi.dublincore.net/>
- [9] Marwah, Alian1, Riad, Jabri: A Shortest Adaptive Learning Path in eLearning Systems: Mathematical View. *J. of American Science*, 5 (6), pp. 32-42 (2009). doi:10.7537/marsjas050609.08
- [10] Neo4j Graph Database (2017). <https://neo4j.com/>
- [11] Ontology of the Semantic Educational Web Portal (2017). <http://drive.google.com/file/d/0B0jk0QU2E5q9Y0x6bTJaOEpXLWM>
- [12] Ontotext S4 GraphDB (2017). <http://docs.s4.ontotext.com/display/S4docs/Fully+Managed+Database>
- [13] Ontotext S4 Text Analytics (2017). <http://docs.s4.ontotext.com/display/S4docs/Text+Analytics>

- [14] Philip Howard: Graph and RDF Databases 2016. Market Report Paper by Bloor. <http://www.bloorresearch.com/research/market-report/graph-and-rdf-databases-2016/>
- [15] Victor Telnov: Semantic Web and Search Agents for Russian Higher Education. A Pilot Project. CEUR Workshop Proc. 1536. Selected Papers of the XVII Int. Conf. on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2015). Obninsk, Russia, pp. 195-204. <http://ceur-ws.org/Vol-1536/>
- [16] Wikidata (2017). <http://www.wikidata.org>
- [17] W3C Semantic Web (2017). <https://www.w3.org/standards/semanticweb/>
- [18] W3C RDF Schema 1.1 (2014). <https://www.w3.org/TR/rdf-schema/>
- [19] W3C OWL 2 Web Ontology Language (2012). <https://www.w3.org/TR/owl2-overview/>