Web service for the dissertation opponents selection based on ontological approach

Tatiana Shulga [0000-0002-5521-5960], Alexander Sytnik [0000-0002-1256-7253], Svetlana Kumova[0000-0002-8385-7047] and Dmitriy Isaev [0000-0002-9691-5250]

Yuri Gagarin State Technical University of Saratov
77 Politechnicheskaya street, Saratov, Russia, 410054
taiss@yandex.ru

Abstract. The report is devoted to the problem of presenting data on Russian scientists in format of linked open data. The domain of science in Russia is still poorly formalized in the form of ontologies, despite the availability of bibliographic and abstract databases, electronic scientific libraries, etc. For example, the information of dissertation councils` work published in the form of static pages or relational databases are closed for effective analysis on the web. One of the main tasks of any dissertation council is selection of official opponents for dissertation. Official opponents are appointed from among competent scientists who have publications on the topics of thesis. This report presents the first prototype of the web service for the dissertation opponents selection based on an ontological approach. This service will allow selecting candidates on the basis of their dissertation scientific specialty, city, academic degree (candidate or doctor of science) and analysis of their publications from the abstract systems and databases. In the first prototype the Russian scientific electronic library elibrary.ru is used, in the future it is possible to connect such systems as Scopus, Webof-Science, etc. Currently the web service has no analogues. Authors describe the structure of OWL-ontology which allows to present data on Russian scientists dissertations. The ontology is filled with data on more than 10,000 theses defended in Russia in 1990-2012. Presentation of data on scientists in the formats of the semantic web will allow the use of machine analysis methods in the field of scientific activity of the Russian Federation and thereby make it more open.

Keywords: Opponents, thesis, dissertation, Linked Open Data, OWL-ontology, RDF-data, Semantic Web

1 Introduction

One of the modern concepts in the field of web technologies is the concept of "open data". It involves the data publication in special machine-readable formats without restrictions of copyright, patents and other control mechanisms. The governments of the United States and many European countries are actively introducing this concept into life publishing in open formats the data processed in the public administration sector, including data related to the education system and science (for example, [1,2]). Web

services for the provision of scientific publications are being actively developed based on an ontological approach. For example, the scientific and educational social network "Socionet" [3] provides opportunities for semantic search and analysis of scientific papers. The system is based on the Socionet ontology. Another example of the ontologies using in the scientific field is the electronic search system PubMed [4] which has a huge publication base (more than 4,000 publications from around the world) on medical subjects.

In recent decades the Russia is actively going through the process of informatization of research activities. In particular, a large number of bibliographic and abstract databases, electronic scientific libraries appear. However, this area is still weakly formalized as ontologies. First of all, this is due to the fact that the system of scientific activity organization in the Russian Federation has quite a lot of formalisms that are absent in countries that are pioneers in the field of ontological modeling.

For example, such field of activity as the work of dissertation councils in Russia is still sufficiently closed for effective analysis on the web. Data on the work of dissertation councils are published either in the form of static web pages or generated on the basis of relational databases and there are no special web applications for analyzing this data. One of such applications can be a system for selecting opponents to defend a thesis. This paper discusses the development of a web service for the selection of official opponents based on an ontological approach.

2 Main Functions of the Web Service for the Official Opponents Selection

In the normative documents regulating the process of defending dissertations in Russia [5], it is indicated that the main role in the assessment of the dissertation belongs to opponents. So their selection and appointment is one of the main tasks of any dissertation council. Note that official opponents are appointed from among competent scientists who have publications on the topics of thesis and who have given their consent to this.

The proposed system of selection of opponents will allow for the data entered (the thesis topic, scientific specialty, keywords, branch of science, etc.) to find a scientist for the role of an opponent in the defense of the dissertation. It is assumed that the system will perform the following main functions:

- search for potential opponents based on the analysis of their (name, scientific specialty, branch of science, city, academic degree - candidate or doctor of science, etc.);
- search for potential opponents based on the analysis of publications from abstract systems and databases (in the first prototype the Russian scientific electronic library elibrary.ru will be used [6], in the future it is possible to connect such systems as Scopus, WebofScience, etc.) ;

- search for data of a specific scientist (the scientific specialty of their dissertation, degree (candidate or doctor of science), and the branch of science, his publications) for its suitability as an opponent of the stated dissertation.

Obviously, the solution of these problems involves the integration of data from various sources (information about scientists and their publications). To solve these problems it is proposed to use the ontological approach which is one of the central ideas in the concept of the semantic web and allows you to get a number of advantages when using web applications, in particular, the possibility of automatic data integration, openness, extensibility, which is difficult to implement using classic relational databases data.

Therefore, it was decided to develop an ontology that will allow the submission of data on dissertations of Russian scientists.

3 Ontology "Russian Scientists Theses"

The developed ontology is intended to provide data on dissertations defended in Russia for the degree of candidate or doctor of science. The peculiarity of the thesis defense in the Russian Federation is that each dissertation is defended in a specific scientific specialty (or in two specialties) from the approved list [7]. Each scientific specialty has a code and a name (for example, 05.13.18 Mathematical modeling, numerical methods and program complexes), and also corresponds to a branch (or several branches) of sciences (for example, technical, physical and mathematical sciences, etc.)

The process of ontology developing consisted of several stages. At the first stage, a list of questions was compiled, which the ontology should answer. For example, "In what scientific specialty was the thesis defended?", "For what degree was the dissertation defended?", "In what year was the thesis defended?", Etc.

When developing ontologies, the following ontologies were taken as the basis.

1. The ontology "Specialties" [8] represents the structure of the official lists of specialties and areas of training bachelors and masters, operating in recent years in Russia, and also a list of scientific specialties. Access to the ontology can be obtained through the SPARQL endpoint [9] or through the Web application Linked Open Specialties (LOSp) [10], whose interface is presented in Russian. Also, the ontology and the corresponding set of related data are available in the open source dictionary (Linked Open Vocabulary)[11].
2. The ontology of the description of a person's personality (FOAF) allows to submit information about the name, surname, patronymic, contacts, human connections, etc.

On the basis of the formed questions and already existing classes in the selected ontologies a list of ontology classes was compiled (Fig. 1).

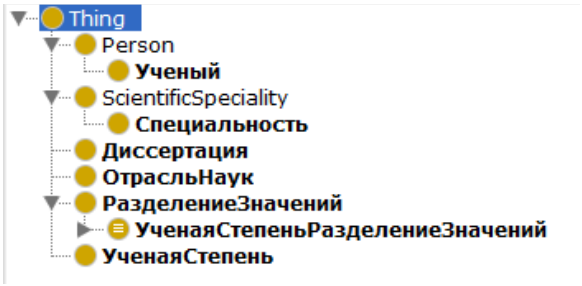

Fig. 1. The structure of the ontology classes "Russian Scientists Dissertations (Theses)".

We briefly describe the purpose of classes. The class "Dissertation" (rus. "Диссертация") is intended to present theses. Class "Scientist" (rus. "Ученный") describes a person who has been awarded a degree. Class "Degree" (rus. "Ученная степень") represents the level of the qualification system in science, which allows ranking of scientists at certain stages of an academic career. In Russia, there are two academic degrees (doctor and candidate of science). The class "Specialty" (rus. "Специальность") represents scientific specialties in which they can defend a thesis in Russia. The class "Branch of Sciences" (rus. "Отрасль наук") describes the scientific branch to which the specialty belongs.

The object properties and the data type properties of the ontology presented in Table 1 and 2.

Table 1. Ontology Object Properties

Property name	Comments
haveacademicdegree	Binds classes "Dissertation" and " Degree." Describes a degree on which the dissertation was defended
wasdefended	Connects classes "Dissertation" and "Scientist". Describes a researcher who defended his thesis
defendeddissertation	Connects classes "Scientist" and "Dissertation". Describes the thesis that the scientist defended
protectedbySpecificity	Connects classes "Dissertation" and "Specialty". Describes the specialty for which the dissertation was defended
referstobranchofscience	Associates classes "Dissertation" and "Branch of Sciences." Describes the branch of science in which the dissertation was defended

Table 2. Data Type Properties of the ontology

Property name	Комментарий
Surname	Describes the last name of a thesis researcher (line)
Introduction	Describes the introduction of the thesis (long line)
CityofDefense	Describes the city of thesis defense (string)
Conclusion	Describes the dissertation conclusion (long line)

NamePatronymicname	Describes the name and patronymic of a thesis researcher (line)
Numbero pages	Describes the number of pages of the thesis (integer)
Title	Describes the title of the thesis, the branch of science to which the thesis relates (line)
Content	Describes the content of the thesis (long line)
YearofDefense	Describes the dissertation year (year)

The presented ontology was developed in the ontology editor Protégé.
The next stage of development is the filling of ontology. To fill the ontology with data about dissertations of Russian scientists, it was written a parser of data from the scientific electronic library of dissertations and abstracts disserCat [12]. The catalog of dissertations in all specialties is presented on the main page of the site, so the task of the parser is to unload all the dissertations that are available in the catalog and enter data about them into the ontology. The parser was written in PHP.

To fill the ontology with data the following data was obtained from the pages of dissertations on the website disserCat:

- title of the thesis,
- author of the thesis,
- the year the dissertation was written,
- the degree for which the dissertation was defended,
- science,
- place of defense of the thesis,
- the name of the specialty for which the dissertation was defended,
- the name of the specialty for which the dissertation was defended,
- introduction of the thesis,
- the content of the thesis,
- dissertation conclusion

To parse HTML pages it was decided to use the PHP Simple HTML DOM library, which allows working with html-code using jQuery selectors.

At the moment ontology is filled with data on more than 10,000 dissertations defended in the Russian Federation in 1990-2012.

4 Web Service Examples

The presented ontology was used to develop the first prototype of a web service for the dissertations opponents.

One of the main options for using the web service is a keyword search. For its implementation, a SPARQL queries on the ontology data are formed.

The search is performed by the introduction, conclusion, content of the thesis. However, we can use filters to find an opponent by academic degree, scientific specialty and branch of science

For example, if you enter the keyword "delivery management areas", the system will display dissertations that satisfy this keyword (Fig. 3). Note that the web service has a Russian-language interface and data on dissertations are also presented in Russian. Figure 2 shows a table of search results with the corresponding columns - The Title of the Dissertation, Full Name, Specialty, Academic Degree, Branch of Science and the more details button.

Название	ФИО	Специальность	Ученая степень	Отрасль наук	
Адаптивные модели и алгоритмы маршрутизации	Перцовский Александр Константинович	Математическое моделирование, численные методы и комплексы программ (05.13.18)	Кандидат	физико-математических наук	Подробнее...

Fig. 2. Theses Search Results

When we click on the link "Details", we can view more detailed information about the dissertation: year of defense, city of defense, degree, branch of science, specialty, code of specialty, introduction, conclusion, content. An example of information about the thesis of the alleged opponent is presented in Figure 3.

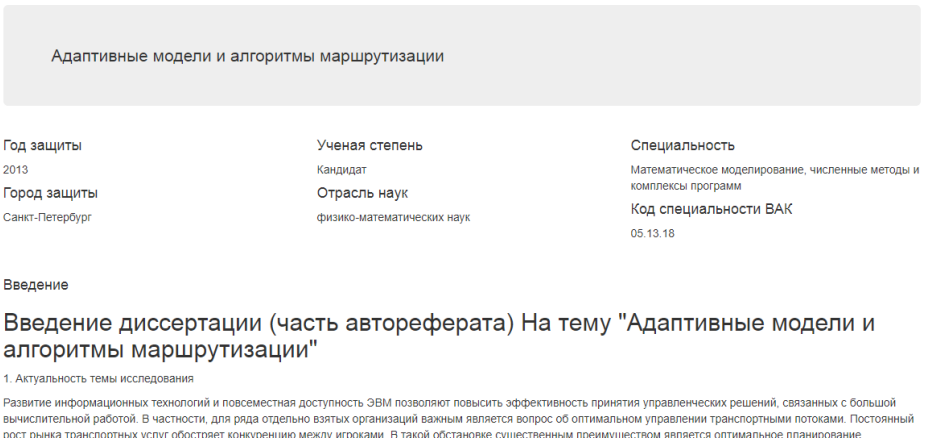

Fig. 3. Thesis Information

5 Conclusion

This paper presents the first prototype of prototype of the web service for the dissertation opponents selection in Russian. The system at the moment has no analogues.

The service was developed on the basis of the ontology "Russian Scientists Dissertations", which is currently filled with data on more than 10,000 dissertations defended in the Russian Federation in 1990-2012. In the future, it is planned to fill the ontology with data on all available dissertations in the Russian Federation based on information published by the Russian State Library.

Currently, it is planned to publish ontology on the web under an open license in one of the syntax of the OWL language. This will allow any developer to use it when creating various web applications in the field of education and science of the Russian Federation. Examples of such applications may be the websites of universities, the websites of ministries and departments, incl. website of the Higher Attestation Commission of the Russian Federation. The authors will be happy to receive comments and suggestions on the refinement and development of ontology from specialists in this subject area and are ready to cooperate.

References

1. Kelle Pereira, Crystiam & Siqueira, Sean & Pereira Nunes, Bernardo & Dietze, Stefan. (2017). Linked data in Education: a survey and a synthesis of actual research and future challenges. IEEE Transactions on Learning Technologies. 1-1. 10.1109/TLT.2017.2787659.
2. Zablith, F.; Fernandez, M.; Rowe, M. The OU linked open data: Production and consumption. In Proceedings of the Semantic Web: ESWC 2011 Workshops, Heraklion, Greece, 29–30 May 2011.
3. Scientific information space "Socionet". Available online: https://socionet.ru/ (accessed on 5 April 2019)
4. Medical Subject Headings. Available online: https://www.nlm.nih.gov/mesh/meshhome.html. (accessed on 5 April 2019)
5. Regulations on the procedure for awarding academic degrees. - Approved by the Government of the Russian Federation of September 24, 2013 No. 842
6. Russian scientific electronic library Elibrary.ru Available online: https://elibrary.ru/ (accessed on 5 April 2019)
7. On approval of the Nomenclature of scientific specialties for which scientific degrees are awarded // Order of the Ministry of Education and Science of Russia February 25, 2009 N 59
8. Sytnik A.A., Shulga T.E. Ontological engineering knowledge in the field of higher education of the Russian Federation // Engineering enterprises and knowledge management (IP & UZ-2018): collection of scientific papers of the XXI-th Russian scientific conference. April 26-28, 2018 / under scientific. ed. Yu. F. Telnova: in 2.t. - Moscow: FGBOU VO "REU them. G. V. Plekhanov", 2018.T1. Pp. 234-239. ISBN 978-5-7307-1359-8. (v.1)
9. SPARQL-endpoint to the ontology "Specialties". Available online: http://sparql.sstu.ru:3030 (accessed on 5 April 2019)
10. Web application "Specialties of higher education of the Russian Federation. Available online: http://los.sstu.ru (accessed on 5 April 2019)
11. Ontology "Specialties" in the open dictionary of related data LOV. Available online: http://lov.okfn.org/dataset/lov/vocabs/losp (accessed on 5 April 2019)
12. Scientific electronic library of dissertations and abstracts disserCat. Available online: http://www.dissercat.com/ (accessed on 5 April 2019)