

Scientific Portal of University Department – Shaping Research Area of Users through their Behavior

Mikhail Navrotskiy, Nataly Zhukova

ITMO University, Saint Petersburg, 197101, Russian Federation,
m.navrotskiy@gmail.com

Abstract. Nowadays open access to scientific data in addition to obtaining information is an essential factor for educational and research processes. A lot of universities publish open data (include scientific and educational data), but not all of them is used.

The paper dwells on the scientific portal of a university department. It contains information about the university research activity (projects, publications, employees) that university employees could use for researches. To represent scientific and educational data the ontology model is developed. This model is based on the following vocabularies: FOAF, VIVO, BIBO and Teach. In addition, a case study of implementing the proposed solution at Computer Science and Applied Mathematics Department of ITMO University is presented.

Keywords: ontology, linked open data, data integration, data search, SPARQL.

1 Introduction

Researchers in their works often deal with the search of scientific information like scientific publications, research projects, educational courses and other, which is necessary for completing many tasks such as searching new scientific discoveries and colleagues progress. For this purpose they use resources such as Google Scholar, Academia.eu, BingAcademic or other publications bases, and these bases include a modern search function. On the other hand, it would be beneficial if researches had a single start point for research, and this system could find information in different sources. One of the solutions to this problem is the technologies of Semantic Web.

Semantic Web is a Web of machine readable data, where each data source can be used in different applications [1]. One of the best solutions for publishing related data is to use the technology of linked [2] data which is based on W3C standards and technologies such as RDF and OWL.

Universities play a major role in the development of this research area. Today leading universities of Europe and the USA are developing projects using Semantic Web principles and Linked Open Data (LOD):

- LODUM, a portal supported by the University of Munster, contains educational and scientific information in RDF format;
- Oxford University publishes the results of scientific activities;
- University of Southampton Open data portal contains the results of scientific activities;
- Harvard University produces publications;
- Aalto University develops the data portal which will store data about staff, courses, research projects, publications, buildings.

Researchers could use information from LOD universities portals for their researches but, it is necessary to develop a research portal that implements the search of scientific information on the open data portals of the universities. The accuracy of the result can be improved by using the description of researchers behaviour through their research interests.

2 Background

Leading European universities have already developed LOD portals (Linked Universities). Linked Universities is an alliance of European universities engaged into exposing their public data as linked data [3]. The organization includes universities such as The Open University, University of Southampton, University of Munster and others. Portals of linked universities publish scientific and educational data such as articles (results of research activities), research projects, open courses, events. In addition, a lot of data was published from libraries (access points to the libraries data are available at datahub.io). On the one hand, LOD portals publish data in a machine-readable format, but on the other hand, the scientific community does not have a single application for scientific and educational data search for the universities open portals and other open data portals.

Russian universities do not have LOD portals. Some universities publish open data but this information is published with non-Semantic Web formats.

Researchers, of course, can use full text search engines or databases such as Google Scholar, Scopus, Academia.eu (Russian researchers can also use elibrary¹). However, these databases do not take into account the scientific interests of the user.

Thus, for developing the scientific search with the users research interests the authors should develop an ontology model. This model should include classes which will store links to data from data sources.

Portal's data sources include:

- university LOD portal (for this project it is a LOD of ITMO University [4]);
- external LOD portals (from other universities);
- external public libraries.

¹ <http://elibrary.ru/default.asp>

The Scientific portal under development has the following function: if the user finds the necessary content (a publication, a project, an employee or courses), the system displays similar content (publications, a project, an employee or courses) on the page with this content. The developed search algorithm simulates the behavior of a user. It implements search data in user's research interests and with keywords.

The search algorithm is the following:

1. the system defines the research interests of the current user;
2. the system defines a set of keywords of the current publication;
3. the system finds publications in the local ontology (maximum value of results is 5 but the user can change this value in settings);
4. if the number of results is lower than 5, the system finds publications in LOD portal of ITMO University (maximum value of results is 5 too);
5. if the number of results is lower than 5, the system finds publications in other data sources like in step 4.
6. the user can apply or cancel each publication which was found in steps 3 - 5. If the user selects to apply, relevance value is incremented by 1 otherwise relevance value is not changed. If this publication is not located in the local ontology, the system will add it.

3 Ontology Model for Scientific Portal

The developed ontology should, firstly, describe a domain, it should include classes of scientific publications, research projects, educational courses and university employees. Secondly, it should implement storage results of searching (links to data from other data sources). In addition, it should implement requests for searching publications, projects, courses, employees by the user research area and keywords.

The authors used modern vocabularies for developing the ontology model: FOAF [8–10], VIVO [11, 12], BIBO [13, 14] and Teach [16].

- FOAF ontology was used to describe employees and research projects;
- VIVO ontology was used to describe a research area, an academic status, an academic degree and an academice qualification;
- BIBO onotology describes publications;
- Teach ontology describes educational courses.

In addition, the authors described classes, properties for searching, and this onotlogy was called LODIFMO.

In fig. 1 the authors have shown part of the ontology model. The part (fragment) shows classes and properties for the process of searching publications. A full ontology includes similar classes, properties for searching projects (foaf:Project), courses(teach:Course) and employees (foaf:Person).

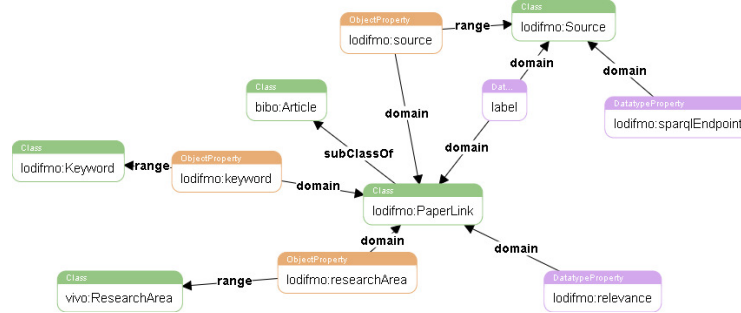


Fig. 1. Ontology model of scientific portal (Fragment)

A fragment of the ontology describes the following:

”Keyword” is a class which describes a keyword of a publication, a project, a course from the LODIFMO ontology.

”Article” is a class describing a scientific paper from the BIBO ontology.

”ResearchArea” is a class describing a research area of the user, a publication, a project, a course from the VIVO ontology.

”PaperLink” describes a scientific publication from the external data source (LODIFMO ontology).

”Source” is a class of a data source (e.g. LOD portal of ITMO University) from the LODIFMO ontology.

”label” Data Property stores a link to a publication (URI).

4 Profile of the Scientific Interests of the User

A model of the user is developed by the university corporate portal and includes general information, a description of his research activity and scientific interests. It is designed [15]:

$$P = \langle D, P_a, P_p, M, K, Apr \rangle \tag{1}$$

where D is set of descriptive characteristics of the user including an academic degree, a status, a major, competence in languages, education, experience of academic advising, P_a is a set of publications, P_p is set of research-oriented projects the user has been taking part in or has supervised, M is a set of scientific events the user participated in, K is a set of users research interests, Apr is a set of scientific profiles.

The authors profile from scientometric databases is expressed in the following way:

$$Apr = \langle K_{ape}, P_{pr}, Ind_{pr} \rangle \tag{2}$$

where P_{pr} is a set of publications in the authors profile, Ind_{pr} is an identifier of the authors profile.

The model of a publication is the following:

$$At = \langle Iz, Aat, Kat \rangle \quad (3)$$

where Iz is a journal profile, Aat is a set of publications authors, Kat is a set of thematic characteristics of publications.

The user imodel s filled with scientific interests in the following way:

$$K = K^p \cup K^a \quad (4)$$

where K^p is a set of keywords defined by the user, K^a is a set of the automatically identified keywords in compliance with their showing frequencies in the information system.

5 System Prototype

The System will perform search until the sufficient number of results is found to meet the users needs or until all the data sources are checked. The user can set the result size; a default value equals 5. For example, the users research interests include Semantic Web and Open Data. When the user searches publications on ontology, the system will not display publications related to philosophy, but only to Semantic Web.

If the results contain new objects not stored in the local ontology, the links to these objects are added to the local ontology with keywords, links to the research area and with URI from data source.

The first example is a request for search publications by a keyword and scientific interest in the local ontology:

```
SELECT *
WHERE
{
  ?publication a lodifmo:PaperLink .
  ?publication rdf:label ?link .
  ?publicaiton lodifmo:researchArea ?ra .
  ?ra rdf:label ?ra_label .
  ?publicaiton lodifmo:keyword ?keyword .
  ?keyword rdf:label ?keyword_label .
  FILTER(STR(?ra_label)="Semantic Web") .
  FILTER(STR(?keyword_label)="linked open data") .
}
```

Then system uploads data about a paper using a programming language.

The second example is a request to add a new link to the scientific object with a keyword:

```
INSERT DATA INTO GRAPH
<http://www.semanticweb.org/mikhailnavrotskiy/ontologies/2016/3>
```

```

{
  <bibo:Publication_137> lodifmo:link
    "http://svn.aksw.org/papers/2015/SEMANTICS_ITMOLOD_DEMO/public.pdf" .
  <bibo:Publication_137> lodifmo:keyword
    <lodifmo:Keyword_linked_open_data> .
  <bibo:Publication_137> vivo:researchArea
    <http://lod.ifmo.ru/ResearchArea_SemanticWeb > .
}

```

The last example is a request to search a paper by a keyword in an external source:

```

SELECT *
WHERE
{
  ?publication a <http://purl.org/spar/fabio/PosterPaper> .
  ?publication rdf:label ?title .
  FILTER(regex(str(?title), ".*(ontology)|(semantic\sweb).*")) .
}

```

The features of the scientific portal include searching different LOD sources such as AKSW LOD portal, LOD portal of ITMO University and extending datasets of local ontology with new data from external data sources.

The developed scientific portal presents a web application. The client is an HTML5 application. The server is a Rack application [6, 7] written with Sinatra (ruby web framework).

The server consists of the following modules:

1. Search Module is a module giving access to ontologies with SPARQL requests;
2. Ontology Module is a module giving access to ontology (the authors use INSERT SPARQL requests);

In Fig. 2 the system architecture is presented.

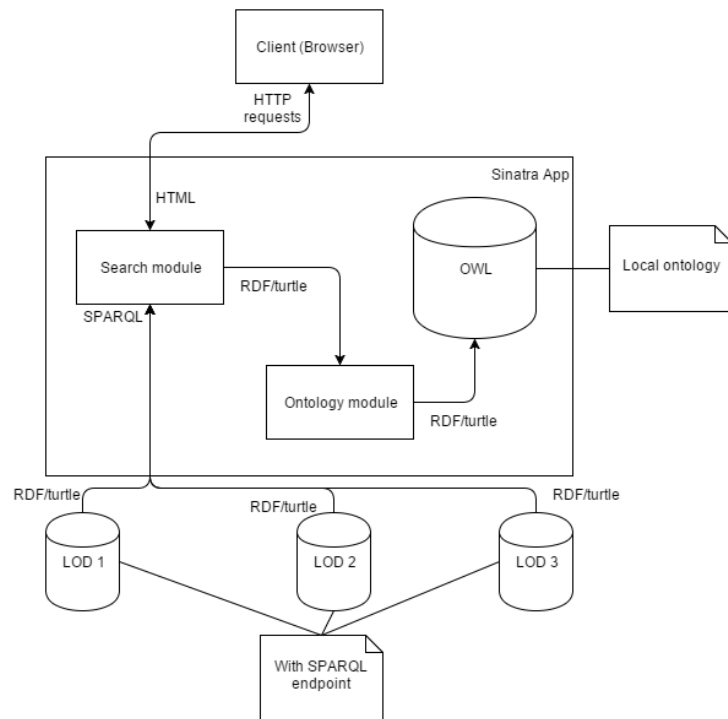


Fig. 2. Scientific portal architecture

For example, the user (researcher) searches a publication about linked open data. And they find an article about linked open data in ITMO university (Fig. 3). The system searches an article with the keywords from the current article: ontology, RDF, linked open data, data integration, data publishing; and with the user research interests: linked data, open data, semantic web. Thus, the input data are: research interests of the user (linked data, open data, ontology) and the article keywords (ontology, RDF, linked open data, data integration, data publishing) and output data are: two articles ("Towards the Linked Russian Heritage Cloud: Data Enrichment and Publishing and Metadata Extraction from Open edX Online Courses Using Dynamic Mapping of NoSQL Queries"). The Figure presents the following:

- (A) The current publication description area;
- (B) A description card with publications found through the user research areas and current publication keywords;
- (C) If a publication is correct, then the user can "apply" this result and relevance value of this article will be incremented;
- (D) If a publication is not correct, then the user can "cancel" this result and relevance value of this article will not be changed.

The screenshot displays a search results page. On the left, there is a detailed view of a publication titled "STUDY OF CURRENT APPROACHES FOR WEB PUBLISHING". This view includes the authors' names (D. I. Mouromtsev, J. Lehmann, I. A. Semerkhanov, M. A. Navrotskiy, I. S. Ermilov), their affiliations (ITMO University and University of Leipzig), and a brief abstract. A blue button labeled "PDF" is visible at the bottom of this section. On the right side, there are two search result cards. The top card, labeled "B", is titled "Towards the Linked Russian Heritage Cloud: Data Enrichment and Publishing" and lists authors Dmitry Mouromtsev, Peter Haase, Eugene Cherny, Dmitry Pavlov, Alexey Andreev, and Anna Spiridonova. It includes keywords: ontology, data publishing, open data, cultural heritage. Below the title is a green checkmark and a red 'X' icon, labeled "C". The bottom card, labeled "D", is titled "Metadata Extraction from Open edX Online Courses Using Dynamic Mapping of NoSQL Queries" and lists authors Dmitry Mouromtsev, Aleksei Romanov, Dmitry Volchek, and Fedor Kozlov. It includes keywords: ontology, open data, eLearning system, edX, semantic web technologies in education. Below the title is a green checkmark and a red 'X' icon.

Fig. 3. Example output of search

The authors developed the ontology model of the scientific portal which is used for the search process based on keywords and the users scientific interests. The developed ontology is used to store links to scientific and educational objects (projects, publications, employees, courses).

This portal is used in the work of graduate students and employees of the Computer Science and Applied Mathematics Department of ITMO University and the portal is in the status of beta testing now.

6 Conclusion and Future Work

The system was deployed on ITMO University server. Now posgraduate students of Computer Science and Applied Mathematics Department of ITMO University are using this system in their research work. It is a closed beta testing process. In the future the authors aspire to deploy this system for all the university departments, laboratories.

Some problems raised in the project require additional research and further development. The most challenging problems are:

1. Now the portal performs searching scientific publications but the authors want to add searching research projects, educational courses and researchers.
2. Increasing the number of search sources.
3. Development of a better search algorithm (for example, case based reasoning).
4. Development of more complex queries.

References

1. Leinberger M., Scheglmann S., Lammel R., Staab S., Thimm M., Viegas E. Semantic web application development with LITEQ // *Lecture Notes in Computer Science*. 2014. V. 8797. P. 212227.
2. Heath T., Bizer C. *Linked Data: Evolving the Web into a Global Data Space*. 1st ed. Morgan & Claypool Publ., 2011. 136 p. doi: 10.2200/S00334ED1V01Y201102WBE001
3. Halaç, Tayfun Gökmen and Erden, Bahtiyar and Inan, Emrah and Oguz, Damla and Gocebe, Pinar and Dikenelli, Oguz, *Publishing and Linking University Data Considering the Dynamism of Datasources, I-SEMANTICS 2013*, (2013)
4. Mouromtsev D. I., Lehmann J., Semerkhanov I. A., Navrotskiy M. A., Ermilov I. S. Study of current approaches for Web publishing of open scientific data. *Scientific and Technical Journal of Informatics Technologies, Mechanics and Optics*, 2015, vol. 15, no. 6, pp. 1081-1087.
5. Ermilov, I., Hffner, K., Lehmann, J., Mouromtsev, D. *kOre: Using Linked Data for OpenScience Information Integration, SEMANTiCS 2015*, (2015)
6. Getting data from the Semantic Web. http://semanticweb.org/wiki/Getting_data_from_the_Semantic_Web. accessed: 12-06-2016.
7. Engblom, J., *Published Linked Data in Ruby on Rails*. Royal Institute of Technology. 2012.

8. Mumtaz M. Ali Al-Mukhtar et al, International Journal of Computer Science Engineering and Technology (IJCSET), January 2014, Vol. 4, Issue 1, 10-14
9. Vacura, M., Svatek, V. Ontological Analysis of Human Relations for Semantically Consistent Transformations of FOAF Data. KIELD-2010, CEUR-WS, 2010, s. 15-27
10. FOAF Vocabulary Specification 0.99., Dan Brickley, Libby Miller, <http://xmlns.com/foaf/spec/>, 2014
11. Ding, Y., Yan, E., Ghazinejad, A., & Jia, H. (2013). Extending the VIVO ontology to iS-chools: Enabling networking of information scientists. iConference 2013 Proceedings (p. 905-908).
12. Mitchell, S., Chen, S., Ahmed, M., Lowe, B., Markes P., Rejack, N., Corson-Rikert, J., He, B., Ding, Y. and VIVO Collaboration. 2011. The VIVO Ontology: Enabling Networking of Scientists.
13. Ruiz-Iniesta A., Corcho O. A review of ontologies for describing scholarly and scientific documents. SePublica2014. May 2014
14. Peroni S and Shotton D (2012). FaBiO and CiTO: ontologies for describing bibliographic resources and citations. J. Web Semantics: Science, Services and Agents on the World Wide Web. Available online 13 August 2012
15. Varenikov D.A., Shley M.D., Muromtsev D.I. Building scientific profiles for scientific and educational process participants in it system of the university. Modern problems of science and education, 2015, vol. 2-2, pp. 178
16. Teaching Core Vocabulary Specification, Tomi Kauppinen, <http://linkedscience.org/teach/ns/>, 2013