

Web application for development of domain information space for thematic information retrieval and reuse

Sarkisova Elena,
Volgograd State Technical University,
Volgograd, Russia,
sarkisova.ln@gmail.com

This paper is devoted to a problem of creating the information space of subject domain for further information retrieval and reuse in various subject domains. An ontology-based approach to the collaborative development of the domain information space using the personal human cognitive spaces was proposed. In the framework of the proposed approach the information space ontology was constructed, the prototype of editor for development of the domain information space using the personal human cognitive space is developed.

Semantic Web, ontology, Open Science, human cognitive space, information space, information retrieval, collaborative knowledge construction

1. INTRODUCTION

One of the important stages in any research work or learning process is an investigation of the subject domain. Thematic information retrieval is a search of information resources to satisfy information need and is one of the ways to explore the subject domain. Exploration of information resources with different coverage of the subject domain, complexity level, form of presentation, which are focused on different goals and outcomes allows improving the efficiency of learning. In the scientific research the use of open information resources is important on the early stages of this process for study the state of the art in subject domain, identification of research problems and goals, statement of the specific research tasks.

The subject domain can have rather complex structure and cover a huge set of concepts and relations between them, that complicates extremely the studying these resources by the person who is not familiar with the domain. In this case, it is advisable to create a domain model during the study of the resources or to use and extend the domain models created by other persons. So, the important and urgent problem is to automate the collaborative development of the domain information space for different subject domains. This domain information space should include the description of the subject

domain structure as well as the available information resources for this subject domain.

2. THEMATIC INFORMATION RETRIEVAL

The accuracy of thematic information retrieval depends heavily on the quality of representation of document semantics in the form of document search pattern. For appropriate definition of the search pattern it is necessary to dispose of document context that defines the basic domain concepts and semantic relationships between them.

Existing approaches (using the key words and phrases, descriptors, classifiers and rubricators) do not allow to appropriately represent the domain concepts and interrelationships, and to describe the domain in terms of information needs of specific user groups.

Thus, a development of the document search pattern that adequately represents the subject domain in terms of specific needs of information retrieval subjects is a very actual modern problem.

3. STATE OF THE ART

Information support of learning implies the use of the appropriate information and educational

resources. The problem is to find the resources that achieve the learning objectives and meet the requirements of the learner among a great number of available resources, and to provide the effective personal learning trajectory to achieve required learning outcomes. The OpenCourseWare approach requires that learning content is presented as a combination of the reusable and remixable learning objects. The number of low-quality learning objects on Internet grows fast, it makes the task of content search and filtering rather challenging and labor-intensive. The efficiency of learning resources retrieval can be improved using the resources metadata annotation on the base of the appropriate metadata standards and ontological models (5) (6) (7) (8). In scientific research the electronic libraries (like ACM Digital Library, IEEE Xplore Digital Library, SpringerLink, ISI Web of Sciences) can be used to access to high quality peer-reviewed scientific papers, but they provide the access only on subscription base and do not provide the ability to perform the semantic search within subject domains. The OpenScience approach involves publishing the open research, campaigning for open access, encouraging scientists to practice open notebook science, and generally making it easier to publish and communicate scientific knowledge. The services like arXiv.org allow to publish the papers and provide the open access for them. The services like Share-Latex (9), Authorea (10) and Intech (11) provide the ability for collaborative writing and/or publishing the papers. Scientific social networks like Research- Gate (12), Academia (13) allow to share publications, connect and collaborate with specialists in the related fields and co-authors. Some software tools (like Mendeley Desktop, Zotero) allow to create personal repositories of the information resources to use it in research and cite in the following publications. During the scientific research process the domain models (conceptual models, mind maps, ontologies) are created using various software tools, some of these tools allow to associate the information resources with the components of the domain models, but none of these allows to share the resulting model with included resources and to search the models and resources created and gathered by other researches.

According (3) (4) , a cognitive space is the set of concepts and relations among them held by a human. The cognitive space can be individual as well as shared by a group of people. Using the modern software tools the cognitive space can be mapped into conceptual model represented as a mind map, topic map, concept map (conceptual diagram) or ontology. An information space (3) (4) is the set of objects and relations among them held by information system. The components of

the information space for the information retrieval task include concepts, documents, words, relations among words and documents. So the information space should be consistent with the cognitive space of particular humans or groups.

The essential and actual problem is to create the information space for some subject domain which is relevant to personal cognitive space of subject of information process.

4. COGNITIVE-INFORMATION SPACE EDITOR

In (1) (2) an approach that allows to represent the subject domain in the form of concepts and relationships between them is proposed, as well as to link these concepts with the appropriate information resources (documents). As a result a cognitive-information space of the subject domain is formed and a collection of information resources in this space is generated which meets the requirements of some user group. To evaluate the appropriateness of proposed approach a prototype of the Cognitive-information space editor was created which implements the following basic functionality (Fig. 1):

1. creation of domain cognitive space as a set of domain concepts and relationships between them;
2. creation of domain cognitive-information space by setting the association relationships between the concepts and information resources with the evaluation of resource relevance and the concept localization in the resource;
3. view of the existing cognitive-information space and its parameters.

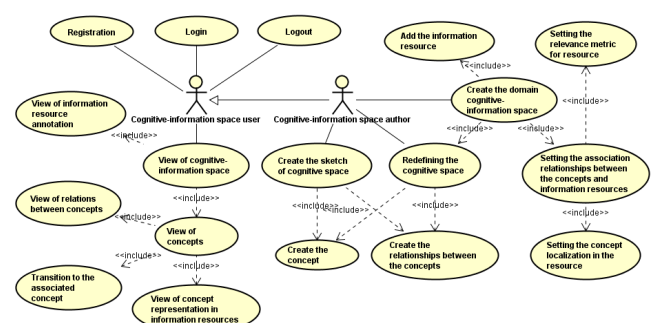


Figure 1: Cognitive-information space editor: Use case diagram

Software architecture (Fig. 2) and software tool for collaborative development of domain information space were designed, the Cognitive-information space editor was implemented as Java server application. It allows to create the cognitive space

within domains as a set of concepts and relations and fill it with the annotations of the information resources to convert into the information space (Fig. 3).

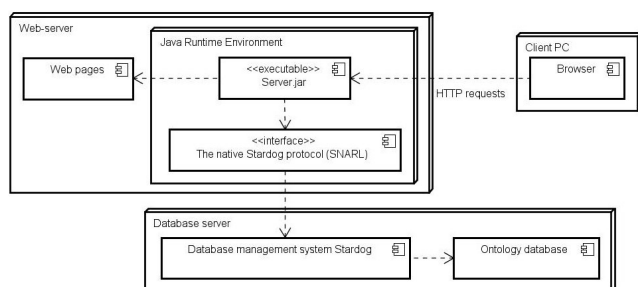


Figure 2: Cognitive-information space editor architecture



Figure 3: Cognitive-information space editor: resource annotation

5. ONTOLOGY VIZUALISATION

Cognitive-information space may include a lot of concepts and resources connected with various types of relations between them. It is not a trivial task to navigate and to search in such large schemes. Therefore, various methods and visualization tools are being developed specifically for this purpose. One of the promising visualization environment is the open-source javascript tool Ontodia, which has the following advantages. It allows to connect to the SPARQL-endpoint as a data source to separate the ontology storage and viewing/editing functions and provides collaborative development, sharing and exploring of the ontology. Also, this tool allows to visualize the ontology derived knowledge, as well as originally defined knowledge because the reasoning engine works separately from the ontology storage. Ontodia provides the possibilities for visualization of all the types of ontology elements, search and visualization of the search results, creation of ontology graph by different ways. For that reasons we chose this evolving project as a basis for the development of the cognitive-information space editor, as an aid in modeling and studying the subject area.

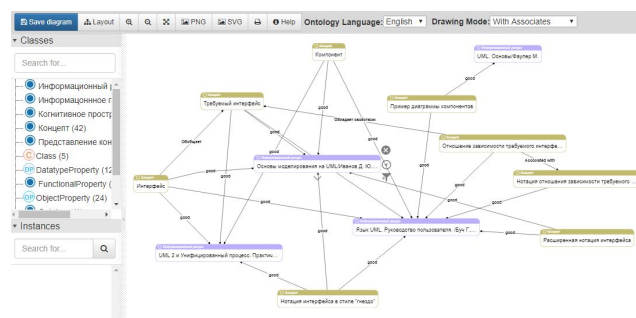


Figure 4: Cognitive-information space in Ontodia

6. CONCLUSION AND FUTURE WORK

The proposed representation of document search pattern as a fragment of cognitive-information space allows to take into account the relationships between the concepts of subject domain and to represent these concepts in the information resources in terms of information needs of different user groups.

The developed prototype of Cognitive-information space editor is planned to be used for creation of cognitive-information space in Programming languages subject domain and for manual and automated thematic search for students learning the programming in VSTU.

REFERENCES

- [1] Postroenie ontologicheskogo predstavleniya kognitivno-informacionnogo prostranstva dlja zadachi tematiceskogo poiska / D.V. Litovkin, A.V. Anikin, M.B. Kul'cova, O.N. Ljapina // Izvestija VolgGTU. Ser. Aktual'nye problemy upravlenija, vychislitel'noj tehniki i informatiki v tehniceskikh sistemah. - Volgograd, 2015. - 13 (177). - pp. 69-74. (2015)
- [2] Anikin, A., Litovkon, D., Kultsova, M.: An Ontology-based Approach to Collaborative Development of Domain Information Space. Economics and Education. Pro-ceedings of the 12th International Conference on Educational Technologies (EDUTE '16). Proceedings of the 10th International Conference on Business Administration (ICBA '16). Barcelona, Spain, February 13-15, 2016. pp.13-19 (2016)
- [3] Gregory B. Newby: Cognitive space and information space. In Journal of the American Society for Information Science and Technology, Vol. 52, Issue 12, pp.10261048 (2001)
- [4] Gregory B. Newby: Metric Multidimensional Information Space. In Proceedings of TREC-5. Gaithersburg, MD: The National Institute of Science and Technology (1996)

- [5] Brusilovsky, P.: Adaptive Hypermedia for Education and Training. In: P. Durlach and A. Lessgold (eds.): Adaptive Technologies for Training and Education. Cambridge: Cambridge University Press, pp. 46-68 (2012)
- [6] Digital libraries in education, science and culture: analytical survey. UNESCO Institute for information technologies in education. Moscow (2007)
- [7] Kogalovsky, M.R., Parinov, S.I.: Technology for Semantic Enrichable Scientific and Educational Digital Libraries. In Economical efficiency of information business systems, Proceedings. Moscow State University, 16 April, 2015. Moscow (2015)
- [8] Vasiliev, V., Kozlov, F., Mouromtsev, D., Stafeyev, S., Parkhimovich, O.: ECOLE: An Ontology-Based Open Online Course Platform. In Open Data for Education. Linked, Shared, and Reusable Data for Teaching and Learning. Springer International Publishing, pp.41-66 (2016)
- [9] <https://www.sharelatex.com>
- [10] <http://www.authorea.com>
- [11] <http://www.intechopen.com>
- [12] <https://www.researchgate.net>
- [13] <https://www.academia.edu>