An Ontology-based Approach to Collaborative Development of Domain Information Space

ANTON ANIKIN, DMITRY LITOVKIN, MARINA KULTSOVA
Volgograd State Technical University
Lenina av. 28, Volgograd, 400005
RUSSIA

Abstract: This paper is devoted to a problem of creating the domain information space for the purpose of further information retrieval and reuse in different subject domains. We propose an ontology-based approach to collaborative development of the domain information space using the personal human cognitive spaces. In framework of the proposed approach the information space ontology is constructed and the algorithm for generation of domain information space is developed on the base of personal cognitive spaces using reasoning on ontology. The application of the proposed approach is illustrated by the example of the information space construction for subject domain of programming languages.

Key-Words: Ontology, reasoning, human cognitive space, information space, information retrieval and reuse, collaborative knowledge construction.

1 Introduction

Information support of research and learning processes involves the thematic information retrieval which implies a creation of the collections of information resources for some topic on the base of their retrieval and integration [5, 6].

The specificity of the thematic information retrieval consists in the following:

- in the beginning of the retrieval the person does not realize clearly his information needs and has only general idea - the topic, so he can not create the right search query for the search engine;

- during the information retrieval the person redefines his information needs. The results of the information retrieval are not only the information resources itself, but also the clarifying the information needs of the person.

The use of general-purpose web search engines based on vector space retrieval model for solving this task is complicated by the fact that these systems, as a rule, do not take into account the semantics of the search query and the document, that reduces the quality of search results. The great disadvantage of the original vector space model is that in theory the model assumes the independence of the terms, that is not the case in the reality of information retrieval [9, 10]. The search result ranking is based, as a rule, on a document page rank and relevance of keywords in the document to the search query. This method of ranking does not always allow to assess the quality of retrieved document both on its relevance to the subject domain, and a compliance of its content and presentation with user requirements.

In addition, the user who just starting to study the subject domain is not always able to formulate the correct search query for such systems due to lack of sufficient knowledge in the subject domain. Clustering algorithms and latent-semantic indexing allow to solve some of these problems (in particular, partially mitigate the problem of polysemy and homonymy) and to identify the latent interdependencies in the set of documents, but also assess the relevance of the document based exceptionally on the lexical similarity measures. They require the large but limited number of documents, that makes difficult the use of these methods for solving the considered problem.These approaches do not allow to create and represent the model of complex structured subject domain in human-readable format.

One of the convenient tools for the thematic information retrieval is the thematic directories. These directories are focused on structured organization of the thematic collections of the information resources with the convenient system of links and resources hierarchy. This tool allows the person to find the required resource using the directory structure navigation or specific retrieval mechanism of the directory.

The significant disadvantage of the thematic directories is rigid enough structure of the directory that reflects the vision of some subject domain of some group of experts in some aspect. So 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 [2].

In this paper we propose an ontology-based approach to collaborative development of domain information space that allows to set the relations and mapping between the cognitive spaces of different persons as well as sharing and collaborative use of this information space.

Ontology-based models are widely used in modern intelligent systems. They provides the ability to describe the objects and their properties and relations in any domain, and the ontology reasoning mechanism allows to produce the new knowledge about these objects.

The ontologies can be used for the knowledge acquisition (that is essential to the knowledge engineering), knowledge sharing, knowledge reuse, verification and validation, and domain theory development [1]. For the specific tasks such as an information retrieval the ontology can provide the expressive terminology to describe the content, and inferences sanctioned by ontology can be used to improve the quality of search.

In the tasks of the learning process support [4, 7, 8] the domain model should provide the breadth and depth of knowledge and skills, granularity and scalability. The domain model should be modular and extensible so that it could cover the new subdomains, and the ontology meets all of these requirements. The processes of scientific research and learning include the tasks of information retrieval and domain theory development, so the domain models should meet the requirements mentioned above, and the ontology can be used for information support of these processes.

2 An Ontology-based Approach to the Information Space Development

According [2, 3], 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 [2,3] 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.

In accordance to the proposed approach we developed the following ontological representation of the cognitive and information spaces.

The cognitive space of some subject domain *CognitiveSpace* is defined as follows:

CognitiveSpace = <*Concepts, IncludesRelation*>, (1)

where: *Concepts* - set of the concepts of the subject domain;

IncludesRelation - set of the subordination relations defined on the set of the concepts.

The information space *InformationSpace* that is relevant to the cognitive space *CognitiveSpace*:

InformationSpace = <*Objects, Relations, Rules*>, (2)

where: *Objects* - set of the objects of the subject domain held by the information system,

Relations - set of the relations between these objects,

Rules - set of the reasoning rules for setting the relations between the objects.

The set of the objects of the subject domain *Objects* is defined as follows:

Objects = <*Concepts, InformationResources*>, (3)

where: *Concepts* - set of the concepts of the subject domain;

InformationResources - set of the information resources associated with the concepts of the subject domain.

The set of the relations *Relations* includes the two types of relations:

Relations = <*AssociatesRelations, IRRelations*>, (4)

where *AssociatesRelations* - set of the association relations between the concepts of the subject domain and information resources:

Concepts × *InformationResources* → *{undefined, bad, good, excellent}*,

where: *excellent* - the resource describes the concept in full; *bad* - the resource contains minimal information about the concept; *good* - intermediate value between the *excellent* and *bad*; *undefined* - the resource describes the concept with relevance which is not defined yet;

IRRelations - set of the relations between the information resources.

The information space can be defined using the three techniques described below or their combinations (Fig. 1):

1) Individual creation of the information space. The person defines his own cognitive space *CognitiveSpace*, then performs the information retrieval and associates the found information resources with the concepts of the subject domain and evaluates their relevance. So, the information space *InformationSpace* which is relevant to the person's cognitive space is created. The person redefines his cognitive and information spaces in course of the structuring and formalizing the subject domain, afterwards these spaces can be used by other persons.

2) Collaborative creation of the information space. The person creates the cognitive and information spaces, then provides to other persons the ability to supplement the information space by defining the new information resources and relations between resources and concepts of the subject domain.

3) Creating the information space on the base of existing information spaces, defined by other persons. The person creates the cognitive space for some subject domain *CognitiveSpace_1*, after that defines the conformity relations *CorrespondsRelations* with the concepts of the cognitive space *CognitiveSpace_2* of some other subject domain: *Concepts_1* × *Concepts_2* → *{equivalent, compatible}*, where: *equivalent* - identity relation; *compatible* - compatibility relation, that includes subordination relations as well as intersection relations [3]. Thus the association relations are defined between the concepts of *CognitiveSpace_1* and information resources of the information space *InformationSpace_2*. The person can redefine these relations later.

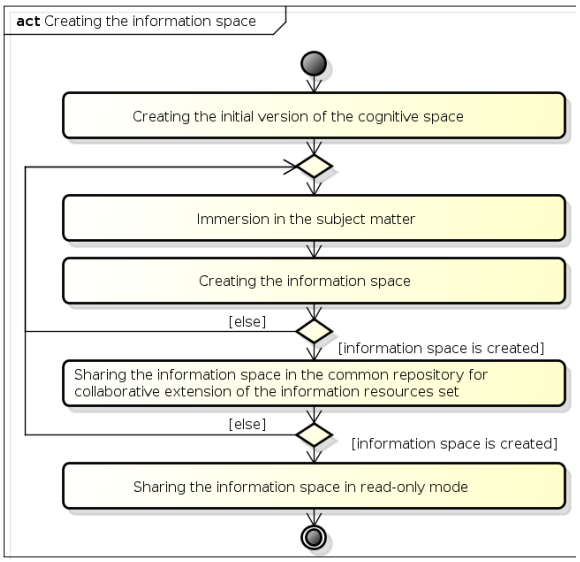

Fig. 1. Cognitive and information space creating process (UML activity diagram)

For formal representation of the cognitive and information spaces we propose to use the ontology described with OWL language (Fig. 2).

In the terms of the proposed ontological model, the creation of the information space for some subject domain boils down to creating the instance of the class CognitiveSpace and instances $concept_i$ of the class *Concept* and defining the relations "includes" between these instances, as well as the creating one instance *cognitive_space* of the class *CognitiveSpace* and relations *hasConcept* between instance *cognitive_space* and all instances $concept_i$.

Such approach allows us to describe the cognitive spaces of the individuals and groups within the common ontological model, and also create the information spaces on the their basis by setting the relations between the concepts of different cognitive spaces. The creation of the new relations between the objects of information space is carried out using logical inference on ontology. These relations can be used for inclusion in the new information space the information resources which were previously included in other information spaces.

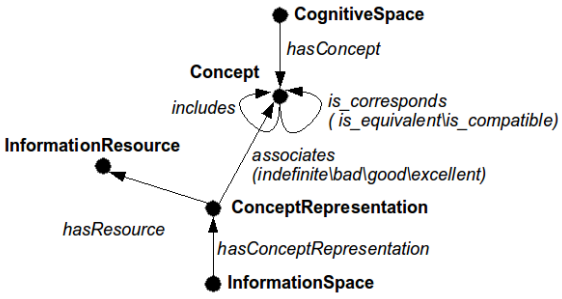

Fig. 2. The structure of the ontology of cognitive and information spaces

We applied the proposed approach to creation of the domain information space on the base of cognitive space for the subject domain of programming languages. The fragment of the cognitive space ontology for domain "Programming Language C" is shown in Fig. 3.

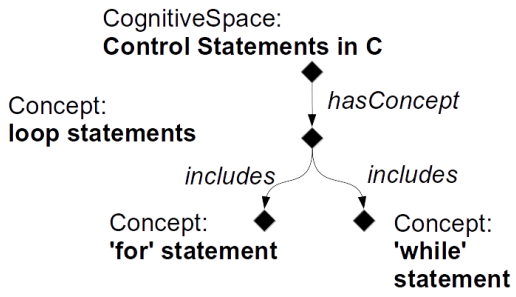

Fig. 3. The fragment of the cognitive space ontology for "Programming Languages. C" domain

In this fragment the cognitive space "Control Statements in C" is represented as an instance of class *CognitiveSpace* in corresponding ontology and includes the concepts of the subject domain "loop statements", "for statement", "while statement" which are represented as the instances of the class *Concept* of the ontology, and the relationship "includes" between them.

The creation of the information space implies defining the instances *information_resoruces* of the class *InformationResource* and one instance *infromation_space* of the class *InformationSpace*. To define the relations *associates* between the concept *concept* $\in$ *concepts* and the information resource *information_resource* $\in$ *information_resources* it is necessary to:

1) define the instance *concept_representation* of the class *ConceptRepresentation*;

2) define the relation *hasConceptRepresentation* between the instances *information_space* and *concept_representation* ;

3) define the relation *hasConceptRepresentation* between the instances *concept_representation* and *information_resource* $\in$ *information_resources*;

4) define the relation from the set *{undefined, bad, good, excellent}* between the instances *concept_representation* and *concept* $\in$ *concepts* .

The ontology of the information space for "Programming Language C" domain is shown in Fig. 4.

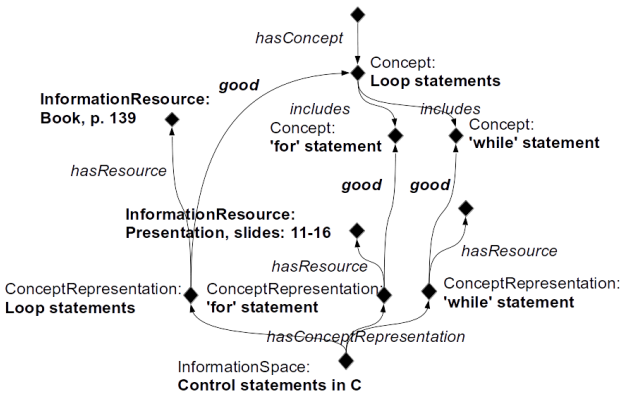

Fig. 4. The ontology of information space for "Programming Language C" domain

This fragment defines the information space "*Control statements in C*" as an instance of the class *InformationSpace* of the ontology, at that the information space is the extension of the cognitive space defined above. The information resources are defined as instances of the class *InformationResource*, also the relations between the

resources and appropriate representation of the concepts in the information space are specified, as well as the relations between the representations of the concepts in the information space (instances of the class *ConceptRepresentation*) and the concepts of the cognitive space.

The task of the determining the conformity between the concepts of the different subject domains using the ontology representation boils down to the defining the relations from the set of {*is_equivalent, is_compatible*} between instances of the class *Concept* that belong to the different subject domains.

To create the information space *InformationSpace_1* on the basis of the existent information space *InformationSpace_2* and the conformity relations between the cognitive spaces *CognitiveSpace_1* and *CognitiveSpace_2* we developed the algorithm described below using the proposed ontology representation:

1) define the instance *information_space_1* of the class *InformationSpace* for the new information space;

2) for each information resource from the *InformationSpace_2* that is relevant for concepts of the cognitive space *CognitiveSpace_2:*

2.1) define the instance *concept_representation_1* of the class *ConceptRepresentation* and

2.2) set the relation *hasResource* between the instance *concept_representation_1* and the instance of the class *InformationResource*, belonging to the *InfromationSpace_2*;

3) for each instance *concept_representation_1* defined above:

3.1) set the relation *hasConceptRepresentation* between the instances *informaton_space_1* and *concept_representation_1*;

3.2) set the relation "*associates*" with one of the instances of the class *Concept* belonging to the *CognitiveSpace_1*.

To define the relevant information resources on the step 2 of the algorithm and the relations "*associates*" on the step 4, the set of the SWRL-rules was developed. These rules allow to take into consideration the conformity relations *is_corresponds* between the concepts of the cognitive spaces *CognitiveSpace_1* and *CognitiveSpace_2* as well as the association relations between the concepts of the cognitive space *CognitiveSpace_2* and information resources of the information space *InformationSpace_2*.

The SWRL-rules for the association of the two information spaces (Fig. 6) are defined as follows:

hasConceptRepresentation (?is1, ?cr1) ^ hasConceptRepresentation (?is2, ?cr2) ^ hasResource(?cr1, ?res)^ good(?res, ?c1) ^ is_equivalent(?c2, ?c1) ^ sameAs(?cr1, ?c1) ^ sameAs(?cr2, ?c2) -> hasResource(?cr2, ?res) ^ good(?res, ?c2); (5)

hasConceptRepresentation (?is1, ?cr1) ^ hasConceptRepresentation (?is2, ?cr2) ^ hasResource(?cr1, ?res)^ good(?res, ?c1) ^ is_compatible(?c2, ?c1) ^ sameAs(?cr1, ?c1) ^ sameAs(?cr2, ?c2) -> hasResource(?cr2, ?res) ^ indefinite(?res, ?c2), (6)

where ?is1, ?is2, ?c1, ?c2, ?cr1, ?cr2, ?res - the variables of the SWRL rules.

The example of the associating the cognitive spaces is shown in Fig. 5. The resulting information space is shown in Fig. 6.

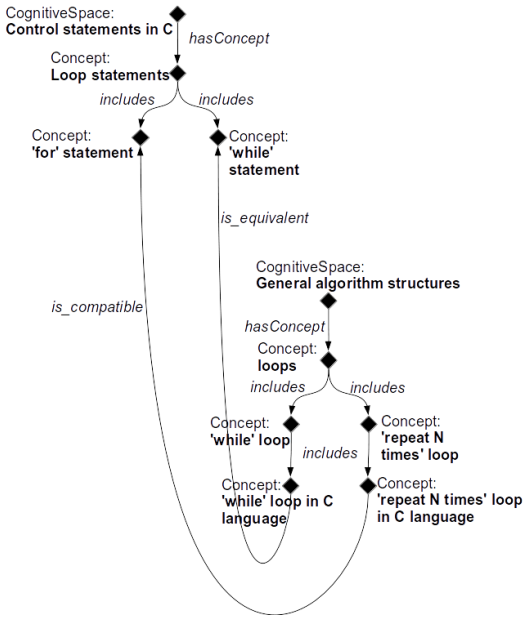

Fig. 5. The example of association of two cognitive spaces

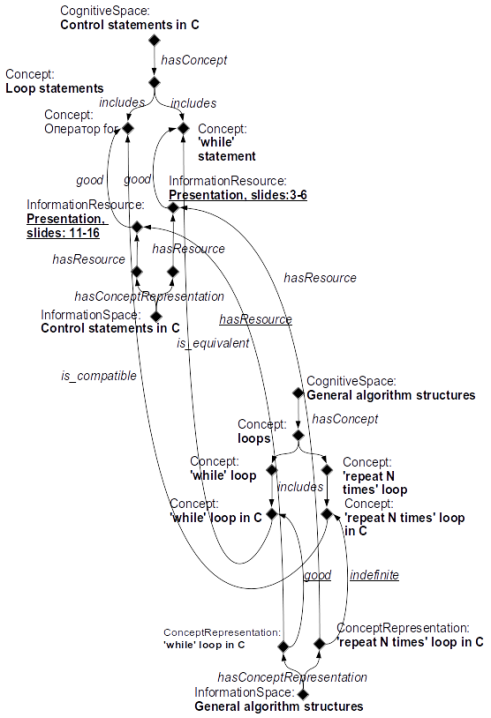

Fig. 6. The example of association of two information spaces

Thus, the relations *is_equivalent* and *is_compatible* are defined between the concepts "*while statement*" and "*while loop in C language*", and the concepts "*for statement*" and "*repeat N times loop in C language*" of the cognitive spaces "*Control statements in C*" and "*General algorithm structures*", represented by the instances of the classes in the ontology (Fig. 5). In addition, the relations for inclusion of the information resources of the information space "Control statements in C" into the information space "General algorithm structures" (Fig.6) are defined as a result of reasoning on ontology using the SWRL-rules (5), (6).

The developed ontological models and algorithms are planned to implement in web-application for collaborative creating and reusing the information spaces using the graph database Stardog 4.0, which allows to store and query the OWL ontologies, use the SWRL-rules and provides multi-user access.

3 Conclusion

The ontology-based approach to collaborative construction of the domain information space on the base of the cognitive spaces of individuals or groups and the existing information spaces was proposed, that allows to decrease the time and increase the efficiency of retrieval and reuse of the information resources which are relevant to the subject domain and the cognitive space of the information process subject.

This work is carried out in the framework of ongoing project, the main provisions of the proposed approach have been already used for the development of the software tool for distributed learning resources retrieval and creation of the personal learning collections [7,8]. The ontological model for knowledge representation was developed including ontologies of learning course domain, learning resource, learner's profile and personal learning collection. The last one included the set of semantic rules for creating the personal learning collection. The new two-stage method for electronic learning resources retrieval and integration into personal learning collection was developed based on ontology reasoning rules. Developed models, method and software tool were successfully applied for creation of information space in the form of personal learning collection for the course "Programming Languages. C++" in Volgograd State Technical University.

Further evolution of the project implies the following activities:

- extending the set of SWRL-rules for the purpose of automatic generation of information space;

- implementation of the repository of information spaces and the software tool (in form of web-application) for creation of these spaces using the graph database StarDog;

- testing and evaluation of the proposed approach and developed software tools for information support of scientific research and learning process.

Acknowledgement

This paper presents the results of research carried out under the RFBR grant 15-07-03541 "Intelligent support of decision making in management of large scale systems on the base of integration of different types of reasoning on ontological knowledge".

References:

[1]. Bench-Carpon T., 2005. Ontologies in AI and Law. Inteligencia Artificial Aplicada al Derecho. Memoria del Congreso Internacional de Culturas y Sistemas Juridicos Comparados, pp.65-84.

[2]. Gregory B. Newby, 2001. Cognitive space and information space. *In Journal of the American Society for Information Science and Technology*, Vol. 52, Issue 12, pp.1026–1048.

[3]. Gregory B. Newby, 1996. Metric Multidimensional Information Space. *In Proceedings of TREC-5. Gaithersburg, MD: The National Institute of Science and Technology.*

[4]. Dimitrios Sklavakis and Ioannis Refanidis, 2014. The MATHESIS meta-knowledge engineering framework: Ontology-driven development of intelligent tutoring systems. *Appl. Ontology* 9, 3-4 (July 2014), 237-265.

[5]. Kozlov D.D., 2004. *Reshenie zadachi tematicheskogo informacionnogo poiska v runet: autoreferat k.t.n.: 05.13.11,* Moscow State University, Moscow, Russia.

[6]. Chugreev V.L., 2003. *Model structurnogo predstavleniya informacii i metod ee tematichecskogo analiza na osnove chastotno-kontekstnoy klassifikacii: autoreferat k.t.n.: 05.13.01,* Saint Petersburg State Electrotechnic University, Saint Petersburg, Russia.

[7]. Anikin, A. et al, 2014. Knowledge Based Models and Software Tools for Learning Management in Open Learning Network. *In Knowledge-Based Software Engineering. Vol. 466 of the series Communications in Computer and Information Science.* Springer International Publishing, Switzerland, pp. 156-171.

[8]. Kultsova, M, et al, 2015. Ontology-Based Learning Content Management System in Programming Languages Domain. *In Creativity in Intelligent, Technologies and Data Science. Volume 535 of the series Communications in Computer and Information Science.* Springer International Publishing, Switzerland, pp. 767-777.

[9]. Wolfgang G. Stock, Mechtild Stock. Handbook of Information Science. Berlin, Germany, de Gruyter Saur, 2013. 901 pp.

[10]. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.