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

Anton Anikin,
Dmitry Litovkin,
Marina Kultsova

Volgograd State Technical University, Russia

Outline

I. Introduction & State of the Art

II. Ontology-based Approach
to the Information Space Development

III. Conclusion

I. State of the Art (1/4)

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**.

The spicifity of the thematic information retrieval:

- In the beginning, the person does not realize clearly his information needs
- During the information retrieval the person redefines his information needs

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

I. State of the Art (2/4)
General purpose search engines

- Do not take into account the **semantics of the search query and the document**. The vector space model assumes the independence of the terms, that is not the case in the reality of information retrieval

- **The method of ranking (PageRank)** does not always allow to assess the quality of retrieved document

- User who just starting to study the subject domain is **not able to formulate the correct search query** for SE

- **Clustering algorithms and latent-semantic indexing** allow to solve some of these problems. But they requires large but limited number of documents. These approaches do not allow to create and represent the model of complex structured subject domain in human-readable format.

I. State of the Art (3/4)
Thematic directories

- Focused on structured organization of the thematic collections of the information resources.

- But the **structure is rigid** enough and reflects the vision of some subject domain of some group of experts in some aspect.

I. State of the Art (4/4). Ontology-based models

Ontology as a model is widely used in modern intelligent systems for:

- knowledge acquisition,
- sharing,
- reuse,
- verification,
- validation,
- domain theory development,
- information retrieval tasks.

The domain model should:

- provide the breadth and depth of knowledge and skills, granularity and scalability
- be modular and extensible to cover the new subdomains

The ontology can be used for the information support of the information retrieval process in the scientific research and learning tasks.

II. The Ontology-based approach to the information space development

- **Cognitive space** is the set of concepts and relations among them held by a human. It can be individual or shared by a group of people.
  - conceptual model
  - mind map
  - topic map
  - concept map (conceptual diagramm)
  - ontology

- **Information space** is the set of object and relations among them held by information system. Components:
  - concepts
  - documents
  - words
  - relations among the words and documents

The information space should be consistent with the cognitive space of particular humans or groups.

Cognitive and Information space models (1/2)

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.

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.

Cognitive and Information space models (2/2)

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.

Relations = <AssociatesRelations, IRRelations>, (4)

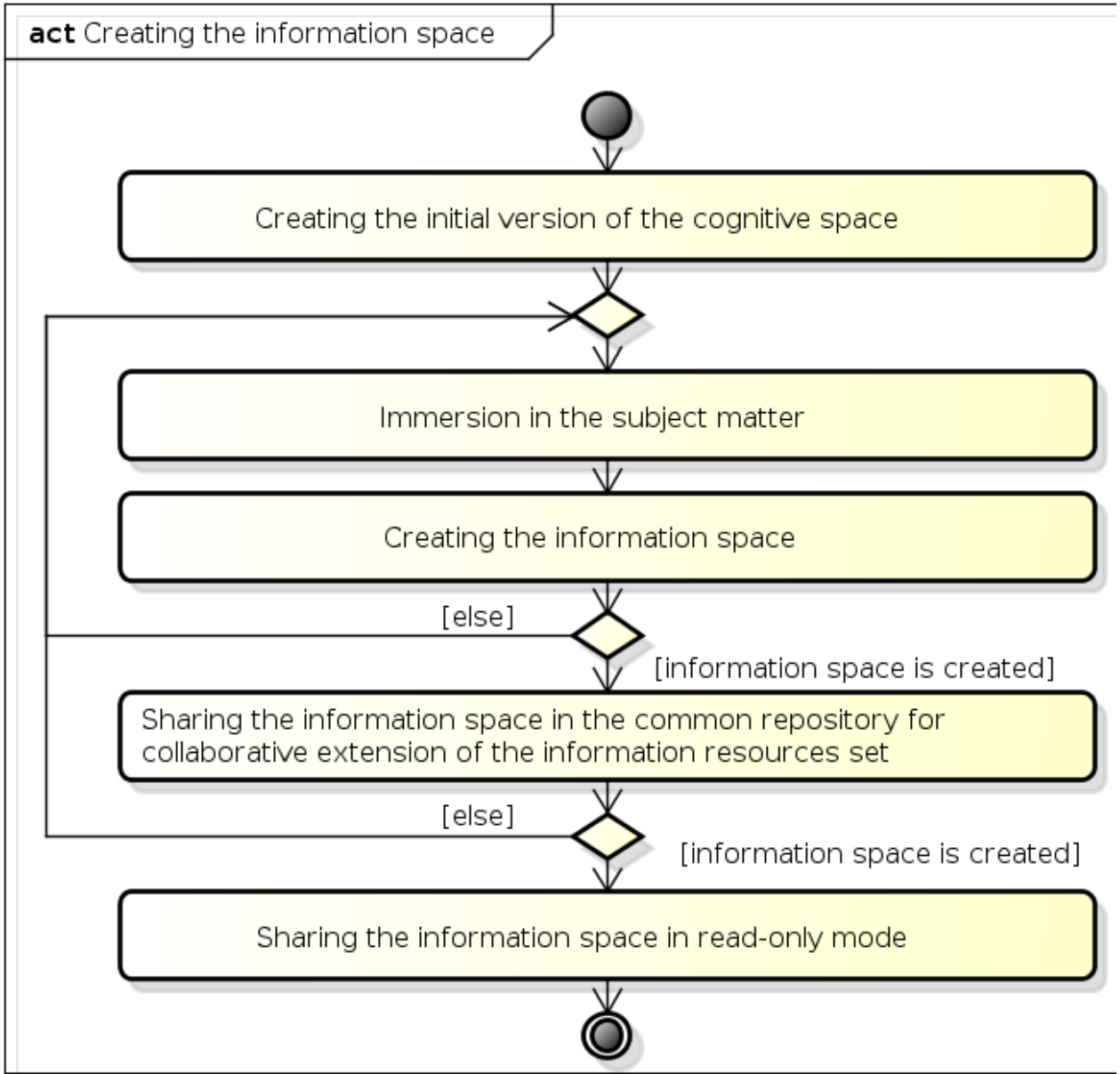
where *AssociatesRelations* - set of the association relations between the concepts of the subject domain and information resources:

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.

Cognitive and information space creating process



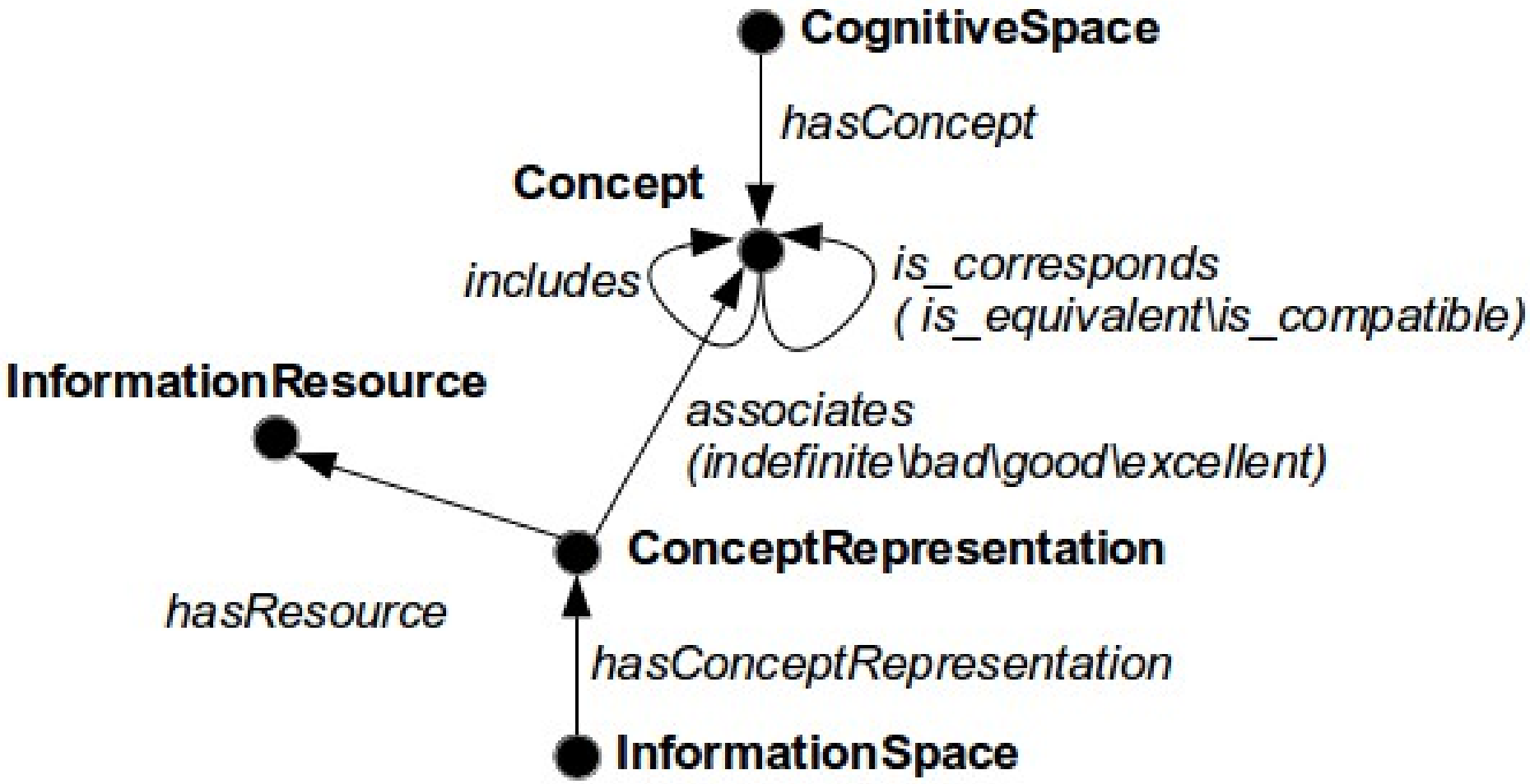
Cognitive and information space creating process scenarios

1) Individual creation of the information space

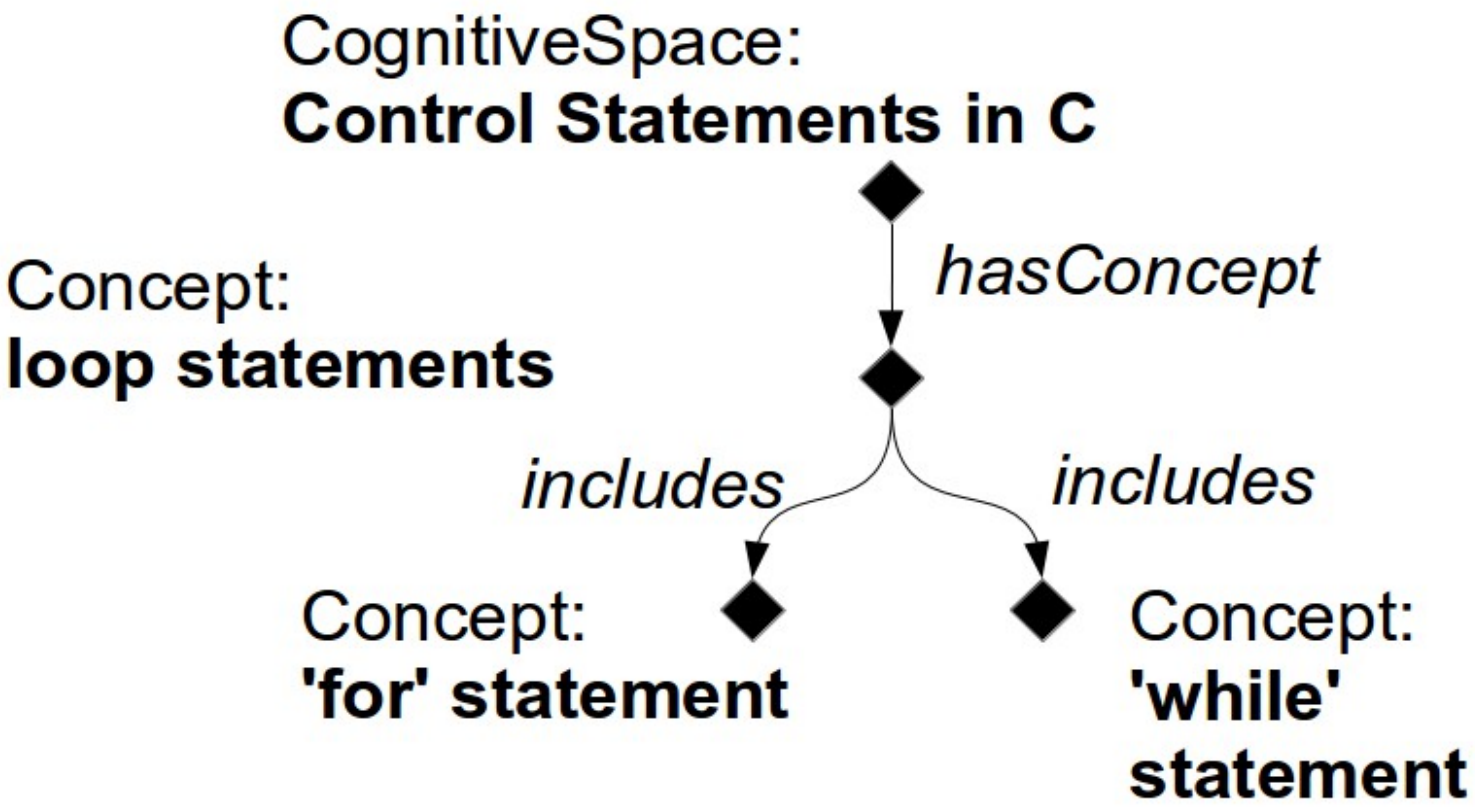
2) Collaborative creation of the information space

3) Creating the information space on the base of existing information spaces, defined by other persons

The structure of the ontology of cognitive and information spaces



The fragment of the cognitive space ontology for «Programming Language C» domain



The creation of the information space

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:

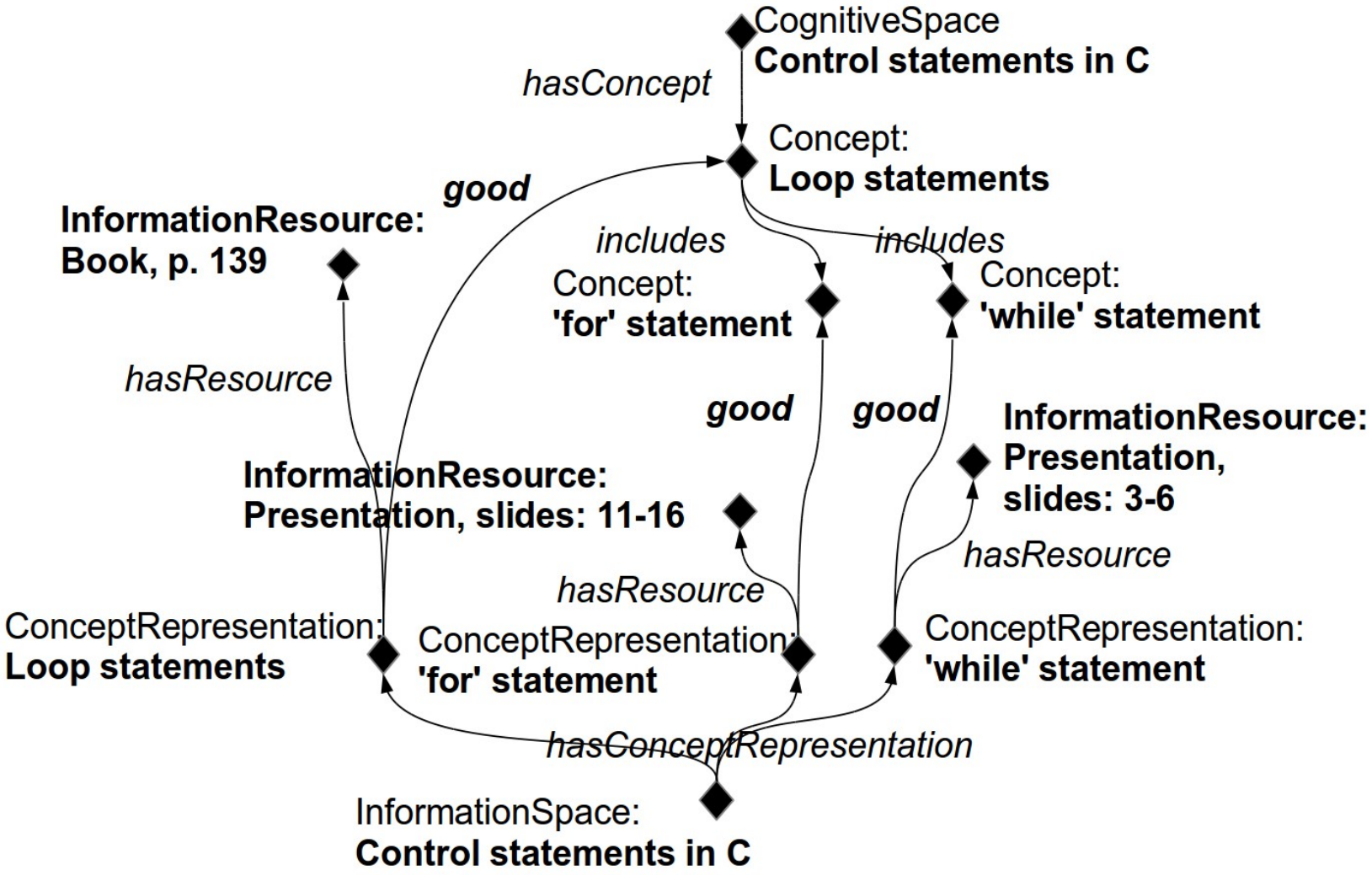
1) define the instance *concept_representation* of the class **ConceptRepresentation**;

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 information space for «Programming Language C» domain



Association of cognitive and information spaces (1/2)

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* the algorithm below can be used:

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 *InformationSpace_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*.

Association of cognitive and information spaces (2/2)

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*.

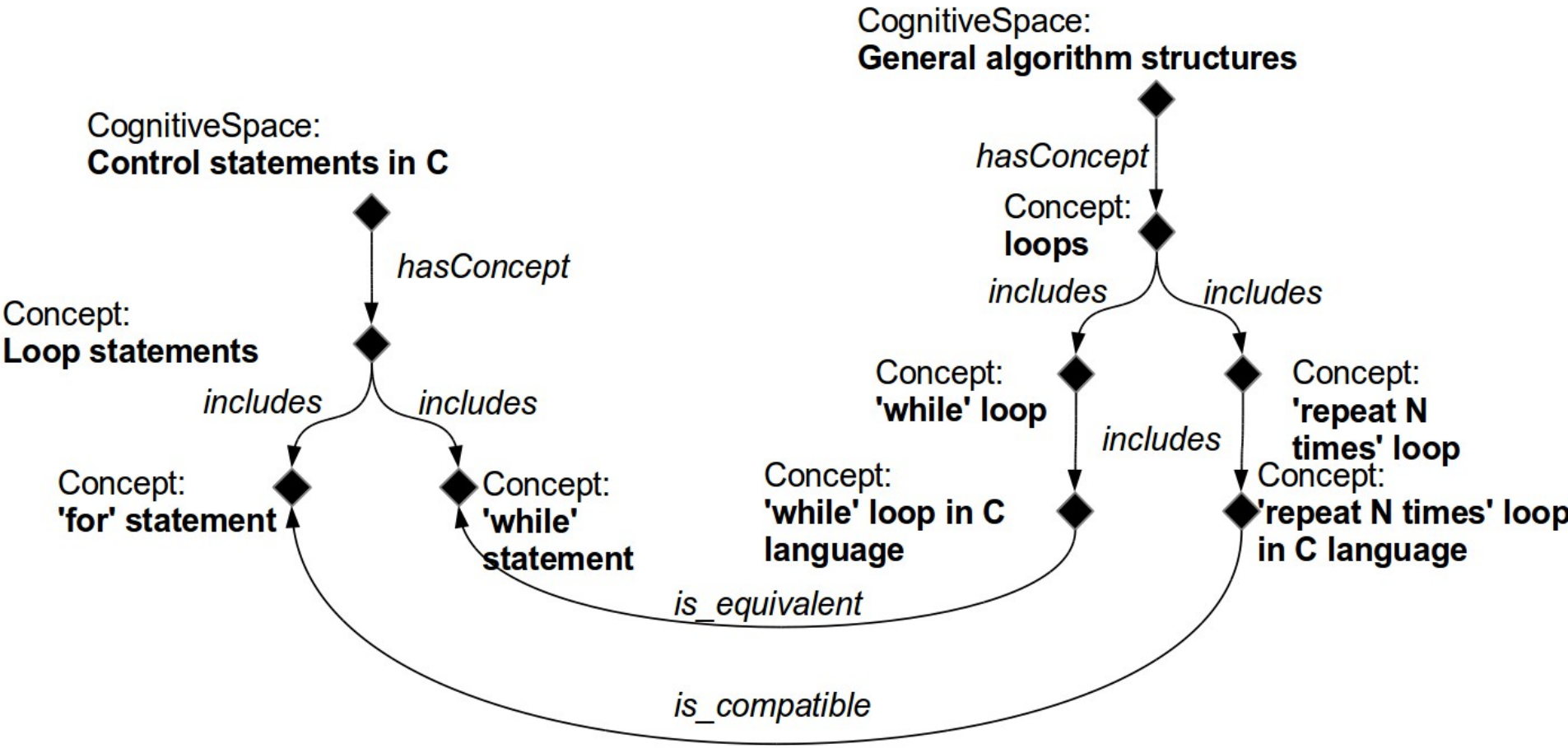
The SWRL-rules for the association of the two information spaces:

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)

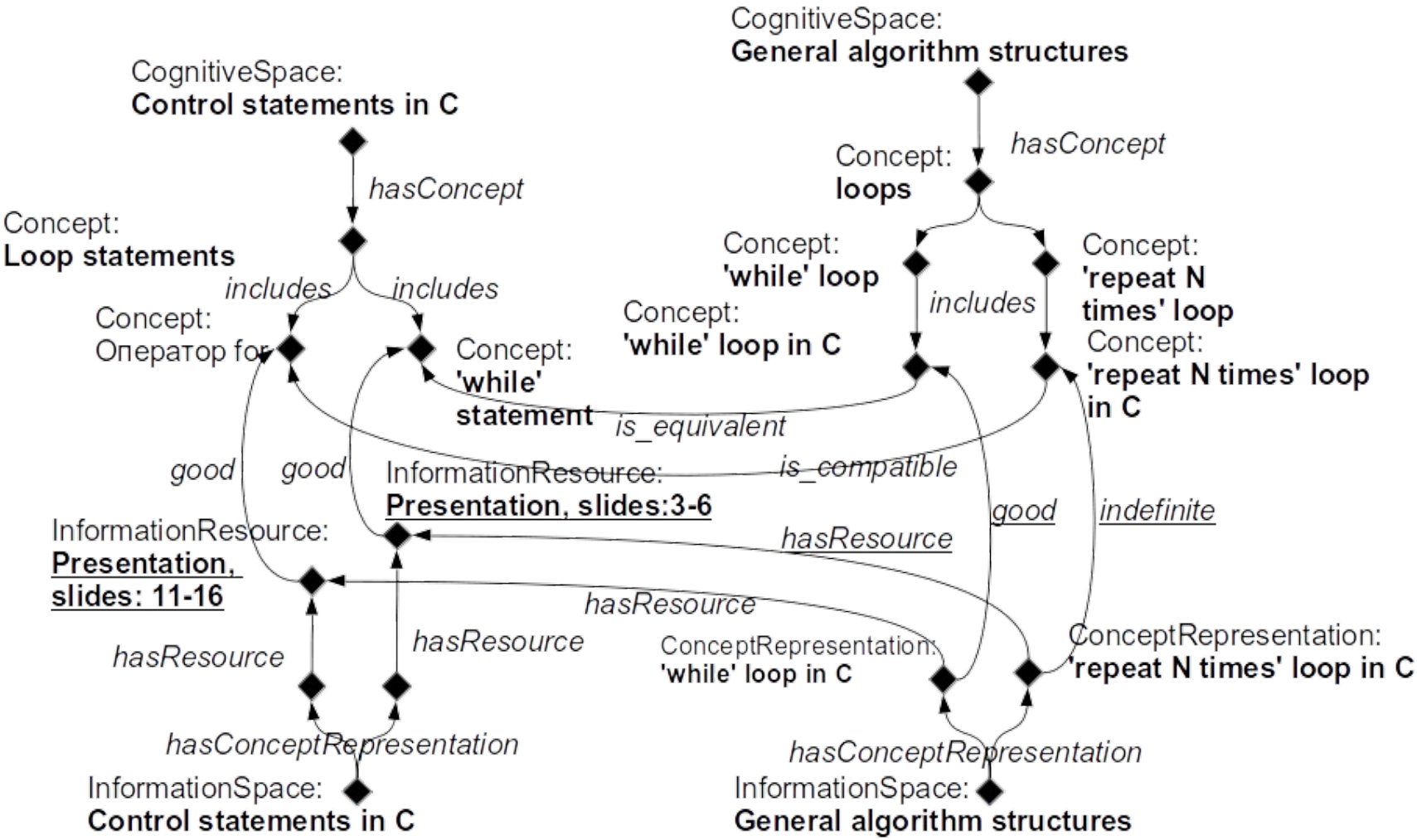
hasConceptRepresentation (?is1, ?cr1) ^ hasConceptRepresentation (?is2, ?cr2) ^ hasResource(?cr1, ?res)^ good(?res, ?c1) ^ is_compatible(?c2, ?c1) ^ sameAs(?cr1, ?c1) ^ sameAs(?cr2, ?c2) -> hasResource(?cr2, ?res) ^ indefinite(?res, ?c2), (6)

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

The example of association of two cognitive spaces



The example of association of two information spaces

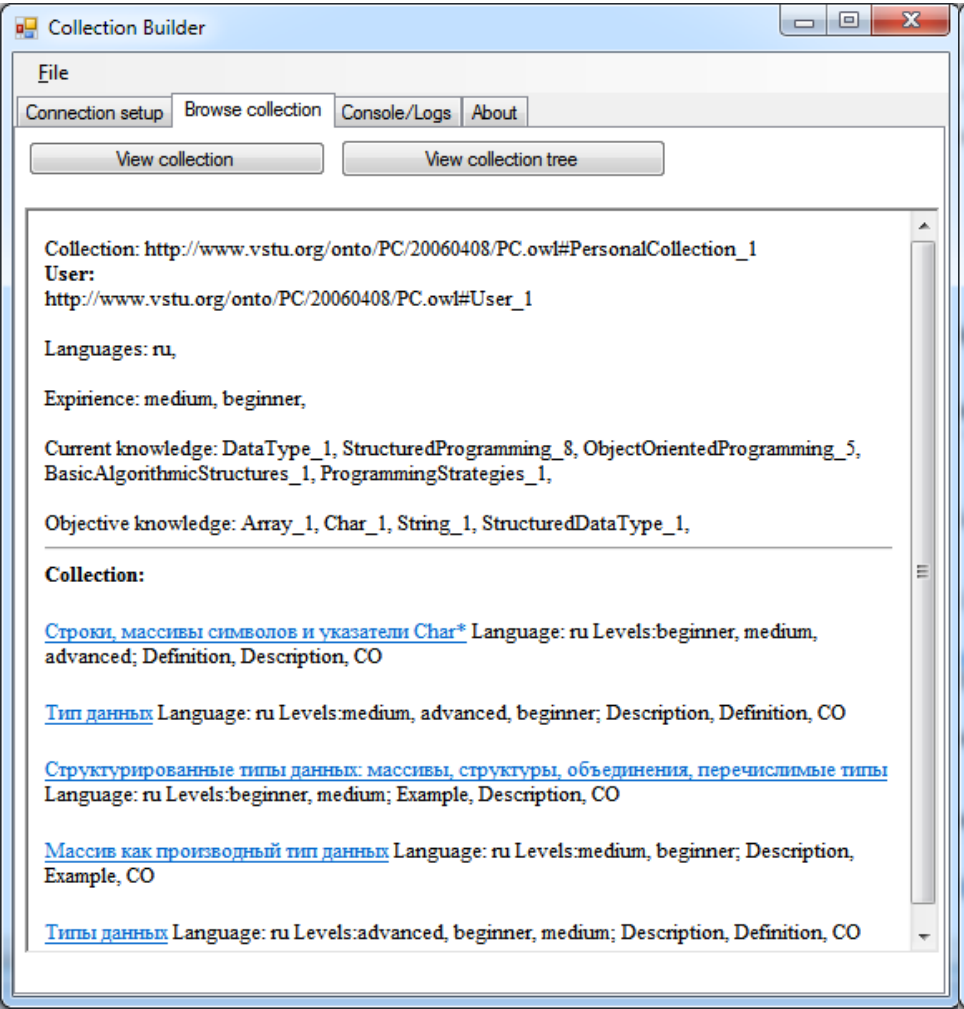


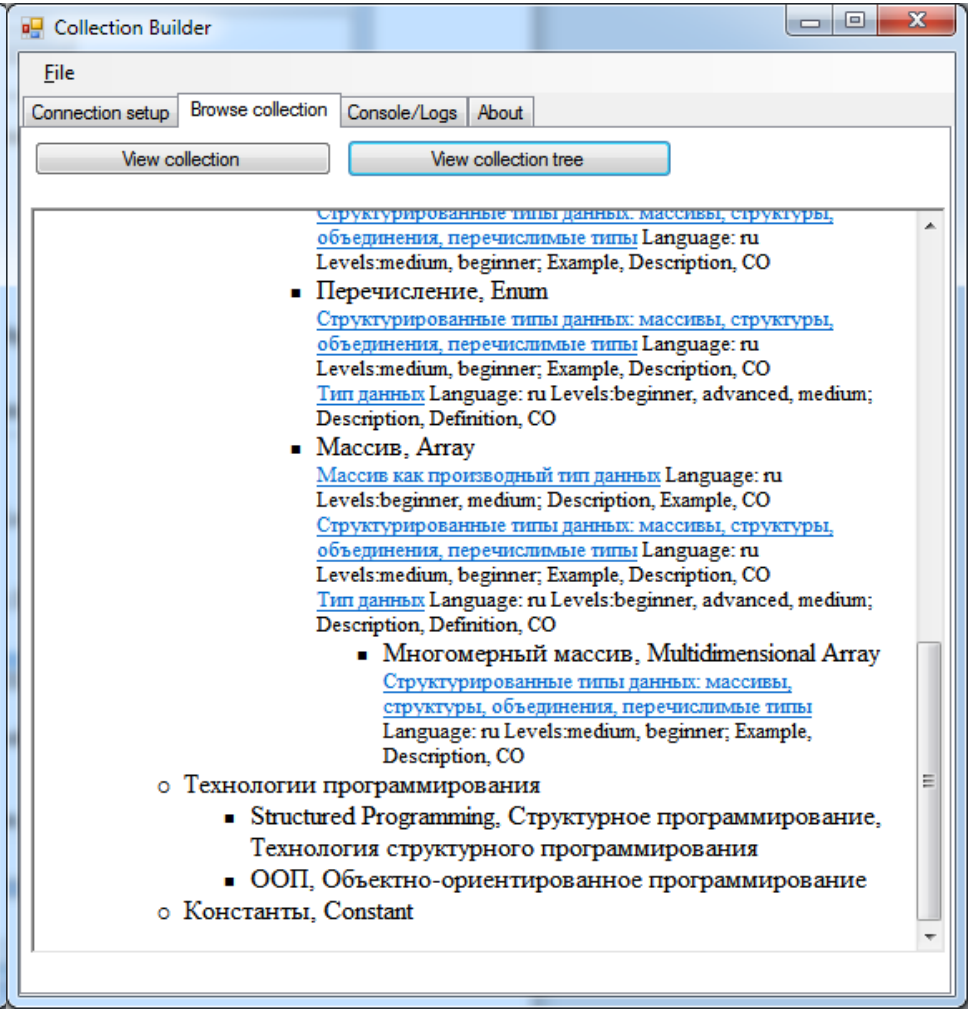
Conclusion

- 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. It 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.

- The main provisions of the approach used for development of the software tool for distributed learning resources retrieval and creation of the personal learning collections. Developed models, method and software tool were applied for creation of information space in the form of personal learning collection for the course «Programming languages. C++» in Volgograd State Technical University.

Personal Collection Builder





Further Work

- 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 for creation of these spaces (using the graph database StarDog 4.0)

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

¿Questions?