

IV. Canonical Form of the Subject Domain Ontology

We will say that the subject domain ontology is in canonical form if the following conditions are met [1]:

- all attributes from the set A participating in the definition of classes, functions, and axioms ontology are atomic;
- all attributes of each class have a functionality flag and have no subordinate attributes;
- the system of FD structure generators between the attributes of the ontology is the elementary basis of this FD structure.

The algorithm for reducing the subject domain ontology to the canonical form consists of the following steps.

1. For each composite attribute, add to set A the atomic attributes that make up the composite attribute. If this composite attribute is part of some class with a flag of functionality, then replace it in this class with atomic attributes with a flag of functionality. If a composite attribute is a part of a class without a functionality flag (it is a list attribute), then represent it as a subclass consisting of atomic attributes with a functionality flag included in the composite attribute. At the same time, determine the keys of a new class depending on the presence of an FD between atomic attributes within a composite attribute.
2. Each composite attribute included in the functions and axioms of the ontology should be replaced with atomic attributes included in its composition.
3. Each atomic attribute that is part of some class without a flag of functionality should be represented in the form of a subclass consisting of this attribute with a flag of functionality, and the class key consists of this atomic attribute.
4. Each atomic attribute that is part of a class and has subordinate attributes in it should be represented as a subclass consisting of this attribute and subordinate attributes with a flag of functionality. If this attribute was without the functionality flag, then the key of the new class consists of this atomic attribute, otherwise all of its attributes are included in the key of the new class.
5. Select non-trivial FD between attributes according to the rules described above and form a system of generators of FD structure on the set of attributes $P = \{P_j = X_j \rightarrow Y_j | X_j \subset A, Y_j \subseteq A, j = 1, m\}$.
6. Remove redundant attributes from the left sides of the FD from set P . Attribute $B \in X_j$ is considered redundant in X_j , if $B \in (X_j \setminus B)^+(P)$.
7. Remove redundant attributes from the right sides of the FD from set P . Attribute $B \in Y_j$ is considered redundant in Y_j , if $B \in X_j^+(P')$, where P' denotes the system of generators of the FD structure, obtained from P by replacing FD

$X_j \rightarrow Y_j$ with $X_j \rightarrow (Y_j \setminus B)$. As a result of performing steps 6 and 7, an elementary basis of the FD structure on a set of attributes $E = \{H_j \rightarrow T_j | H_j \subset A, T_j \subseteq A, j = 1, m\}$ will be obtained.

8. Bring the subject domain ontology in accordance with the obtained elementary basis of the FD structure between attributes by performing the following steps:
 - remove from the classes the attributes that turned out to be redundant in the right parts of the corresponding FD of the elementary basis;
 - remove from the composition of the keys of the classes the attributes that turned out to be redundant in the left parts of the corresponding FD of the elementary basis;
 - unite into one class those ontology classes that have the same closures of their keys concerning the elementary basis of the FD structure;
 - remove functions from the set F , in which all the attributes of the right-hand sides in the corresponding FD of the elementary basis turned out to be redundant;
 - from the left-hand sides of the functions from the set F , remove the attributes that turned out to be redundant in the left-hand sides of the corresponding FD of the elementary basis.

V. Optimal Canonical Form of the Subject Domain Ontology

The canonical form of a subject domain ontology is called optimal if it contains a minimum number of classes with a minimum number of occurrences of attributes in them. The task of bringing the subject domain ontology to optimal canonical form comes down to finding the optimal elementary basis of the structure FD between the attributes of the ontology, which contains the minimum number of FD with the minimum number of occurrences of attributes in them.

In article [9], the concept of P-dependencies in the elementary basis of the FD structure was introduced and studied. Let $E = \{H_j \rightarrow T_j | H_j \subset A, T_j \subseteq A, j = 1, m\}$ be the elementary basis of the FD structure $S(E)$ on the set A . We will say that in the FD $(H_s \rightarrow T_s) \in E$ there is a P-dependence of a nonempty $T'_s \subseteq T_s$ on H_s if there exists a $P \subset H_s^+(E)$, $P \neq H_s$ such that $T'_s \subseteq (P^+(E) \setminus P)$ and no subset of P possesses these properties. Let E' denote the system of generators of the FD structure obtained from E by replacing the FD $H_s \rightarrow T_s$ with $H_s \rightarrow T_s \setminus T'_s$. We will distinguish three types of P-dependence: P_1 -dependence occurs if $P \subseteq H_s^+(E')$, P_2 -dependence if simultaneously $P \not\subseteq H_s^+(E')$ and $T'_s \not\subseteq (P^+(E') \setminus P)$, P_3 -dependence if simultaneously $P \not\subseteq H_s^+(E')$ and $T'_s \subseteq (P^+(E') \setminus P)$.

We formulate the main properties of P-dependencies in the form of the following statements, the proof of which is carried out by checking the fulfillment

of sufficient conditions for the equivalence of the FD structures.

If in the FD $(H_s \rightarrow T_s) \in E$ there is a P_1 -dependence of $T'_s \subseteq T_s$ on H_s , and the system of generators of the structure of the FD Q is obtained from E by replacing the FD $H_s \rightarrow T_s$ with $H_s \rightarrow T_s \setminus T'_s$ and adding the FD $P \rightarrow T'_s$ to the E , then the structures of the FD specified by the systems of generators E and Q are equivalent.

If in the FD $(H_s \rightarrow T_s) \in E$ there is a P_2 -dependence of $T'_s \subseteq T_s$ on H_s , and the system of generators of the FD structure Q is obtained from E by replacing the FD $H_s \rightarrow T_s$ with $H_s \rightarrow ((T_s \setminus T'_s) \cup (P \setminus H_s))$ and adding the FD $P \rightarrow T'_s$ to the E , then the FD structures specified by the systems of generators E and Q are equivalent.

If in the FD $(H_s \rightarrow T_s) \in E$ there is a P_3 -dependence of $T'_s \subseteq T_s$ on H_s , and the system of generators of the FD structure Q is obtained from E by replacing the FD $H_s \rightarrow T_s$ with $H_s \rightarrow ((T_s \setminus T'_s) \cup (P \setminus H_s))$, then the FD structures specified by the systems of generators E and Q are equivalent.

The given properties of P -dependencies make it possible to move from one elementary basis of the FD structure to other elementary bases and find the optimal elementary basis of the FD structure.

In article [10], the concept of a cycle in the elementary basis of the FD structure is introduced and it is proved that the presence of cycles in the elementary basis of the FD structure is a necessary condition for the existence of P -dependencies in the elementary basis. The elementary basis of the FD structure on set A can be associated with a bipartite oriented graph (A, E, H, T) , in which A is the set of vertices of the first part of the graph, E is the set of vertices of the second part of the graph, H is the set of arcs of the graph directed from the vertices of the first part to the vertices of the second part of the graph (showing the occurrence of elements from A to the left parts of the FD), and T is a set of arcs of the graph directed from the vertices of the second part to the vertices of the first part of the graph (showing the occurrence of elements from A in the right parts of the FD). It is easy to verify that each cycle in an elementary basis corresponds to a family of cycles in the corresponding bipartite graph.

In general, the problem of finding all cycles in a bipartite directed graph is NP-hard, but its difficulty is determined by the fact that the maximum possible number of cycles in a graph depends exponentially on the dimension of the graph. The search time for one cycle in a directed graph using a standard depth-first search algorithm depends linearly on the dimension of the graph. In real optimization problems of the canonical form of a subject domain ontology, with the number of attributes on the order of 10^3 , the number of cycles in the elementary basis of the FD structure does not exceed 10^2 , therefore all cycles in the elementary basis of the FD structure can be found in an acceptable time.

VII. Software for Bringing the Subject Domain Ontology to Canonical Form

The input data of the software is the subject domain ontology in the OWL-2 language in the input file. The output data is the equivalent subject domain ontology, which is in the canonical form, presented as an owl-file. The software extracts from the original owl-file and presents attributes, classes, class hierarchy, links between attributes and classes in the form of database tables.

Then, atomic attributes and the system of forming the FD structure between the attributes are extracted from the database tables. The software for bringing the subject domain ontology to the canonical form includes the Attribute, AttributeSet, FuncDepen, and FDStructure classes developed in C++.

The Attribute class is used to represent a single atomic attribute. The class data is an attribute identification number, an attribute name, and an attribute purpose.

The AttributeSet class is used to represent any subset of atomic attributes. The class data is an array of attribute identification numbers. The class methods return the number of attributes in a subset, add an attribute to a subset, remove an attribute from a subset, check for the presence of an attribute in a subset, check if a given subset of attributes is in a subset, get the union and intersection of a given subset of attributes with a subset.

The FuncDepen class is used to represent a single functional dependency between atomic attributes. The data of the class are an object of the AttributeSet class, corresponding to the left part of the FD, and an object of the AttributeSet class, corresponding to the right part of the FD. Class methods add an attribute to the left or right part of the FD, remove an attribute from the left or right part of the FD.

The FDStructure class is used to represent a system of generators and an elementary basis for a structure of functional dependencies between atomic attributes. The class data is an array of objects of the FuncDepen class. The class methods return the number of FDs in the structure, add FDs to the structure, remove FDs from the structure, obtain the closure of a given subset of attributes with respect to the FD structure, and find the elementary basis of the FD structure.

Software for bringing the subject domain ontology to the optimal canonical form is under development. The software for bringing the subject domain ontology to canonical form was used in the development of the domain ontology of radio communication networks. As a result, the main ontology classes and their subclasses were obtained.

The TransmitterTypes class and its subclasses are designed to store data about transmitter types.

VII. Conclusion

To eliminate ambiguity and redundancy in the domain ontology, the concept of the canonical form of the domain ontology was introduced and algorithms

were proposed for bringing the domain ontology to a canonical form and an optimal canonical form. Software has been developed that implements bringing the domain ontology to a canonical form.

Список литературы

- [1] A. A. Karpuk and A. V. Havorka, "Bringing the Subject Domain Ontology of Radio Communication Networks to Canonical Form," *Problems of Infocommunications*, № 2(14), pp. 25–30, 2021, (in Russian).
- [2] A. V. Havorka and A. A. Karpuk, "Reduction the Subject Domain Ontology to Canonical Form," *International Journal of Information and Communication Technologies. Special Issue*, pp. 43–47, May 2022.
- [3] A. V. Palagin, S. L. Kryvy and N. G. Petrenko, "Ontological Methods and Means of Processing Subject Knowledge: Monograph," Lugansk, 324 p., 2012, (in Russian).
- [4] A. A. Karpuk and A. V. Havorka, "Construction of an Applied Ontology of Radio Communication Networks," *Vestnik svyazi*, № 6, pp. 36–40, 2021, (in Russian).
- [5] "W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes. W3C Recommendation", 5 April 2012 [Electronic resource] URL: <http://www.w3.org/TR/xmlschema11-2/>.
- [6] W. W. Armstrong, "Dependency structure of data base relationships," *Proc. IFIP Congress. Geneva, Switzerland*, pp. 580–583, 1974.
- [7] S. D. Kuznecov, "Database. Models and Languages", M., Binom-Press, 720 p., 2008, (in Russian).
- [8] A. A. Karpuk and V. V. Krasnoprishn, "Methodology of Data Domain Description for Databases Design in Complex Systems", *International Academy Journal Web of Scholar*, Vol. 1, № 4(13), pp. 11–20, 2017.
- [9] A. A. Karpuk, "Analysis of Structure of Functional Dependencies between Attributes of a Relational Database", *Economics and Management of Control Systems*, № 3(25), pp. 64–70, 2017, (in Russian).
- [10] A. A. Karpuk and V. V. Krasnoprishn, "Cycles in Structures of Functional Dependencies", *International Journal of Information Technologies and Knowledge*, Vol. 5, № 7, pp. 38–44, 2017.

Дано формальное определение онтологии предметной области. Рассмотрено понятие канонической формы онтологии предметной области, которая строится на основе анализа функциональных зависимостей между понятиями и свойствами понятий онтологии. Описан алгоритм приведения онтологии предметной области к канонической форме. Введено понятие оптимальной канонической формы онтологии предметной области, содержащей минимальное количество классов и минимальное количество атрибутов в классах. Предложен метод приведения онтологии предметной области к оптимальной канонической форме.

Received 13.03.2024

**ПРИВЕДЕНИЕ ОНТОЛОГИИ
ПРЕДМЕТНОЙ ОБЛАСТИ К
ОПТИМАЛЬНОЙ КАНОНИЧЕСКОЙ
ФОРМЕ**
Карпук А.