Ensuring Semantic Interoperability Based on the Merging of Ontological Models[1]

Dmitry Korneev 1[0000-0001-7260-4768], Alexander Boichenko 1[0000-0003-3113-9446] and Vasily Kazakov 1[0000-0001-8939-2087]

[1] Plekhanov Russian University of Economics, Moscow, Russia
Korneev.DG@rea.ru

Abstract. The article describes an ontologies merging algorithm used to ensure the semantic interoperability of information systems (IS). The algorithm is based on a set-theoretic approach for calculating measures of semantic proximity of vertices of homogeneous ontologies at the level of subject areas and the level of tasks. The measure of semantic proximity is calculated taking into account the comparison of the attributes of the compared concepts of ontologies and the values of these attributes, the location of the selected nodes within the corresponding ontologies, and also taking into account the comparison of the presence and types of links of the evaluated concepts.

Keywords: semantic interoperability, ontological engineering, an algorithm for integrating ontologies

1 Introduction

Interoperability in ISO / IEC 24765-Systems and Software Engineering-Vocabulary [1] refers to “the ability of two or more systems or elements to exchange information and to use information obtained as a result of the exchange”.

Interoperability standards and studies address different levels of interoperability between systems. Most often in scientific research they refer to the European Interoperability Framework v2.0 (EIF stack) [2], in which the following logical levels of interaction are distinguished:

1. Regulatory - involves the interaction of systems in a single regulatory and legislative environment;
2. Organizational - refers to the organizational aspects of the functioning of information systems and presupposes the commonality of business processes and regulations for their functioning;
3. Semantic - the ability of systems to understand the meaning of the information that they exchange;
4. Syntactic - the ability to exchange data, the ability of systems to integrate;
5. Technical - the organization of the relationship between systems.

[1] The article was prepared with the support of the Russian Foundation for Basic Research (grants No. 18-07-01053 and No. 20-07-00926).

Proceedings of the of the XXIII International Conference "Enterprise Engineering and Knowledge Management" (EEKM 2020), Moscow, Russia, December 8-9, 2020.

At the first two levels of the EIF stack, initial requirements for the design of information systems are set, organizational measures are taken to unify the relevant regulatory documents and business processes.

To ensure the fourth and fifth levels of EIF stack in the design and development of information systems, they must include certain software tools. The indicated levels of interoperability are well enough studied and their practical implementation does not cause serious difficulties at present.

Currently, the greatest scientific and practical significance is the solution of problems of ensuring the semantic interoperability of information systems (IS). This is also due to the fact that in recent years the intelligence of IS, including devices operating on IoT technology, has sharply increased. Information systems are being created that are capable of replacing a person in many respects, including in the field of making intelligent decisions. Understanding the meaning of the request (and not just the syntax of the request) that comes to the IS from another system will allow you to give a more correct answer, which, in turn, should be as correctly understood as possible by the system that generated the request. Ensuring semantic interoperability is associated with the need to apply ontologies of concepts used in processes and describing the processes of functioning of an information system.

Based on the studies carried out [3, 4, 5], the authors formalized the requirements for the structure of the ontology to ensure semantic interoperability: the basic concepts that allow describing both the static state and dynamic changes in the states of objects in the subject area, sets of attributes (properties) of concepts and the main types of links between concepts, sets of attributes (properties) links. In particular, it is proposed to use the following types of concepts: "Object class", "Object" and "Entity". Concepts can be linked together by the following types of unidirectional or bidirectional relationships: "Inheritance", "Association", and "Action". In [4], the language OWL-DL was chosen as the optimal means for describing ontologies, and the ORACLE-11g DBMS was chosen as the storage medium for ontologies.

Based on the results of the studies carried out [4, 5], the following ontology construction algorithm was proposed to ensure the semantic interoperability of SIS:

1. Allocation of ontology concepts and definition of the semantics of links in accordance with the rules [4].
2. Description of the ontology by means of the OWL DL language using the Protégé 5.0 ontology editor (creating an OWL file).
3. Creation of structures for storing ontologies in the ORACLE 11g DBMS.
4. Filling out the structures in accordance with the description of item 2 (loading the OWL file into the ORACLE 11g DBMS).
5. Creation of additional user rules for obtaining implicit knowledge in the ORACLE 11g DBMS environment.

2 Ontology merging algorithm used to ensure the semantic interoperability of information systems

To ensure the semantic interoperability of information systems, it is necessary to compare the ontologies that underlie them and find out their commonality and differences. This problem is solved by using methods for assessing the semantic proximity of ontology concepts. Many well-known methods for finding a measure of proximity between ontology concepts are based on Tversky's set-theoretic approach, based on comparing the properties of concepts [6]. In works [7-12] the mutual arrangement of vertices within the ontology is analyzed. The lengths of paths between pairs of concepts are calculated. The length of the shortest path is determined as the number of concepts in the ontology located between the two nodes under consideration, which are interconnected. It is believed that the shorter the path length between the vertices, the semantically closer the pair of concepts of the considered ontology [7]. In [13], the frequency of occurrence of a concept and its subclasses in one and another ontology is taken as the basis for calculating the measure of semantic proximity of two concepts of different ontologies. The methods described above for calculating proximity measures between ontology nodes are symmetric. The work [14] describes a calculation method, the essence of which is that the closeness of two concepts depends on the closeness of concepts with which there are hierarchical relationships, and is calculated recursively.

The most promising for use in algorithms used to calculate measures of semantic proximity of ontology concepts are the so-called hybrid measures. The hybrid measure proposed in [15] consists of three parts - taxonomic, relational, and attributive. Difficulties in comparing different ontologies of subject areas lie in the difference in the names of concepts and relations, as well as in the approaches to the definition of concepts. When mapping two ontologies, a search is performed for each concept of one ontology of a similar concept of another ontology, taking into account the synonymy of concepts. In works [16, 17], a method for calculating a measure is proposed, taking into account the lexical proximity of concepts, properties, domains and ranges of relations (ranges of values of the arguments of relations), parent/child concepts. The main disadvantage of most methods for determining semantic proximity is the need to involve an expert to confirm the correctness of detecting similarities and differences in semantic concepts.

Below we will consider the problems of integrating ontologies that reflect either different points of view on the same subject area, or different points of view on the same problem (i.e., we will integrate homogeneous ontologies at the level of subject areas and levels of tasks). The purpose of the integration is to preserve the existing and define new semantic dependencies of the concepts contained in both ontologies.

In accordance with the results of works [4, 5], the following formal definitions can be given regarding ontologies used to ensure the semantic interoperability of IS.

1) A lot of concepts are defined as follows:

C = {C1, C2, C3}, (1)

where: C1– concept of the “Object class” type;

C2 – concept of the "Object" type;

C3 – concept of the "Entity" type.

2) The set of relationships between concepts is defined as follows:

$R = \{R1, R2, R3\}$, (2)

where: $R1$– relationship "Inheritance" (relationship "class-subclass");

$R2$ – relation "Association";

$R3$ – is the "Action" relation.

3) The ontology used to ensure the semantic interoperability of the IS can be formally presented in the following form:

$O = \{C_i (A_{ij}, S_{ik}), R_{ij}, P_m\}$, (3)

where: C_i - ontology concepts;

A_{ij} – j-th attribute of the i-th concept;

S_{ik} – the k-th synonym for the i-th concept;

R_{ij} – is the relationship between concepts i and j;

P_m – inference rules.

It is proposed to build an algorithm for integrating ontologies to ensure the semantic interoperability of the IS based on the calculation of the semantic proximity of the vertices of two ontologies $O1$ and $O2$. For each concept C_{1i} of the ontology *O1*, we calculate the measures of semantic proximity with the concepts C_{2j} of the ontology $O2$.

In the algorithm described below, the calculation of measures of semantic proximity of ontology concepts used to ensure the semantic interoperability of IS will be based on the set-theoretic approach [6]. The main idea of this approach is that to calculate the measures of semantic similarity, it is necessary to take into account not only the general properties of objects, but also their differences. The proposed algorithm will calculate the measures of semantic proximity of homogeneous concepts, that is, concepts that have the same names or names that are synonyms. The proximity measure will consist of three parts:

• Attributive measure, which is calculated based on the comparison of the attributes of the compared concepts and the values of these attributes;

• Geometric measure, which is calculated taking into account the location of the selected vertices within the corresponding ontologies;

• Relational measure, which is calculated on the basis of comparing the presence and types of relationships of the evaluated concepts with other concepts of the corresponding ontologies.

Let us introduce the following characteristics of measures of semantic proximity:

• Equivalence. We will assume that the vertex C_{1i} of the *O1* ontology is equivalent to the vertex C_{2j} of the *O2* ontology if: 1) the composition of attributes and their values coincide or differ by intervals not exceeding the minimum threshold values (attributive measure); 2) the selected concepts are located in ontologies in such a way that the lengths of the minimum chains (bridges) between these concepts and two other equivalent concepts in each ontology does not exceed the minimum allowable threshold value (that is, in ontologies the selected concepts are "surrounded" by concepts the evaluations carried out are equivalent (geometric measure); 3) the evaluated concepts are associated with concepts with the same types of links, or the number of different types of links does not exceed a certain minimum threshold value (relational measure).

• Conformity. Determined according to the rules described above. In the case when the minimum allowable threshold value of the corresponding measure of proximity is exceeded, a comparison is made with the maximum allowable threshold value of the corresponding measure of proximity of concepts. In this case, the maximum threshold value must not be exceeded. Vertices possessing the above characteristic of the proximity measure will be called corresponding.

• Difference. Determined according to the rules described above. In the case when the maximum permissible threshold value of the corresponding proximity measure is exceeded. Vertices possessing the above-described characteristic of the measure of proximity will be called different.

To construct an ontology merging algorithm, it is also proposed to use the concept of a bridge - a chain of ontology vertices that correspond to equivalent concepts used to establish a mapping of two ontologies in [18].

To integrate ontologies used to ensure semantic interoperability, the following sequence of actions is proposed.

Step 1. In the ontologies *O1* and *O2*, bridges are computed, consisting of vertices that in pairs have equivalent or corresponding proximity measures. The lengths of the bridges (the number of vertices) must coincide.

Step 2. Calculate the weight of each bridge. Assigning to the vertices with an equivalent measure of proximity, the maximum coefficient is equal to 1, and to the vertices with the corresponding measure of proximity - a coefficient in the range from 1 to 0.5, depending on the approach to the threshold values of the estimated parameters of the attributive, geometric, and relational measures of proximity.

Step 3. As the base for merging, we choose the ontology in which there is the largest number of vertices. Let in our case it be *O1* ontology. We select in it all the bridges defined in the previous steps.

Step 4. For differing vertices in the O2 ontology, we find the bridges with the largest weight and the smallest length to the vertices included in the bridges defined in Step 2.

Step 5. Add the bridges found in Step 4 to *O1* ontology.

Steps 4 and 5 are repeated iteratively. We start looking for bridges with a length equal to 1, sequentially increasing the length of the bridge by one vertex at each iteration.

Moreover, if the new vertex C_{2i} is already included in the O1 ontology as a result of performing Steps 3 and 4 (it became a new vertex C_{1j}), then the vertices C_{2i} and C_{1j} are considered equivalent.

In this case, the algorithm, due to the formalization of the ontology structure (see formulas 1-3 above), makes it possible to avoid the semantic conflicts described in [19], which arise during the merging of ontologies at the time of transferring vertices connected by the types of links of the type ". In ontologies used to ensure semantic interoperability and constructed in accordance with the rules described in [4, 5], the vertices indicated in [4] will be linked by links of the "Association" type. This will allow avoiding semantic conflicts when using the ontology merging algorithm described above.

3 Summary

Currently, there are algorithms and their software implementations for the automatic merging of ontologies. Information about such algorithms is given, for example, in [20]. Each of the currently existing algorithms for automatic merging of ontologies has, along with advantages, a number of significant disadvantages. The disadvantages are primarily associated with attempts to create a universal algorithm for combining ontologies that describe concepts of one subject area, but have different structures and algorithms for their initial construction. The article presents an original algorithm for the integration of ontologies representing structured knowledge about each of the interacting ISs, taking into account the fact that their structures and construction algorithms are clearly defined and unified.

4 Acknowledgements

The authors of the article are grateful to the Russian Foundation for Basic Research for their support in writing the article (grants No. 18-07-01053 and No. 20-07-00926).

References

1. ISO/IEC 24765-Systems and Software Engineering-Vocabulary, URL: http://www.cse.msu.edu/~cse435/Handouts/Standards/IEEE24765.pdf, last accessed 2019/03/30.
2. EIF - European Interoperability Framework, http://ec.europa.eu/idabc/en/document/2319/5938.html, last accessed 2019/03/30.
3. Korneev D.G., Gasparian M.S., Kiseleva I.A., Mikryukov A.A. Ontological engineering of educational programs. Revista Inclusiones. 2020; 7(S2-3).
4. Korneev D.G., Gasparian M.S., Mikryukov A.A., Yaroshenko E.V., Golkina G.E. The technology for semantic interoperability based on a cognitive approach / International Journal of Advanced Trends in Computer Science and Engineering. 2020; 9(3):3637-3640.
5. Korneev D., Boichenko A., Kazakov V. Warehouse development of ontology for providing semantic interoperability. CEUR Workshop Proceedings. Selected Papers of the 22nd International Conference "Enterprise Engineering and Knowledge Management", EEKM 2019. 2019; 2413:70-76. URL: http://ceur-ws.org/Vol-2413/paper09.pdf
6. Kuznetsov O.P., Sukhoverov V.S., Shipilina L.B. Ontology as a systematization of scientific knowledge: structure, semantics, tasks. URL: http://cmm.ipu.ru/proc (date of access: 02.09.2020).
7. Rada R., Mili H., Bicknell, E. et al. IEEE Trans. on Systems, Man and Cybernetics. 2018; 19:17.
8. Leacock C., Chodorow M. WordNet: An electronic lexical database. Cambrige, 2019. pp. 265.

9. Wu Z., Palmer M. Proc. 32nd Annual Meeting of the Association for Comput. Linguistics. Las Cruces, 2018. p. 133.

10. Li Y., Bandar Z. A., McLean D. IEEE Trans. on Knowledge and Data Engineering. 2017. p. 871.

11. Hirst G., St-Onge D. // WordNet: An electronic lexical database. Cambrige, 2018. p. 305.

12. Lukashevich N.V. et al. Ontologies for automatic text processing: a description of concepts and lexical meanings. In: Computational linguistics and intellectual technologies: proceedings of the international conference "Dialogue 2006". Publishing house of the Russian State University for the Humanities, Moscow; 2016. pp. 138-142.

13. Resnik P. Using information content to evaluate semantic similarity in a taxonomy. In: Proc. 14th Int. Joint Conf. on Artificial Intelligence. Montreal, 2015. p. 448.

14. Maedche A., Staab S. Proc. 13th EKAW Conf. LNAI. Berlin: Springer, 2018. p. 251.

15. Maedche A., Zacharias V. Proc. 6th European PKDD Conf. LNCS V. 2431. Berlin: Springer, 2017. p. 348.

16. Rodríguez M.A. Thesis for Degree of Doctor of Philosophy. University of Maine, 2018.

17. Karpenko A.P. et al. Methods for mapping ontologies. Review. Science and Education. 2009. URL: http://technomag.edu.ru/doc/115931.html (Date of treatment 10/02/2020).

18. Nguyen H.A. Thesis for the Degree Master of Science. University of Houston − Clear Lake, 2018.

19. Vostrov A, Kurochkin M. Conflict detection in the integration of expert systems on ontological models. Scientific and technical bulletin of SPbSPU. Informatics. Telecommunications. Management. 2018; (2).

20. Swati Negi, Sanjay Kumar Malik An Algorithm for Merging Two Ontologies: A Case Study/ International Journal of Applied Engineering Research. 2018; 13(12).