

The Method of Automated Building of Domain Ontology

Igor Antonov

Faculty of Computer Science and
Electrical Power Engineering
Pskov State University
Pskov, Russia
igorant63@yandex.ru

Iuliia Bruttan

Faculty of Computer Science and
Electrical Power Engineering
Pskov State University
Pskov, Russia
bruttan@mail.ru

Dmitry Andreev

Faculty of Computer Science and
Electrical Power Engineering
Pskov State University
Pskov, Russia
dandreev60@mail.ru

Lilia Motaylenko

Faculty of Computer Science and
Electrical Power Engineering
Pskov State University
Pskov, Russia
lvs@mail.ru

Abstract—This article is devoted to the tasks of automating the construction of domain ontologies. In the beginning, the limitations and problems of constructing the ontology of the domain using the well-known methods are discussed. Next, a model of the domain ontology is proposed, which provides the ability to automatically build the ontological hierarchy, including the automatic synthesis of generalized concepts. Then, the article discusses the method of building an ontology based on the proposed model using machine learning, and discusses its capabilities and limitations.

Keywords—domain ontology, information systems, information retrieval, data mining, machine learning.

I. INTRODUCTION

The concept of ontology came to information technology from philosophy. In the information disciplines ontology is currently understood as a formalized model of a system of concepts covering a certain subject area. Such model includes a formal representation of the content of concepts and relationships between them. Ontologies have practical application in various fields. Among other things, ontologies are the means of documenting a consistent point of view of specialists on the structure and terminology of the subject area. Ontologies are used as part of software tools that provide information support for the design of complex technical systems. Also, ontologies are used as teaching materials, clearly demonstrating the system of connections between the concepts of the domain, thus presenting it as a coherent systemic whole [1].

Information systems developed on the basis of ontologies have shown their effectiveness in practice and interest in them is constantly growing. However, the existing methods of ontology building require expert knowledge in the studied domain, and the construction of ontologies based on their use takes a significant amount of

time, therefore, the actual task is to automate the process of ontology building.

II. LIMITATIONS AND PROBLEMS OF THE METHODS FOR AUTOMATED ONTOLOGY BUILDING

There are several areas in which research is being conducted for creating means for building and maintenance of domain ontologies automating.

The authors of [2] - [11] propose methods for constructing ontologies based on automatic processing of the content of web resources and natural language texts. As a priority source of source data, the authors often point out Wikipedia as a source with a high density of meaningful information about the relationship between the concepts of the subject area. These papers consider the problem of statistical, syntactic and lexical text processing in order to extract data used in the construction of ontology. The task of a multi-level ontological hierarchy construction automating in the scope of these works is not being solved. The extraction of ontological information is performed by identifying the semantic relations of the terms of the subject domain based on the analysis of the syntactic structure of sentences. According to the results of statistical analysis, the most frequently used terms are singled out as candidates for inclusion in the ontology. The results of computer texts processing go through the stage of verification, adjustment and additions by the domain expert, after which they are used as an ontological model of the subject domain. The main difficulty of the practical use of the solutions considered in these works is that the ambiguity and contextual dependencies of natural language constructs limit the possibilities of their computer processing and do not provide completeness and correctness of extracting semantic information from texts using formal methods.

Print ISSN 1691-5402

Online ISSN 2256-070X

<http://dx.doi.org/10.17770/etr2019vol2.4065>

© 2019 Dmitry Andreev, Sergey Lyokhin, Olga Poletaeva, Victor Nikolaev.

Published by Rezekne Academy of Technologies.

This is an open access article under the Creative Commons Attribution 4.0 International License.

The works [12] - [18] analyze the possibility of using the tools developed within the framework of the Formal Concept Analysis (FCA) method for the automatic construction of the domain ontology. As initial data for the ontological analysis of the subject area, these works suggest using a two-dimensional "object - attributes" table, which is based on empirical data about the objects of the subject area or with results of performing certain procedures that reveal the attributes of the objects of the subject area, including data extraction operations from text arrays. Such "objects - attributes" table is transformed into the formal context of the FCA method by combining objects with homogeneous composition into classes. Based on the analysis of the formal context matrix, algorithms that implement the FCA method generate all possible formal concepts. Formal concepts are connected by hierarchical relations. Each concept in the hierarchy represents the objects sharing some set of properties; and each sub-concept in the hierarchy represents a subset of the objects (as well as a superset of the properties) in the concepts above it. The authors of the noted papers do not consider the problem of assessing the semantic significance of concepts included as a result of applying the FCA method to the domain ontology. When constructing ontologies intended for practical use, one should take into account that the hierarchy of concepts synthesized on the basis of the formal method will include formal concepts that do not have semantic significance and correspond to any repeated combinations of lower-level concept attributes. Therefore, the methods considered in these works do not allow synthesizing ontology directly suitable for use as domain ontology in real-world applications.

In general, we can conclude that at present there is a certain lack of technology and tools that provide automated construction and maintenance of domain ontologies. Further progress in this area requires the development of new methods for the automated construction of domain ontology and software tools that provide opportunities for automating the construction and maintenance of ontologies.

III. DOMAIN ONTOLOGY MODEL

As a possible solution to a part of the problems discussed above, this paper proposes a domain ontology model that provides the ability to effectively automate the ontology construction process, and proposes a method for automated ontology construction. The ontology model is represented by a tuple (1).

$$Q = \langle C, M, R \rangle \quad (1)$$

where $C = \{c_i\}$ - the set of concepts forming ontology, $1, I$;

$M = \{M_i\}$ - the set of sets of concept attributes;

$M_i = \{m_1, \dots, m_{d_i}\}$ - the set of attributes describing the i -th concept,

d_i - number of attributes describing the concept ;

$R \subseteq C \times C$ - the relationship of the direct inheritance of concepts.

The relationship R can be defined by a matrix with dimension $I \times I$. If the concept c_k directly inherits from concept c_p i.e. $(c_p, c_k) \in R$, then the matrix element $r_{ik} = 1$, otherwise, if $(c_p, c_k) \notin R$, then $r_{ik} = 0$. The relationship of direct inheritance between two concepts means that there are no intermediate concepts between them:

$$\exists c_i \exists c_k ((c_i, c_k) \in R) \Rightarrow \exists c_h (((c_i, c_h) \in R) \wedge ((c_h, c_k) \in R)) \quad (2)$$

If two ontology concepts are in inheritance relation, then the corresponding parent and child concepts in the proposed model are related by a strict partial order relation, denoted by '<'. This relationship is transitive:

$$\forall c_i \forall c_j \forall c_k ((c_i < c_j) \wedge (c_j < c_k) \Rightarrow c_i < c_k) \quad (3)$$

The proposed model differs from the traditional ontology model by the presence of a unique correspondence of a certain set of attributes to a certain ontology concept. The concepts related by inheritance in the proposed model always differ in the composition of their attributes. This solution allows to formalize the operation of incorporating a new concept into the ontological hierarchy. At the same time, mentioned above properties of the model do not create restrictions for the use of ontology based on this model within information systems.

IV. METHOD OF BUILDING OF ONTOLOGY

The method of constructing an ontology based on the model discussed above allows to automatically determine the position of the inclusion of a new concept in the ontological hierarchy and automatically generate new generalized concepts caused by the addition of a new concept to the ontology.

The proposed method provides the implementation of an iterative ontology construction by successively entering into the ontology model new concept models c_x represented by a set of attributes M_x , where $M_x \subset M$. For newly introduced concepts, the place of the concept in the ontological hierarchy is determined automatically.

If the set of attributes of a new concept is not a superset of the set of attributes of any of the ontology concepts, then the concept will occupy a position at the level immediately following the root ontology concept c_0 (Fig.1):

$$\exists c_j ((c_j \in C) \wedge (M_j \subset M_x)) \quad (4)$$

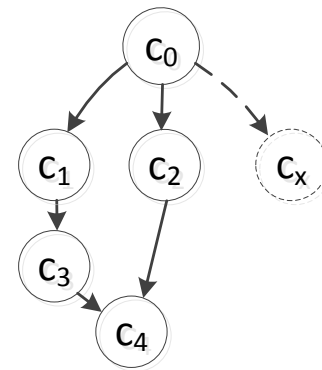


Fig. 1. Adding a new top-level concept.

In other cases, the new concept will have to inherit from any other concepts present in the ontology and inheriting from the concept c_0 . To automatically determine the parent concepts, a sequential “top-down” traversal of the graph representing the ontology concept links is performed. In the process of traversing the graph, the parent concept for the concept c_x is determined based on the following criterion: “The concept c_x will have the parent c_p for which a set of attributes M_p is a subset of the set of attributes of the concept c_x , and the set of attributes of any of the child concepts of a concept c_p is not a subset of the set of attributes of the concept c_x ”:

$$((M_p \subset M_x) \wedge \exists c_d ((r_{pd} = 1) \wedge (M_d \subset M_x)) \Rightarrow r_{px} := 1 \quad (5)$$

If (5) is violated and $\exists c_d ((r_{pd} = 1) \wedge (M_p \subset M_x) \wedge (M_d \subset M_x))$ then c_d is considered as the next candidate for parent concepts for the concept c_x and replaces c_p in the condition check operation (5). In this case, the search for the parent concept for the concept c_x in the general case is carried out along several branches of the graph. If condition (5) is fulfilled when traversing a branch of a graph, the search for this branch is complete at the concept c_p level, and the concepts indicated in (5) as c_p form a set of parent concepts of the concept c_x . If the set of attributes of the concept c_d in the case of truth of condition (5) is a superset of the set of attributes of the concept c_x , then c_d is a child of the concept c_x and is included in the corresponding set:

$$\forall c_d ((r_{pd} = 1) \wedge (M_p \subset M_x) \wedge (M_x \subset M_d)) \Rightarrow (r_{px} := 1, r_{xd} := 1, r_{pd} := 0) \quad (6)$$

On the graph the concept c_x is placed in the «gap» between the concepts c_p and c_d (Fig. 2).

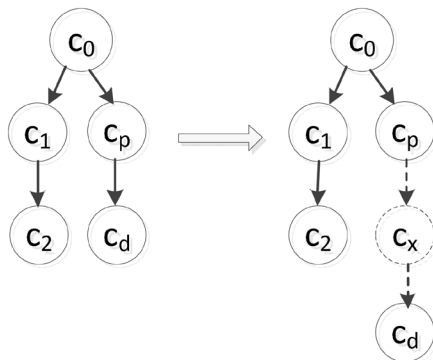


Fig. 2. Adding a new concept with breaking the inheritance link.

If the condition specified for c_d is not fulfilled, then c_x becomes a neighbour concept for the c_d , being located at the same level of the hierarchy of concepts and also inheriting from c_p (Fig. 3).

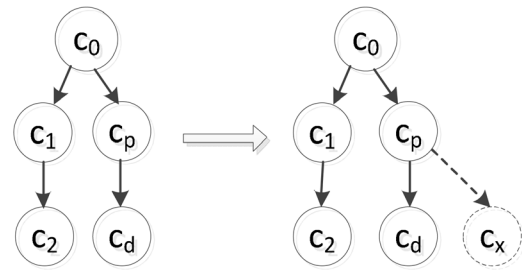


Fig.3. Adding a new child concept without breaking the inheritance relationship.

After determining the place of a new concept in the ontological hierarchy, new generalized concepts of higher levels can be automatically determined on the basis of identifying sets of attributes that recur in various concepts. The discovery of such sets of attributes that are not indicated in the ontology by any concepts means the appearance of new generalizations that can be included in the ontology as new concepts. Fig. 4 shows the inclusion of the concept c_x in the ontology with the formation of new generalizations c_g and c_g' .

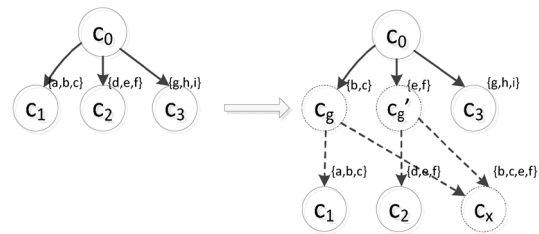


Fig.4. Inclusion of the concept in ontology with the formation of new generalizations.

Automatically detected generalizations before being included in the ontology should be evaluated by an expert for their semantic significance. If the expert considers the generic concept to be of no value, it is not included in the ontology and automatically considered later as ignored when performing the following inclusions of new concepts in the ontology. Iterative construction of a list of generalizations excluded from ontology implements machine learning for automatic recognizing concepts that are not of semantic significance.

When using the proposed method, the domain ontology is formed in the process of successively adding to the ontology models of concepts represented by a set of attributes characterizing the concept. Based on the formal criteria, an analysis of the composition of the new concept is performed and the place of the new concept in the ontological hierarchy is objectively determined. When a new concept is included in ontology, its links with parent and child concepts are automatically determined and new generalized concepts are formed, which, after evaluating their semantic significance, are either included in the ontology or supplement the list of non-valuable generalizations for the ontology. The method reveals the position of the concept in the ontology, regardless of the level that the concept should occupy in the general hierarchy. The position of a new concept can be located at the top level, at the bottom level, or at any of the

intermediate levels of the ontology. The method can be used to form the ontology from scratch, starting with an empty ontology, and to expand an existing domain ontology.

V. CONCLUSION

On the basis of the model and method proposed in the work, ontology editors can be created that provide automated support for ontological engineering operations. The proposed method provides ontology machine learning for automatic recognition of semantically insignificant generalized concepts, which makes it possible to exclude multiple manual processing of concepts that do not have semantic significance in the process of ontology construction. The method complements the traditional construction of ontology “from top to down” with the possibility of entering into the ontology of new concepts “from below” based on the input into the system the information about the attributive composition of the new concept. The source of information about the structure of the concepts of the lower levels can be empirical data about the objects of the subject area, in particular, the experimental data obtained as a result of measurements and studies of objects of the subject area. The method is limited in application in relation to top-level ontologies and humanitarian ontologies, where the non-formalized interpretation of concepts is mainly used.

REFERENCES

- I. V. Antonov and Iu. V. Bruttan, “Ontological approach to building control systems of electronic learning objects,” *Bulletin of Pskov State University, Economics. Law. Management*, No. 5, pp. 230-234, 2014. (in Russian)
- E. A. Rabchevsky and G. I. Bulatova, “Automatic ontology construction,” *Scientific and technical bulletins of SPbSPU*, No. 4, pp. 22-26, 2007. (in Russian)
- [1] A. E. Ermakov, “Automation of ontological engineering in systems of knowledge extraction from text,” *Computer linguistics and intellectual technologies: in proceedings of The International Conference Dialogue, Bekasovo*, June 4-8, 2004, Moscow: RSUH, 2008, pp.154- 159. (in Russian)
- [2] Song, Q., Liu,J., Wang, X. A Novel Automatic Ontology Construction Method Based on Web Data: *Proceedings of Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kitakyushu, Japan*, 2014, pp.762-765.
- [3] Khalatbari, S., Mirroshandel, S. A.: Automatic construction of domain ontology using wikipedia and enhancing it by google search engine. *J. Inf. Syst. Telecommun.* 3, 2015, pp. 248-258.
- [4] D. Küçük, Y. Arslan, Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles: *Renewable Energy*, 62, 2014, pp. 484-489.
- [5] Damir Jurić, Marko Banek, Zoran Skočir, Automated Construction of Domain Ontology Taxonomies from Wikipedia: *International Conference on Database and Expert Systems Applications DEXA, Database and Expert Systems Applications*, 2011, pp. 439-446.
- [6] P. Buitelaar, P. Cimiano, Ontology learning and population: Bridging the gap between text and knowledge: *Frontiers in Artificial Intelligence and Applications Series*, Vol. 167, IOS Press, 2008.
- [7] P. Cimiano, J. Völker, *Text2Onto. NLDB*, 2005, pp. 227–238.
- [8] Dahab, M.Y., Hassan, H.A., Rafea, A., *TextOntoEx: Automatic Ontology Construction from Natural English Text: Expert Systems with Applications*. Volume 34, Issue 2, 2008, pp. 1474–1480.
- [9] Azevedo, R. R., Freitas, F., Rocha, R. G. C., Menezes, J. A. A., Oliveira Rodrigues, C. M., and F. P. Silva, G., An Approach for Learning and Construction of Expressive Ontology from Text in Natural Language: *Proceedings of 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, 2014, pp. 149-156.
- [10] S. V. Smirnov, “Ontological analysis of domain modeling,” *Proceedings of Samara scientific center of RAS*, Vol. 3, No. 1, pp. 62-70, 2001. (in Russian)
- [11] V. A. Pronina and L. B. Shipilina, “Using relations between attributes to construct domain ontology,” *Control problems*, No. 1, pp. 27–32, 2009. (in Russian)
- [12] Ning, L., Guanyu, L., Li, S., Using formal concept analysis for maritime ontology building: *International Forum on Information Technology and Applications 2*, 2010, pp. 159–162.
- [13] P. Cimiano, A. Hotho, G. Stumme, J. Tane, *Conceptual Knowledge Processing with Formal Concept Analysis and Ontologies: Eklund, P.W. (ed.) ICFA 2004. LNCS (LNAI)*, vol. 2961, Springer, Heidelberg, 2004, pp. 189–207.
- [14] Obitko, M., Sinsel, V., Smid, J., *Ontology design with formal concept analysis: Proc. of the CLA2004 International Workshop on Concept Lattices and their Applications*, 2004, pp. 1377–1390.
- [15] H.M. Haav, An Application of Inductive Concept Analysis to Construction of Domain-specific Ontologies: *Proceedings of the Workshop of VLDB2003*, 2003, pp. 63-67.
- [16] R. Bendaoud, A. Napoli, Y. Toussaint, *Formal Concept Analysis: A unified framework for building and refining ontologies: A. Gangemi and J. Euzenat editors, Knowledge Engineering: Practice and Patterns – Proceedings of the 16 International Conference on Knowledge Engineering and Knowledge Management, EKAW’2008*, pp. 156-171.