REALIZATION OF THE ONTOLOGICALLY BASED METHOD FOR CHECKING STRUCTURAL INCONSISTENCES OF RELATIONAL DATABASES

Alexander A. Kropotin, Andrei V. Grigoryev, Yuliya V. Bidulya,
Alexander G. Ivashko & Nikita S. Durnin

This Publication has to be referred as: Kropotin, A[lexander]; Grigoryev V., A[ndrei]; Bidulya V., Y[uliya]; Ivashko G., A[lexander] & Durnin S., N[ikita] (2016). Realization of the Ontologically Based Method for Checking Structural Inconsistences of Relational Databases, Proceedings of the 27th DAAAM International Symposium, pp.0762-0767, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-08-2, ISSN 1726-9679, Vienna, Austria
DOI: 10.2507/27th.daaam.proceedings.110

Abstract

In paper the problem of use of technologies of a semantic web for automatic detection of structural inconsistencies of relational databases in case of software integration is solved. The description of the conflicts of relational databases in case of their integration is provided in chapter 1, methods of their determination and reasons for development of own method of the solution of an objective. The description of a method of automatic detection of structural mismatches of relational databases with use of technologies of a semantic web is provided in chapter 2. The description of experimental evaluating of the developed method is provided in a chapter 3 and its results are represented. In the inference plans for future development in this area are defined.

Keywords: Ontology; description logics; database; ER-model.

1. Introduction

At present the integration of information systems and their databases is one of the popular research task [1, 2, 3]. Data integration may be formally defined like synchronization of data between databases of the software which can be provided by different data schemes in different sources.

A lot of scientific schools around the world take part in researches related to information systems integration, such like: Database Technology University of Zurich [1], Wright State University [2], Dublin City University School of Computing [3] and other. Besides, there are different methods and approaches to integration of data [4, 5, 6, 7], but all of them face with the problems of integration of data connected to a structural mismatch of data schemes [4], which may be defined as conflicts related to

- a) schemes inconsistencies, it may arise in solving problems of integrating schemes;
- b) data inconsistencies, it may arise in solving problems of integrating schemes.

The following data schemes differences may be associated with schemes inconsistencies[4, 5]:

- a) a difference in terminologies: homonymy and synonymy of data in different sources;
- b) a difference in data semantics: equivalent real-world entities are organized at various levels of abstraction in different sources;
- c) a difference in data structure: equivalent real-world entities organized in different data structures in different sources.

In turn, semantic (b) and structure (c) differences in these schemes may be defined as:

- d) a difference in the cardinality of relationships set: equivalent relationships set from different sources have a different cardinality;
- e) a difference in the degree of participation of relationships set: equivalent relationships set from different sources have different degrees of participation;
- f) an absence of elements of schemes: an absence of value sets, entities sets, relationships sets in different sources;
- g) a difference in «entities sets $\rightarrow$ value sets»: equivalent real-world entities represented by a entities sets in some sources and a value sets in the other;
- h) a difference in the «value sets $\rightarrow$ set value sets»: properties of equivalent real-world entities are represented in value sets of some sources and a set of value sets in the other;
- i) a difference in the restrictions value sets: restrictions on permissible and existing values in a difference of sources;
- j) a difference in the types of data: equivalent value sets of different sources presented different types of data;
- k) a difference in the set of valid values: equivalent value sets of different sources identified different sets of constants;

The difference in «entities sets $\rightarrow$ value sets» (a) as well as differences in relationships sets (c, d), the absence of value sets, entities sets and relationship sets may be associated with a different view on the logical level of conceptual designs as:

- a) entities sets that come into connection 1:1;
- b) entities sets that come into relationships sets superclass / subclass;

Also, the absence of value sets, entities sets and relationships sets (c, d) can be associated with a different view on the logical level of conceptual designs as unary (recursive) connections.

The following data differences may be associated with data inconsistencies [4, 5]:

- a) difference of data format $\rightarrow$ presentation equivalent values in different sources define different formats;
- b) difference values $\rightarrow$ equivalent values in different sources have different views.

From the assumption that by means of application of means of a semantic web [8] perhaps implementation of verification of several diagrams of relational databases regarding existence of their structural distinctions the task of their verification using a formalism of description logics from each other appeared.

In the course of search of the existing solution of an objective operations in which mathematical models and algorithms of ontological submission of diagrams of relational databases in terms of description logics are described [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21] were found and probed. The study of these operations led to an output that submission of diagrams of relational databases in ontology is possible, however this representation doesn't allow to solve an objective.

All these facts prove actuality of this paper. And thus, the main idea of paper is a development and realization of the ontologically based method for checking structural inconsistencies of relational databases [25] during software integration.

2. Method description

We propose the OWL DL implementation to detect problems during the process of data schemes integration. The developed method includes three steps:

- a) at the first step, each conceptual schema of relational database must be mapped into ontology.
- b) at the second step these ontologies have to be merged into one.
- c) at the last step the reasoning for this ontology have to be implemented.

Special for realization of the method steps were developed three models:

a) the ontological model of a conceptual domain objects representation has been developed special for realization of the method first step. The model is presented as a rules for describing database concepts into description logics terms. The model based on the expressive description logic SROIQ.

b) the model of ontologies union of an conceptual domain objects has been developed special for realization of the method second step.

c) the model analysis of an united ontology has been developed special for realization of the method last step.

The detailed description of the developed models has been submitted in [22] therefore further, we will provide their short description in the form of rules of the mapping binary entity-relationship diagram elements into ontology elements.

2.1. The mapping conceptual schema of relational database into ontology

Following we present some rules for mapping of semantic network "binary entity-relationship" models to OWL DL ontology. The set of terminological axioms and assertional axioms is proposed to term the conceptual domain objects ontology (1).

$$\mathcal{K} = \mathcal{T} \cup \mathcal{A}, \quad (1)$$

where $\mathcal{K}$ is the conceptual domain objects ontology; $\mathcal{T}$ is a terminological axioms; $\mathcal{A}$ is an assertional axioms.

Entity set and value set. Both entities sets and values sets should be mapped into class. While transforming entities sets and values sets, it should be taken care of tagging classes names to correspond to entities sets names and values sets names. Each class should be made as disjoint with other.

Entity and value. Both entities and values would be mapped into individuals. While transforming entities and values, it should be taken care of making local unique names. Each individual should be made as different with other in frame of one class, when it putted in line with entity.

Relationship set and attribute. Both relationships sets and attributes should be mapped into object properties. While transforming relationships sets and attributes, it should be taken care of tagging object properties names to correspond to relationships sets and attributes names.

While transforming attributes, in addition to object properties datatype properties should be created. Thus, each attribute should be mapped into composition of object property and datatype property. While transforming attributes, it should be taken care of making local unique names.

Each object property and datatype property should be made as functional, otherwise OWL DL allows object property and datatype property to take many values by default.

While transforming relationship sets cardinality of «one-to-many», in addition for object properties would be created and tagged as inverse object properties. While transforming relationship sets cardinality of «one-to-one», object properties would be tagged as symmetric. Thus relationship sets would be mapped into two inverse object property or one symmetric object property for each.

The domain and the range of each object property is the class. The domain of each data property is the class and range is the actual datatype (int, string, etc).

Primary key and unique constraint. Each datatype property and object property, when it putted in line with primary key attribute, should be tagged as hasKey of their class.

2.2. The merging rules of ontologies conceptual domain objects

The merging ontologies domain conceptual objects (1) is proposed to term ontology of set of domain conceptual object different representation (2).

$$\mathcal{K} = \bigcup_{1 \leq l \leq L} \mathcal{K}_l, \quad (2)$$

where $\mathcal{K}$ is the ontology of set of domain conceptual object different representation; L is the count of different ontology union $\mathcal{K}_l$: $L \geq 2$; $\mathcal{K}_l$ is the domain conceptual objects ontology.

Then in this case in integrated terminological axiom set $\mathcal{T} \subseteq \mathcal{K}$ should be defined equal atomic concepts (1) of different ontologies in terminological axiom terms of a concepts equivalence (3).

$$\{\{R_i^l \equiv R_i^h\}_{i=1, I^l} \}_{l < h \leq L} \}_{l=1, \overline{L}}, \quad (3)$$

where R_i^l, R_i^h are object properties and datatype properties defined on the ontology $\mathcal{K}_l$ and $\mathcal{K}_h$ respectively; I^l is a count of all object properties and datatype properties defined on the ontology $\mathcal{K}_l$; L is a count of all ontologies $\mathcal{K}_l$.

In addition, in integrated terminological axiom set $\mathcal{T} \subseteq \mathcal{K}$ equal classes (4) of different ontologies should be defined in terminological axiom terms of a concepts equivalence (4), (a) if the count of ontology $\mathcal{K}_l$ object properties is greater than count of ontology $\mathcal{K}_h$ or (b) otherwise.

$$\left\{ \begin{array}{l} \{\{A_n^l \equiv A_n^h\}_{n=1, N^l} \}_{l < h \leq L} \}_{l=1, \overline{L}} \quad (a) \\ \left\{ \left\{ A_n^l \prod_{i=1}^I \exists S^l. C^i \equiv A_n^h \right\}_{l < h \leq L} \right\}_{l=1, \overline{L}} \quad (b) \end{array} \right\} \quad (4)$$

where A_n^l, A_n^h are classes of the ontologies $\mathcal{K}_l$ и $\mathcal{K}_h$ respectively; S^l is the transitive object property of the ontology $\mathcal{K}_l$; I is a count of all properties of the concept A_n^l are putted in line with properties of the concept A_n^h which have not equal roles from the ontology $\mathcal{K}_l$; N^l is a count of all classes of the ontology $\mathcal{K}_l$; L is a count of the all ontologies $\mathcal{K}_l$.

In case of integrated assertional axioms the set $\mathcal{A} \subseteq \mathcal{K}$ contains at least two assertional axioms sets $\mathcal{A}_l \subseteq \mathcal{A}$: $\mathcal{A}_l \neq \emptyset$, in which individuals were defined, in these assertional axioms set the individuals equivalence of different ontologies $\mathcal{K}_l$ should be defined in assertional axioms terms (5).

$$\{\{a_j^l \approx a_j^h\}_{j=1, J^l} \}_{l < h \leq L} \}_{l=1, \overline{L}}, \quad (5)$$

where a_j^l, a_j^h are individuals of the ontologies $\mathcal{K}_l$ и $\mathcal{K}_h$ respectively; J^l is a count of all individuals of the ontology $\mathcal{K}_l$; L is a count of all ontologies $\mathcal{K}_l$.

2.3. The checking structural inconsistencies of relational databases

We propose to implement the description logics calculus [8, 23, 24] to ontology of set of domain conceptual object different representation for checking structural inconsistencies of relational databases.

Thus it is necessary to resolve three algorithmic problems of description logics calculus, such that:

- a) the inconsistency algorithmic problem of an ontology terminological (only Tbox);
- b) the classification algorithmic problem of an ontology terminological (only Tbox);
- c) the inconsistency algorithmic problem of an ontology (Tbox and Abox);

The resolving of inconsistency algorithmic problem of an ontology terminological could be able to follows detecting relational databases integration problems [4, 5, 6, 7, 8]:

- a) difference in the cardinality of relationships set the domain and/or domain range do not overlap;
- b) difference of the restrictions value sets range of roles (concrete domain) do not intersect;
- c) difference of the types of data range of roles (concrete domain) do not intersect;
- d) difference of the set of valid values range of roles (concrete domain) do not intersect.

The resolving of classification algorithmic problem of the terminological ontology could be able to follow detecting semantic network "binary entity-relationship" models integration problems:

a) an absence of elements of schemes 1) if there is no composition of atomic roles, then transitive role is not completed in the place of the absent, and one concept is embedded in the another, but about another we could allege this, so they are not equivalent; 2) absence of equivalent concepts from concepts of one of terminology; 3) absence of equivalent concepts from concepts of one of terminology;

b) difference of «entities sets $\neg$ value sets» equivalent concepts belong simultaneously to concepts EntitySet and ValueSet;

c) difference of the «value sets $\neg$ set value sets» 1) equivalent concepts has no concepts in a one terminology; 2) one concept in one terminology has many equivalent concepts in the other;

The resolving of inconsistency algorithmic problem of an ontology could be able to follows detecting relational databases integration problems:

- a) difference in the modality of relationships set null value is defined as {nominal} implication for all Role with not a Concept;
- b) difference values equivalent individuals belong to equivalent roles with different values.

3. Evaluating

Software implementing presented method was developed. It get database descriptions from Papyrus XML file and transforms it to knowledge bases. Software was developed using Java language, because of there is a library called OWL API that allows creating knowledge bases in OWL language. Besides, there are many open source reasoning software allowing solving task of logical reasoning. To check same names of the tables we use greedy algorithm with similarity checking. In the future researches we consider to use a method based on the Kun algorithm for max weighted matching in bipartite graph.

HermiT reasoner is using as reasoning tool, because of HermiT shows great results on past OWL Reasoning Evaluation Workshop [24].

Unfortunately, we test our software only during development on test databases with no more than 10 tables and tests shows 100% regular results (table 1).

Distinction title	Count of structural distinctions	The method result
distinction "entity set – value set"	5	5
distinction "value set – set of value set"	5	5
distinction in a set of admissible values	5	5
distinction in data type	5	5
distinction in semantics of communications	5	5

Tab. 1. The method evaluating result.

4. Conclusion

This paper describes a problem of automatic checking structural inconsistencies of relational databases. This paper presents a framework for the detecting relational databases integration problems. We have provided rules to transform relational databases concepts into equivalent OWL ontology for semantic web.

This proposed framework helps software engineers in upgrading the structured analysis and design relational databases. Our ongoing research on this topic is to handle other cases of relationships that are not binary, which require reification.

Software implementing presented method was developed. It get database descriptions from Papyrus XML file and transforms it to knowledge bases. HermiT reasoner is using as reasoning tool. Software was developed using Java language and OWL API.

Unfortunately, we test our software only during development on test databases with no more than 10 tables and tests shows 100% regular results.

In the future, we will extend our researches, we plan to develop, and present ready software tested on enterprise databases. Moreover, there are few systems with opened database creating scripts, and to build knowledge bases we will use data description using data definition in web services. For example, WSDL-definition with a XSD schemas can be considered as a source of the database schema.

5. Acknowledgements

We would like to thank our colleagues who have helped in the development of independent parts of the software, but who wish to remain outside the scope of the co-authors of the article: Shaykhullina Irina, Michael Samoilov, Alexander Stupnikov and Dusso Bruno Jean-Marie.

6. References

- [1] Patrick Ziegler, Klaus R. Dittrich, "Data Integration — Problems, Approaches, and Perspectives." In *Conceptual Modelling in Information Systems Engineering*, Springer, 2007. (Conference or Workshop Paper). p. 39-58.
- [2] Adila Alfa Krisnadhi. "Ontology Pattern-Based Data Integration", PhD Thesis, Wright State University. December 18, 2015.
- [3] Pahl, Claus and Zhu, Yaoling. "Data integration in mediated service compositions" In *Computing and Informatics*, Volume 31 Issue 6, 2001, pp. 1129-1149.
- [4] Beloshitsky D. A. Data Integration in Information Systems - Electron. Dan. – [Electronic resource] / D. A. Beloshitsky. – Access: <http://sntbul.bmstu.ru/doc/602635.html> – (reference date: 20. 03. 2015).
- [5] Ramon Lawrence, Ken Barker. "Integrating relational database schemas using a standardized dictionary." In *Proceeding SAC '01 Proceedings of the 2001 ACM symposium on Applied computing*. ACM New York, NY, USA, 2001, pp. 225-230.
- [6] Stephen Mc Kearney, "Schema Integration.", 2002. Available at URL: <http://www.smckearney.com/adb/notes/lecture.schema.integration.pdf>.
- [7] Batini C., Lenzerini M., Navathe S. B. "A comparative analysis of methodologies for database schema integration." In *Journal ACM Computing Surveys*, Volume 18 Issue 4, Dec. 1986, pp. 323-364.
- [8] Gruber, T. A Translation Approach to Portable Ontology Specifications / T. Gruber // *Knowledge Acquisition* – 1993. – pp.199-220.
- [9] Astrova, I. Mapping of SQL relational schemata to OWL ontologies / I. Astrova, A. Kalja // *Proceedings of the 6th WSEAS International Conference on Applied Informatics and Communications*. 18-20 August 2006, Elounda, Greece / Institute of Cybernetics Tallinn University of Technology Akadeemia tee 21, 12618 Tallinn Estonia. – P. 375-380.
- [10] Barrasa Rodríguez, J. R2O, an Extensible and Semantically based Database-to-Ontology Mapping Language / J. Barrasa Rodríguez, O. Corcho, A. Gómez-Pérez // *Proceedings of the Second Workshop on Semantic Web and Databases*. August 2004, Toronto, Canada / Springer-Verlag, Berlín, Alemania. – P. 1069-1070
- [11] Berners-Lee, T. The Semantic Web / T. Berners-Lee, J. Hendler, O. Lassila, // *Scientific American* – 2001. – Vol. 284 – pp. 34-43.
- [12] Bumans, G. Mapping between Relational Databases and Owl Ontologies: an example / G. Bunmas // *Computer science and information technologies* – 2010. – Vol. 756 – pp. 99-117.
- [13] Chujai, P. On Transforming the ER Model to Ontology Using Protégé OWL Tool / P. Chujai, N. Kerdprasop, K. Kerdprasop // *International Journal of Computer Theory and Engineering* – 2014. – Vol. 6 – pp. 484-489.
- [14] Colomb, R. M. Issues in Mapping Metamodels in the Ontology Development Metamodel Using QVT / R. M. Colomb, A. Gerber, M. Lawley // *Proceedings of the First International Workshop on the Model-Driven Semantic*

- Web. The First International Workshop on the Model-Driven Semantic Web. 20-24 September 2004, Monterey, California.
- [15] Cullot, N. DB2OWL: A Tool for Automatic Databases-to-Ontology Mapping / N. Cullot, R. Ghawi, K. Yétongnon // Proceedings of the 15th Italian Symposium on Advanced Database Systems. 2007. – P. 491-494.
- [16] Fahad, M. ER2OWL: Generating OWL Ontology from ER Diagram / M. Fahad // Intelligent Information Processing IV – 5th IFIP International Conference on Intelligent Information Processing. The International Federation for Information Processing Volume 288. 19-22 October 2008, Beijing, China / Mohammad Ali Jinnah University, Islamabad, Pakistan. – P. 28-37.
- [17] Louhdi, M. R. C. Transformation Rules For Building Owl Ontologies From Relational Databases / M. R. C. Louhdi, H. Behja, S. O. El Alaoui // Second International Conference on Advanced Information Technologies and Applications. November 2013. P. 271-283.
- [18] Shihan, Y. Semi-automatically building ontologies from relational databases / Y. Shihan, Z. Ying, Y. Xuehui // Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology. 2010. – P. 150 - 154.
- [19] Telnarova, Z. Relational database as a source of ontology creation / Z. Telnarova // Proceedings of the International Multiconference on Computer Science and Information Technology. 2010. – P. 135-139.
- [20] Natalya F. Noy. “Semantic integration: a survey of ontology-based approaches.” In Newsletter ACM SIGMOD Record Homepage archive, Volume 33 Issue 4, December 2004. pp. 65-70.
- [21] AnHai Doan, Alon Y. Halevy. “Semantic-integration research in the database community.” In Journal AI Magazine - Special issue on semantic integration archive, Volume 26 Issue 1, March 2005 pp. 83-94.
- [22] Andrey V. Grigoryev, Alexander A. Kropotin, Alexander G. Ivashko. Database Schema Method for Automatic Semantic Errors Resolving During Information Systems Integration. Proceedings of the 10th AICT2016 International Conference, pp.187-190, IEEE CFP1656H-PRT, 2016, Baku, Azerbaijan.
- [23] Ivashko A.G., Grigoryev A.V., Grigoryev M.V. “Tableau algorithm modification based on complex concepts disjointness checking.” In Bulletin of the Tyumen State University. - 2012, №4. - P. 143-150.
- [24] Birte Glimm, Ian Horrocks, Boris Motik, Giorgos Stoilos, Zhe Wang. “Hermit: An OWL 2 Reasoner” In Journal of Automated Reasoning October 2014, Volume 53, Issue 3, pp 245-269.
- [25] Salibekyan Sergey & Panfilov Peter. Database Architecture for Specifying and Modeling Spatio-Temporal Relations, Proceedings of the 26th DAAAM International Symposium, pp.0589-0598, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-07-5, ISSN 1726-9679, Vienna, Austria