Application of an Instance Migration Solution to Industrial Ontologies

Maxim Davidovsky[1], Vadim Ermolayev[1] and Vyacheslav Tolok[2]

[1] Department of IT, Zaporozhye National University,
66 Zhukovskogo st., 69600 Zaporozhye, Ukraine
m.davidovsky@gmail.com, vadim@ermolayev.com

[2] Department of Mathematical Modeling, Zaporozhye National University,
66 Zhukovskogo st., 69600 Zaporozhye, Ukraine
vyacheslav-tolok@yandex.ru

Abstract. The paper presents the results of evaluating the software solution for ontology instance migration problem in the use case involving the ontologies used in construction industry – freeClass and eClassOWL with the BauDataWeb dataset representing the individuals. Ontology instance migration problem is understood as a sub-problem of ontology alignment. Our methodology assumes (semi-) automated iterative process possibly involving a human for validating the results. The process consists of the two steps: (1) schema-based mappings discovery done by the agent-based matcher software; and (2) ontology instance transformation and migration according to the discovered mappings done by the ontology instance migration engine software. The evaluation experiment has been conducted in two phases and yielded results of acceptable quality in terms of precision, recall, and f-measure.

Keywords. Ontology alignment, industrial application, ontology instance migration, evaluation experiment.

Key terms. Industry, Integration, Interoperability, KnowledgeManagementProcess, AgentBasedSystem,

1 Introduction

Ontologies are being widely adopted today in the academic world and increasingly attract the attention of researchers and practitioners in information technology and knowledge-based system development and applications. Many authors, e.g. [1], argue that ontologies constitute the substance of the advanced technologies for solving the problems of interoperability, communication, and cooperation between different applications within the same environment. Indeed, ontologies conceptualize semantics of the domains within a discourse that are common for interoperating systems. Thus, ontologies serve as a bridge for "understanding" between the systems or their parts. Despite that, application of ontologies in industry still faces several problems.

The first group of problems concerns the inertia that is typical for the process of application of advanced technologies in industry, e. g. [2]. This paper reflects the views of the practitioners who have witnessed incomprehension and opposition in trying to solve customer problems using ontologies. These problems are attempted to be resolved through establishing a closer contact with domain knowledge stakeholders and their more active involvement in the development of ontologies – e.g. [3]. Another complementary and important activity is lowering the effort for developing ontologies which could be done via providing the tool support for domain experts taking part in ontology development.

The other important stratum of problems in the application of ontologies in industry is related to the re-use of existing large industrial knowledge bases, collections, or ontologies and the exploitation of those knowledge assets within large enterprise information systems (IS). Obviously it is obligatory to provide stable interoperation of ISs in industrial settings to prevent substantial errors in maintenance, production, and sales. However the use of ontologies per se doesn't completely solve interoperability issues as it essentially raises heterogeneity problems to a higher level [4]. So, the methods for aligning ontologies need to be provided to understand and explicitly specify semantic mappings between these different conceptualisations. Industrial ontologies as a rule contain large quantity of individuals (or instances). Hence, an important and typical sub-problem of ontology alignment in industrial settings is ontology instance migration that is the process of transferring instances between aligned ontologies. The numbers of the individuals in industrial knowledge bases is very often high, so their manual alignment is not feasible. Therefore it is important to provide the tools that at least partially automate the process of alignment and do that with the quality acceptable for industries. Another important aspect of the use of ontologies in industrial settings is that industrial ISs are often distributed and belong to autonomous business entities. In such settings using intelligent software agents for ontology alignment and ontology instance migration in particular becomes an attractive implementation pathway.

The remainder of the paper is organized as follows. In section 2 we provide a classification of industrial applications of ontology alignment types of problems and describe some typical use cases. Based on this classification, we analyze industrial requirements to ontology alignment solutions. Section 3 outlines our software solution for ontology instance migration problem. Section 4 reports about the setup and results of our evaluation experiments. Finally the conclusion is given and the plans for the future work are outlined.

2 Related Work, Applications, and Use Cases

Surveys of ontology alignment for a wide range of applications can be found in [5], [6], [7]. Applications of agent-based ontology alignment and respective requirements are analyzed in [8]. This paper focuses on industrial applications of ontology alignment in broad and ontology instance migration as its sub-problem [8].

The following industrial application categories may be outlined that require ontology alignment and instance migration solutions.

1. Industrial knowledge-driven simulation models. Simulation models are widely used in industry ([9], [10], [11]). The complexity level of modern simulation systems requires the use of knowledge-based models. This knowledge may be related to various branches of science, engineering disciplines, can contain different models satisfying different demands. This requires the use of ontologies and related activities such as ontology merging and alignment.

2. Industrial information systems in the context of Semantic Web and eCommerce. eCommerce is a type of industry where buying and selling of product or service is conducted over electronic systems such as the Internet and other computer networks. In order to perform such an exchange of business information, this information must contain product (or service) descriptions. As a rule such information is presented in the form of product or service ontology [12]. Good examples of such ontologies are [13], [14], [15]. When a business process involves more than one party or in a case of using more than one source respective ontologies obviously have to be aligned. This situation is also typical for The Semantic Web where ontologies along with intelligent software agents are the main pillars [16].

3. Integration and interoperability of heterogeneous enterprise ISs. Today information ecosystem of a modern enterprise as a rule contains numbers of applications from different vendors and used for different purposes. In order to effectively use these heterogeneous applications together with distributed data and knowledge repositories they must be integrated into a single system. Likewise implementation and deployment of new software solutions must be reconciled and integrated with legacy software systems. Here ontologies may be used not only as domain knowledge representation models, but also as mediators for integration of heterogeneous applications. Enterprise integration attracts substantial interest of research community and a number of solutions are proposed (e.g., [17], [18]).

4. Knowledge sharing and migration between enterprise ISs. Interaction and cooperation of modern enterprises often implies knowledge sharing and migration. In such a way enterprise may enrich and harmonize their knowledge assets. In this case knowledge models obviously must be reconciled and aligned. This issue is not widely addressed in literature (but some early efforts, e.g. [19], are described) as it usually requires some (combination of) typical ontology management activities (such as ontology evolution and knowledge sharing – please see some details above).

Each of the application categories sets up some requirements to specific alignment methods used within the category. Due to the wide variety of ontologies used in industry it is difficult to set up a detailed set of requirements for ontology matching methods. These requirements may substantially vary depending on ontology size and structure so we outline only the most general observations. We analyze the requirements for ontology alignment regardless to industrial application in [8].

Run-time. 1st and 2nd categories assumed the matching process to be performed at run-time. In that case the maximum level of automation must be reached. In 3rd and 4th categories it is allowed to perform matching and relative activities previously and separately. This allows active involvement of experts to the matching process (for alignment validation, relevance verification, etc.).

Completeness. Completeness is of the most importance in the 1st and 3rd cases. It is important not to miss knowledge in these cases. At the same time, in the 2nd category the response time of method implementation to a system query is more critical as in that case matching is usually performed during runtime.

Relevance. In the 4th case, the relevance of knowledge is the most critical (particularly during migration from an older system to a newer one). Here it is first of all important to save actual knowledge, but some obsolete knowledge may be discarded.

3 Solution Overview

The main focus of the paper is evaluation of ontology alignment and instance migration methodology in industrial settings. The methodology assumes (semi-) automated iterative process of ontology alignment and instance migration with possible human intervention for checking the correctness and setting up the process. The overall methodology consists of two steps: (1) mapping discovery and determination of structural differences between ontologies and (2) ontology instance transformation and migration according to the determined differences.

The first step is essentially the process of ontology matching with the only difference that it results not only in ontology alignment but also produces an output of a set of transformation rules that further drive the process of ontology instance migration. The solution for the first step is based on the implementation of meaning negotiation between intelligent agents (we call this agent-based solution ABOA matcher [8]). The matching process embodies the strategy that originates from [20] and is described in detail in [21]. Negotiations among the agents are conducted in an iterative way and with an aim to reduce the semantic distance between the negotiated structural contexts of the respective ontology schemas. A negotiation is stopped when the distance reaches a commonly accepted threshold or the parties exhaust their propositions and arguments.

At the second step agents use Instance Migration Engine in order to transfer instances between ontologies based on the transformation rules generated at the first step. Instance migration results in the transfer of all the assertions that do not require the resolution of the problem cases by the ontology engineer. The cases that caused problems are recorded in the migration log. The details on the second step of the methodology are described in [22].

4 Evaluation Experiment

To test our methodology and solution of ontology alignment and instance migration we choose real industrial ontologies: freeClass[1] ontology for construction and building materials and services and eClassOWL[2] [14] – the web ontology for products and services. The dataset of the European building and construction materials market for the Semantic Web (BauDataWeb[3]) has been selected as the set of assertions for migration. Structural parameters of the ontologies are presented in Table 1. General experimental set-up specified in ISO/IEC 24744 notation for describing methodologies [23] is pictured in Figure 1.

Table 1. Structural parameters of industrial ontologies used in the second experiment

	Total number of axioms	Number of logical axioms	Number of classes	Number of object properties	Number of data properties	Number of individuals
freeClass	78414	9622	5231	168	3	1335
eClassOWL	360243	117090	60662	4900	2453	4766
BauDataWeb	-	-	-	-	-	Over 60 million instances

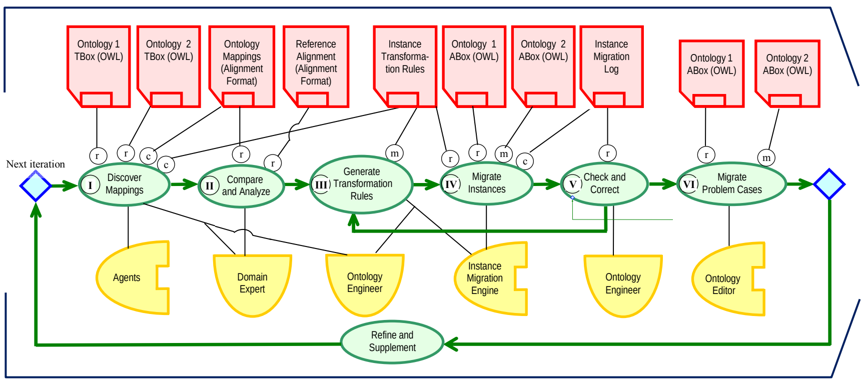

Fig. 1. The set-up of the evaluation experiment

The test case doesn't contain any reference alignment. Hence, we had to determine reference mappings manually in order to objectively judge about the obtained results. For convenience both freeClass and eClassOWL ontologies may be divided into 2 parts. The first parts are actually the sets of entities directly inherited from the

[1] http://www.freeclass.eu/ – the ontology for construction and building materials and services
[2] http://www.heppnetz.de/projects/eclassowl/ – the web ontology for products and services
[3] http://semantic.eurobau.com/ – BauDataWeb: the European Building and Construction Materials Database for the Semantic Web

GoodRelations ontology[4] [13] and also some concepts from other common-sense vocabularies. The schemas of those parts of the ontologies are almost identical, so the difference is mostly in the sets of individual assertions. Further, those parts do not cause any problems in the discovery of the reference mappings as the entities mainly have human-understandable names and labels. Based on the analysis of the above-mentioned parts of the ontologies we constructed the set of reference mappings (further mentioned as Alignment 1). The second parts of the ontologies consist of internal entities that do not have understandable names (the names represent some identifiers composed of numbers and characters), but some of them still have labels with descriptions. Due to the big quantity of those entities we did not analyze the whole sets and choose the 20 entities that are semantically similar. Then we discovered respective mappings for those chosen entities (further mentioned as Alignment 2). The parameters for both alignments are presented in Table 2 where for brevity we include only the information about the classes and properties.

Table 2. Parameters of reference alignment for the experiment with the BauDataWeb dataset

	Number of mapped entities		
	Classes	**Object properties**	**Datatype Properties**
Alignment 1	53	55	53
Alignment 2	20	11	0

Thus, the experiment with the BauDataWeb dataset has been performed in two phases. Within the first phase we constructed the reference alignment (Alignment 1) and started the matching process using the ABOA matcher. Then we found the mappings that correspond to Alignment 1 and compared them to the reference ones. Alignment quality values for the results of this step are very high (Table 3, row 1) as these parts are almost identical.

Table 3. Matching results

Experiment Step	**Alignment Quality Measures**		
	Precision	**Recall**	**F-Measure**
1	0.99999	0.99999	0.99999
2	0.69552	0.42384	0.52671

It might be considered that the Alignment 1 in our experiment is not a topically interesting case as the semantic differences are tiny and could be easily discovered manually. However, this experimental phase represents a good case for validating the generated instance transformation rules and instance migration quality. In this phase all of the generated transformation rules were correct. More details on the transformation rules could be seen in [22]. Within the second phase we determined the Alignment 2 and tried to find the respective mappings within the alignment discovered by the matcher. The alignment quality measures for the second phase are lower than for

[4] http://www.heppnetz.de/projects/goodrelations/ – the web vocabulary for e-commerce

the phase 1, which is conditioned by the relatively weak semantic similarity between the structural contexts [20] that correspond to these parts of ontologies. It is also worth noticing that the Precision value is noticeably higher than the Recall one within phase 2. It is so because string-based structural similarity measurement methods yield high values on labels. Labels can contain parts (e.g. words) that are common for many of them, but respective entities in general are not semantically similar. For example the comparison of labels "construction technology" and "pump technology" will give noticeably high similarity values. However those labels belong to the entities that are obviously not that similar semantically.

5 Concluding Remarks and Future Work

The paper presented the experiment evaluating our methodology and software solution for ontology instance migration on real-world industrial ontologies. The experiment shows acceptable results that allow a positive judgement about the applicability of our methodology in industrial settings. The results also suggest some directions for the future work. The experiment with large ontologies (BauDataWeb dataset) shows that the ontology instance migration engine allows migrating about several million instances using a conventional desktop computer. Hence, a technique to overcome this upper limit is needed for scaling the tool up to the volumes characteristic to Big Semantic Data. Looking for such a technique is on our research and development agenda. In the future we also plan to conduct a series of experiments with the ontologies specified in OWL sublanguages[5] and OWL 2 profiles[6]. Another important direction for the future research is evaluating our approach on ontologies having different structural patterns like a taxonomy (tree-type) structure, a network structure (ontologies rich with object properties), OWL graphs with high and low vertex degrees, etc.

Acknowledgments

The authors are grateful to the colleagues who provided the industrial ontologies for our experiments – Univ. Prof. Dr. Martin Hepp and Dipl. Ing. Andreas Radinger from E-Business and Web Science Research Group of the Universität der Bundeswehr, München.

References

1. Bittner, T., Donnelly, M., Winter, S.: Ontology and Semantic Interoperability. In: D. Prosperi and S. Zlatanova (ed.) Large-scale 3D Data Integration: Problems and Challenges. London, CRCPress (2005)

[5] http://www.w3.org/TR/2004/REC-owl-features-20040210/#s1.3
[6] http://www.w3.org/TR/owl2-profiles/

2. Malzahn, D.: Industrial Application of Ontologies. In: eKNOW 2011, The Third International Conference on Information, Process, and Knowledge Management, (2011)
3. Tatarintseva, O., Ermolayev, V., Fensel, A.: Is Your Ontology a Burden or a Gem? – Towards Xtreme Ontology Engineering. In: Ermolayev, V. et al. (eds.) Proc. ICTERI 2011, CEUR-WS.org/Vol-716, 65-81 (2011)
4. Euzenat J., Shvaiko P.: Ontology Matching. Berlin Heidelberg, Springer-Verlag (2007)
5. Corcho, O.: Methodologies, tools and languages for building ontologies. Where is their meeting point? Data & Knowledge Engineering 46, 41–64 (2003)
6. Ehrig, M.: Ontology Alignment: Bridging the Semantic Gap (Semantic Web and Beyond). Springer (2006)
7. Zhdanova A. V., de Bruijn, J., Zimmermann, K., Scharffe, F.: Ontology Alignment Solution. Deliverable D14 v2.0, (2004)
8. Ermolayev, V., Davidovsky, M.: Agent-Based Ontology Alignment: Basics, Applications, Theoretical Foundations, and Demonstration. Tutorial Paper. In: Dan Burdescu, D., Akerkar, R., Badica, C. (eds.) Proc. WIMS 2012, 11-22, ACM (2012)
9. Silver, G., Hassan, O.H., Miller, J.: From domain ontologies to modeling ontologies to executable simulation models. In: Proc. of the 2007 Winter Simulation Conference, (2007)
10. Novák, P, Šindelář, R.: Applications of ontologies for assembling simulation models of industrial systems. In: Proc. of the 2011th Confederated international conference on the move to meaningful internet systems (OTM'11), pp.148–157, Springer-Verlag Berlin, Heidelberg (2011)
11. Ermolayev, V., Keberle, N., Matzke, W.-E.: An Upper Level Ontological Model for Engineering Design Performance Domain, LNBIP, vol. 20, pp.127–141. Springer, Heidelberg (2008)
12. Ding, Y., Fensel, D., Klein, M., Omelayenko, B., Schulten, E.: The Role of Ontologies in eCommerce. In: Steffen Staab, Rudi Studer (eds.): Handbook on Ontologies. International Handbooks on Information Systems. pp. 593-616, ISBN 3-540-40834-7, Springer (2004)
13. Hepp, M. GoodRelations: An Ontology for Describing Products and Services Offers on the Web, LNCS, vol. 5268, pp. 332–347. Springer Berlin Heidelberg (2008)
14. Hepp, M.: Products and Services Ontologies: A Methodology for Deriving OWL Ontologies from Industrial Categorization Standards. In: Int'l Journal on Semantic Web & Information Systems 2(1) (2006), pp. 72–99, (2006)
15. Morgenstern, L., Riecken, D.: SNAP: An Action-Based Ontology for E-commerce Reasoning. In: Proc., Formal Ontologies Meet Industry, Verona, Italy, (2005)
16. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web (Berners-Lee et. al 2001). Scientific American 284, 28–37 (2001)
17. Izza, S., Vincent, L., Burlat, P.: A Unified Framework for Enterprise Integration: An Ontology-Driven Service-Oriented Approach. In: Pre-proc. of the First International Conference on Interoperability of Enterprise Software and Applications (INTEROP-ESA’2005), pp. 78–89. Geneva, Switzerland, February 23 – 25, (2005)
18. Stoutenburg, S. et al.: Ontologies in OWL for Rapid Enterprise Integration. Time 122, 82–89 (1994)
19. Lochovsky, F. H., Woo, C. C., Williams, L. J.: A micro-organizational model for supporting knowledge migration. In: Proc. of the ACM SIGOIS and IEEE CS TC-OA conference on Office information systems, pp.194–204. Cambridge, Massachusetts, US, (1990)
20. Ermolayev V., Keberle, N., Matzke, W.-E., Vladimirov, V.: A Strategy for Automated Meaning Negotiation in Distributed Information Retrieval. In: Y. Gil et al. (Eds.): ISWC 2005 Proc. 4th Int. Semantic Web Conference (ISWC'05), 6–10 November, Galway, Ireland. LNCS 3729, pp. 201–215 (2005)

21. Davidovsky, M., Ermolayev, V., Tolok, V.: Agent-based implementation for the discovery of structural difference in OWL DL ontologies. In: Mayr, H. C., Ginige, A., Liddle, S. (ed.) Proc. Fourth Int United Information Systems Conference (UNISCON 2012). LNBIP 137, Berlin, Heidelberg: Springer-Verlag (2013)
22. Davidovsky, M., Ermolayev, V., Tolok, V.: Instance Migration between Ontologies having Structural Differences. Int. J. on Art. Int. Tools. 20(6), 1127-1156 (2011)
23. Henderson-Sellers, B., Gonzalez-Perez, C.: Standardizing Methodology Metamodelling and Notation: An ISO Exemplar. In: Kaschek, R., Kop, C., Steinberger, C., Fliedl, G. (eds.) UNISCON 2008. LNBIP, vol. 5, pp. 1–12. Springer, Berlin/Heidelberg, (2008)