DE GRUYTER OPEN
Information Technology and Management Science
doi: 10.1515/itms-2014-0012
2014 / 17

Evaluation of the Ontological Knowledge Model

Darya Plinere [1], Arkady Borisov [2], [1, 2] *Riga Technical University*

Abstract **– An ontology is an explicit formal conceptualization of some domain of interest. Ontology application is widely used in multi-agent systems. Ontology provides the channel through which software agents interact; therefore, if the ontology definitions have not been sufficiently evaluated, communication between software agents may not succeed. Ontology evaluation guarantees to end users the correctness and completeness of ontology definitions and software. The goal is to detect wrong, incomplete or missed definitions in the ontology. The study is based on multi-agent system application for raw materials management task. The common ontology is used for agent cooperation; the quality of ontology is measured with ontology evaluation techniques.**

Keywords **– Application-based ontology evaluation, data-driven ontology evaluation, golden standard, ontology evaluation.**

I. Introduction

Ontology evaluation task is to measure the quality of the ontology. Ontology makes sure that the meaning of the data exchanged between and within systems is consistent and shared by computers (expressed by formal models) and humans (as given by their conceptualization). Ontology allows all participants to speak a common language [1].

Ontology, as well as all engineering artifacts, needs a thorough evaluation. However, ontology evaluation poses some unique challenges: due to the declarative ontology nature developers can not just compile and run it like most of other software artifacts. This is data that have to be shared between various components and used for potentially different tasks [1].

Its role is to provide a formal semantics to terms in order to use them in the machine processable way. Ontology allows to share and formalize conceptualizations, and thus to enable humans and machines to easily understand the meaning of the data exchanged. This allows automatically aggregate, use and reuse distributed data sources, thereby creating an environment in which agents and applications can work together for the benefit of the user to still inexperienced level [1].

The central role of ontology in the agent system makes the evaluation of the ontology an important and worthwhile task: mistakes or omissions in the ontology can lead to not realizing the full potential of the data exchanged in the application. Good ontology leads to a higher degree of reuse and better interaction [1].

Some examples of disadvantages of the low-quality ontology: ontology readability may be affected if the vocabulary or syntax contains errors; reasoners may not be able to infer the answers in case of conflicting semantics. Unspecified ontology prevents automated ontology mapping approaches. On the other hand, high quality ontology can be easily reused, can be featured more easily in the existing application, and will be easier to detect and actively omit the errors in the data [1].

II. Raw Materials Management Task for Ontology Evaluation Technique Application

This paper presents a study based on the use of multi-agent system for the task of raw material management. The example from the field of microelectronics, in particular, the company producing chips is discussed in this paper. Chip production requires raw materials, for example, crystals timely saturation in the warehouse. Today, the purchase of crystals is made by the purchasing department manager by phone, fax and e-mail.

Multi-agent systems are widely used for the tasks of supply chain management, particularly for raw materials management tasks. Multi-agent system provides a decentralized system; the global behavior of the system is determined by individual behavior of agents who use their own behavior rules, exist in a shared environment and interact with the environment and other agents. This makes multi-agent system most suitable for this task. Agents use a common ontology or taxonomy for negotiations, thus allowing agents to understand the substance of the negotiations. Ontology in the multi-agent system provides the channel through which software agents interact; therefore, if the ontology definitions have not been sufficiently evaluated, communication between software agents may not succeed. The outputs of application or its performance depend on used ontology in it. Ontology evaluation guarantees to end users the correctness and completeness of ontology definitions and software [2].

The following infrastructure was used for constructing the system: Java and JADE – platform in which agents exist and interact, it gives agents the basic services necessary to their existence, Protégé ontology editor for ontology development, Ontology Bean Generator to convert the domain ontology in JADE classes, MySQL for database support, Apache Ant to compile the code and NetBeans IDE as an integrated development environment [2].

III. Classification of Ontology Evaluation Approaches

Ontologies are not artifacts in a narrow sense, but are expressed by ontology documents, which in turn are artifacts. Evaluation methods are descriptions of procedures that assess a specific quality of an ontology. Since methods cannot asses an ontology directly (since they are not artifacts), methods always directly evaluate ontology documents.

Only indirectly it is possible for an evaluation method to assess an ontology (i. e., by assessing the ontology document that expresses the ontology) [1].

Various approaches to the evaluation of ontologies have been considered in the literature, depending on what kind of ontologies is evaluated and for what purpose. Broadly speaking, most evaluation approaches fall into one of the following categories [3], [4]:

• those based on comparing the ontology to a “golden standard”, which may itself be an ontology [5], [6];

• those based on using the ontology in an application and evaluating the results [7];

• those involving comparisons with a source of data (e. g., a collection of documents) about the domain to be covered by the ontology [8];

• those where evaluation is done by humans, who try to assess how well the ontology meets a set of predefined criteria, standards, requirements, etc. [9], [10], [11].

TABLE I

EVALUATION APPROACHES IN DIFFERENT LEVELS

Level	Golden standard approach	Application-based approach	Data-driven approach	Assessment by humans
Lexical, vocabulary, concept, data	+	+	+	+
Hierarchy, taxonomy	+	+	+	+
Other semantic relations	+	+	+	+
Context, application		+		+
Syntactic	+			+
Structure, architecture, design				+

In addition to the categories of evaluation above, ontology evaluation approaches can be grouped based on the level of evaluation, as described below. An ontology is a fairly complex structure and it is often more practical to focus on the evaluation of different levels of the ontology separately rather than trying to directly evaluate the ontology as a whole. The individual levels have been defined variously by different authors, but these various definitions tend to be broadly similar and usually involve the following levels:

- Lexical, vocabulary, or data layer;
- Hierarchy or taxonomy;
- Other semantic relations;
- Context or application level;
- Syntactic level;
- Structure, architecture, design.

Table I summarizes which approaches from the categories above are commonly used for which of these levels. The detailed description of ontology evaluation approaches based on the level of evaluation is described in [3].

IV. AN OVERVIEW AND APPLICATION OF THE APPROACHES

A. Data-driven Approach

Data-driven approach is that, where the ontology can be evaluated by comparing it to textual documents about the domain that is to be covered by ontology. The authors [8] proposed to compare one or more ontologies with corpus of documents, rather than to compare one ontology to another in order to find the most appropriate one. They extracted a set of relevant domain-specific terms from the corpus of documents, applying Latent Semantic Analysis and clustering method. Then they used WordNet to add two levels of hypernyms to each term in a cluster. And finally, the set of terms is identified in the corpus, mapped to the ontology. The amount of overlap between the domain-specific terms and the terms appearing in ontology can then be used to measure lexical keyword coverage by ontology labels. The ontology can be penalized for terms present in the corpus and absent in the ontology, and for terms present in the ontology, but absent in the corpus. They also proposed a “tennis measure” for two ontologies with the set of identical concepts, which have different organization of concepts, thus have different distance from each other [8], [3].

The document of raw material procurement process is available. Five different ontologies were constructed for the evaluation approach. Having the corpus and five ontologies, the comparison of each one to the corpus was done.

First of all, two questions were asked: how many words from the corpus were in the ontology? How many were not? The overlap between the domain-specific terms and the terms appearing in ontology used to measure lexical keyword coverage by ontology labels. We used the precision and recall method. Precision (1) in this context is the percentage of the ontology lexical entries that also appear in the corpus, relative to the total number of ontology words. Recall (2) is the percentage of corpus lexical entries that also appear as concept identifiers in ontology, relative to the total number of corpus lexical entries.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

Where TP = true positive, FN = false negative, FP = false positive, TN = true negative.

Table II shows the results of precision and recall method for five developed ontologies.

TABLE II

PRECISION AND RECALL METHOD RESULTS FOR FIVE ONTOLOGIES

	1	2	3	4	5
Precision	64%	41%	59%	72%	32%
Recall	84%	50%	72%	89%	41%

The next step was to find similarity in concept spelling. One set of concepts was compared to the set of corpus terms. The second step was performed by Levenshtein edit distance method. Edit distance – the minimum edit distance between two strings is the minimum number of editing operations needed to transform one into the other, like insertion, deletion or substitution.

The Levenshtein edit distance formula and examples can be found in [12].

TABLE III

LEVENSHTEIN EDIT DISTANCE FOR FIVE ONTOLOGIES

	1	2	3	4	5
Levenshtein edit distance	16	30	14	16	0

The results in Table II and Table III have shown that the fourth ontology is the most appropriate for a current task of raw material procurement.

B. Application-based Approach

Ontology in the multi-agent system provides the channel through which software agents interact; therefore, if the ontology definitions have not been sufficiently evaluated, communication between software agents may not succeed. The authors [7] propose evaluating ontologies by putting them into the application and evaluating the results. Good ontology is one, which produces good results on the given task. This approach has several drawbacks: 1) it can be seen that the ontology is good or bad when used in a particular way for a particular task, but it is difficult to generalize this observation; (2) the ontology could be only a small component of the application and its effect on the outcome may be relatively small and indirect; (3) comparing different ontologies is only possible if they can all be plugged into the same application.

Five ontologies were developed in Protégé ontology editor, then converted by Ontology Bean Generator in JADE classes; NetBeans IDE as an integrated development environment was executed in order to evaluate the produced results.

Five ontology-based agent systems have shown the following results:

1. The use of the first, second, third and fifth ontology is not enough for solving the problem.
2. The fourth ontology has shown the best evaluation results, cooperation between agents is achieved (see Fig. 1).

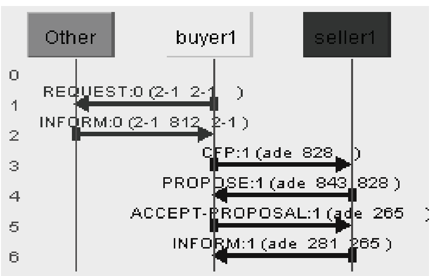

Fig. 1. Agent cooperation in JADE, using the fourth ontology.

C. Golden Standard Approach

The idea of golden standard approach is in comparison with concepts in the evaluated ontology to “golden standard”, which can be another ontology, considered a good representation of the concepts of the problem domain under consideration, or it could be taken from a corpus of documents, or prepared by a domain expert. Evaluation of the ontology can also be based on the precision and recall method, comparing ontology with a human-provided golden standard. In this context, precision is the fraction of the labels that also appear in the golden standard relative to the total number of labels. Recall is the percentage of the golden standard lexical entries that also appear as labels in the ontology, relative to the total number of golden standard lexical entries. Similarity between strings can be measured by the Levenshtein edit distance. The authors [5] propose several measures, such as the semantic cotopy of two hierarchies, for comparing structural aspect of two ontologies. Evaluation of an ontology on the semantic aspect can also be based on precision and recall measures [1]. The drawback of golden standard approach is the requirement for a lot of manual human work. However, once the golden standard is defined, comparison of two ontologies can proceed entirely automatically [3].

The fourth ontology was taken as a golden standard and was compared with others.

TABLE IV

PRECISION AND RECALL MEASURES FOR GOLDEN STANDARD APPROACH

	1	2	3	5
Precision	100%	68%	86%	41%
Recall	77%	68%	86%	32%

The results depicted in Table IV and Levenshtein edit distance have shown that the first ontology and the third ontology have better results than the second and the fifth ontology, having the fourth ontology as a golden standard.

D. Predefined Criterion Ontology Evaluation

This approach of ontology evaluation deals with the selection of a good ontology from a variety of ontologies as a decision-making problem. For the evaluation of ontology, several decision criteria or attributes must be defined and a numerical score for each criterion is given. An overall score for the ontology is then calculated as a weighted sum of its per-criterion scores. A drawback is that a lot of manual involvement by human experts may be needed.

Ontology evaluation can address a number of several different criteria. Therefore, the first task of the evaluator is to choose the criteria relevant for the given evaluation and then the proper evaluation methods to assess how well the ontology meets these criteria [1].

Different criteria and methods were analyzed by [1]. Five criteria have been chosen from this literature survey: accuracy, clarity, completeness, conciseness and consistency (see Fig. 2).

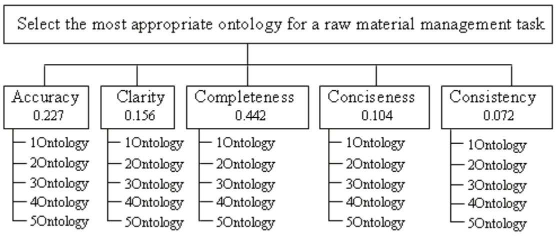

Fig. 2. A hierarchy tree with criteria calculated eigenvector.

- *Accuracy:* A higher accuracy comes from correct definitions and descriptions of classes, properties, and individuals.
- *Clarity*: Definitions should be objective and independent of the context. When a definition can be stated in logical axioms, it should be. All entities should be documented with a natural language.
- *Completeness*: Completeness measures if the domain of interest is appropriately covered. All the knowledge that is expected to be in the ontology is either explicitly stated or can be inferred from the ontology.
- *Conciseness*: Conciseness is the criteria that states if the ontology includes irrelevant elements with regard to the domain to be covered or redundant representations of the semantics.
- *Consistency*: Consistency describes that the ontology does not include or allow for any contradictions.

The authors [9] offer Ontometric – applying the Analytic Hierarchy Process (AHP) in the Ontology Choice. It can be used to: 1) select the most appropriate ontology among various alternatives or, 2) decide on the suitability of a particular ontology for the project.

Taking into account the general steps of AHP, described in [7], they have adapted the method to be used in the reuse of ontologies:

Step 1: specify the objective of the problem – "Select the most appropriate ontology for a raw materials management task";

Step 2: build a hierarchy tree in this way: the root node is the objective of the problem, the intermediate levels are the criteria, and the lowest level contains the alternatives.

Step 3: for each set of brother nodes, make the pairwise comparison matrices with the criteria of the decision tree. For each comparison matrix, an eigenvector must be calculated.

Step 4: for each alternative ontology, assess its characteristics. For each one of these characteristics, the engineer should establish a scale of appropriate ratings.

Step 5: lastly, combine the vectors of weights obtained in step 3 with the values of the alternatives. Finally, the suitable ontology is chosen based on the results obtained.

Please refer to [9] for a better understanding of AHP method and Ontometric method.

TABLE V

APPLIED AHP METHOD FOR SELECTING BEST ONTOLOGY

	Final Result
Ontology1	0.575
Ontology2	0.305
Ontology3	0.551
Ontology4	**0.852**
Ontology5	0.340

The results in Table V have shown that the fourth ontology is the most appropriate one.

V. RESULTS

The study of ontology evaluation techniques has shown that evaluation approaches can be categorized by evaluation means: data-driven, application-driven, by humans or comparing to golden standard, or based on the level of evaluation: lexical, hierarchical, syntactical etc. Five different ontologies were developed for the multi-agent system for raw materials management task:

1) The first ontology consists only of hierarchical relationships – taxonomy;
2) The second ontology with grammatical mistakes;
3) The third ontology is incomplete;
4) The fourth ontology is complete;
5) The fifth ontology refers to a chip manufacturing domain area.

These five ontologies were evaluated by four approaches, and the following results were achieved:

1) Data-driven approach showed that the most appropriate ontology for the task was the fourth one, because of its highest indicators of precision and recall measures comparing corpus terms with ontology concepts and because of Levenshtein edit distance measure in concept spelling similarity.

2) Application-based approach could show the performance of these five ontologies applied in a real multi-agent system. As a result – only the fourth ontology could be used in an agent system, because of its completeness, others – due to incompleteness and grammatical mistakes did not show good performance.

3)The fourth ontology was taken as the expert-provided golden standard and was compared with other four ontologies. Better results were shown by taxonomy and incomplete ontology.

4)Predefined criteria ontology evaluation had the following result – the fourth ontology was the most appropriate ontology using the five criteria in its evaluation by means of an analytical hierarchy process.

VI. CONCLUSION

This paper has proposed a study of ontology evaluation techniques. Practical experiments were based on the developed ontologies for a multi-agent system, solving raw materials management task.

The ontology evaluation results have shown that only complete, with high accuracy developed ontology can be applied in ontology-based systems for its performance; otherwise, ontology application cannot offer its full potential in an agent system.

REFERENCES

[1] D.Vrandečić, "Ontology Evaluation," PhD thesis, Karlsruher Institut für Technologie, 2010.
[2] D. Plinere, A. Borisov, "A Negotiation-Based Multi-Agent System for Supply Chain Management," *IT and Management Science*. vol. 49, pp.128–132, 2011.
[3] J.Brank, M.Grobelnik, D. Mladenic, "A survey of ontology evaluation techniques," *Conference on Data Mining and Data Warehouses (SiKDD 2005)*, 2005.
[4] J. Brank, D. Mladenic, M. Grobelnik, "Automatic Evaluation of Ontologies," In: Kao, A., Poteet, S. (*eds.*), *Text Mining and Natural Language Processing*, Springer, 2006.
[5] A. Meadche, S. Staab, "Measuring Similarity between Ontologies," *European Conference Knowledge Acquisition and Management (EKAW),* pp. 251–263, Oct. 1–4, 2002, Madrid, Spain.
[6] K. Dellschaft, S. Staab "On How to Perform a Gold Standard Based Evaluation of Ontology Learning," *In International Semantic Web Conference*, pp. 228–241, 2006.
[7] R. Porzel and R. Malaka, "A Task-Based Approach for Ontology Evaluation," *ECAI Workshop Ontology Learning and Population (OLP),* Aug. 2004, Valencia, Spain.
[8] C. Brewster, H. Alani, S. Dasmahapatra, and Y. Wilks, "Data Driven Ontology Evaluation," *International Conference on Language Resources and Evaluation* (LREC 2004), May 24–30, 2004, Lisbon, Portugal.
[9] A. Lozano-Tello, A. Gómez-Pérez, "Ontometric: A method to choose the appropriate ontology," *Journal of Database Management*, vol. 15, issue 2, pp. 1–18, 2004. http://dx.doi.org/10.4018/jdm.2004040101
[10] A. Gómez-Pérez, "Some ideas and examples to evaluate ontologies," Knowledge Systems Laboratory, Stanford University, 1994.
[11] A. Gómez-Pérez, "Towards a framework to verify knowledge sharing technology," *Expert Systems with Applications*, no. 4, pp. 519–529, 1996. http://dx.doi.org/10.1016/S0957-4174(96)00067-X
[12] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady* vol. 10, no. 8, 1966, pp. 707–710.

Darya Plinere received her *Mg. sc. ing* from Riga Technical University in 2006. She is a Researcher at the Faculty of Computer Science and Information Technology, Riga Technical University. Her research interests include multi-agent systems, ontology engineering and supply chain management.
E-mail: darja.plinere@rtu.lv.

Arkady Borisov received his Doctoral degree in Technical Cybernetics from Riga Polytechnic Institute in 1970 and *Dr. habil. sc. comp*. degree in Technical Cybernetics from Taganrog State Radio Engineering University in 1986. He is a Professor of Computer Science at the Faculty of Computer Science and Information Technology, Riga Technical University. His research interests include fuzzy sets, fuzzy logic, computational intelligence and bioinformatics. He has 235 publications in the field. He is a member of IFSA European Fuzzy System Working Group, Russian Fuzzy System and Soft Computing Association, honorary member of the Scientific Board, member of the Scientific Advisory Board of the Fuzzy Initiative Nordrhein-Westfalen (Dortmund, Germany).
Address: 1 Kalku Str., Riga, LV-1658; phone: +371 67089530
E-mail: arkadijs.borisovs@cs.rtu.lv.

Darja Plinere, Arkādijs Borisovs. Ontoloģiskā zināšanu modeļa novērtēšana
Šajā rakstā ir izpētīta ontoloģijas novērtēšana, uzbūvēta izejvielu pārvaldības uzdevumam, pielietojot multi-aģentu sistēmu. Ontoloģija aģentu sistēmas nodrošina ar vienādu izpratni par informāciju, ar kuru apmainās aģenti, risinot uzdevumu. Ontoloģija ir kanāls, caur kuru aģenti mijiedarbojas. Līdz ar to, sistēmas efektīgai darbībai ir nepieciešams novērtēt uzbūvētās ontoloģijas kvalitāti un tās piemērotību uzdevumam. Ontoloģiskā zināšanu modeļa novērtēšanas pētījuma rezultātā tika identificētas četras pieejas novērtēšanai – vairāku kritēriju, salīdzinājums ar zelta standartu, balstoties uz datiem un balstoties uz pielietojumu pieeju. Pieejas var arī klasificēt attiecībā uz novērtējuma līmeni – sintakses, hierarhijas, attiecības u.c. Lai veiktu pētījumu, tika uzbūvētas piecas ontoloģijas, un katra no tām tika pakļauta novērtēšanai četrās pieejās. Pirmajai ontoloģijai ir tikai hierarhiskās attiecības (taksonomija), otrā ir – ontoloģija ar gramatiskām kļūdām, trešā – neaptver visu priekšmeta apgabalu (nepabeigtā), ceturtā – aptver visu priekšmeta apgabalu un piektā – ir ontoloģija, kas saistīta ar ražošanu, nevis ar izejvielu pasūtījumu. Pieeja, kas balstīta uz datiem, salīdzināja pasūtījuma procesa dokumentu un uzbūvēto ontoloģiju. Tika izmantotas šādas metodes – precizitāte un pilnīgums, lai novērtētu priekšmeta apgabala segumu ar ontoloģiju, kā arī Levenšteina attālumu, lai noteiktu, cik lielā mērā ontoloģijas koncepti sakrīt ar dokumenta terminiem. Pieejā, kas balstīta uz pielietojumu, ontoloģija ir iekļauta aģenta sistēmā, un aģentu mijiedarbības rezultāti tika pētīti. Tikai ontoloģija, kas konstruēta ar augstu precizitāti un kas sedz visu priekšmeta apgabalu, var dot rezultātus no tās izmantošanas. Pēc tam tika izdarīta salīdzinošā analīze ar zelta standartu un noteikta daudz-kritēriju ontoloģijas kvalitāte. Salīdzinājums ar zelta standartu tika veikts ar tām pašām metodēm, kā pieejā, kas ir balstīta uz datiem. Daudz-kritēriju pieejā tika izmantots algoritms – analītiskais hierarhiskais process. Visas četras pieejas novērtēja ceturto ontoloģiju kā vispiemērotāko ontoloģiju šim uzdevumam.

Дарья Плинере, Аркадий Борисов. Оценка качества онтологической модели знаний
В данной статье исследуется оценка качества онтологии, построенной для решения задачи управления материальными ресурсами предприятия с помощью многоагентной системы. Онтология в агентной системе обеспечивает одинаковое понимание агентами информации, которой они обмениваются для решения поставленных задач. Онтология представляет собой канал, через который агенты взаимодействуют. Следовательно, для функционирования системы необходимо оценить качество предлагаемой онтологии и ее пригодность к данной задаче. В результате исследования оценки качества онтологической модели знаний были выявлены четыре подхода оценки качества – многокритериальный, сравнение с золотым стандартом, основанный на данных и основанный на применении. Подходы также можно классифицировать относительно уровня оценки – синтаксис, иерархия, отношения и т.п. Для проведения исследования было построено пять онтологий, каждая из них подвергалась оценке четырьмя подходами. Первая онтология имеет только иерархические отношения (таксономия), вторая – онтология с грамматическими ошибками, третья – не покрывающая предметную область (неполная), четвертая – покрывающая всю предметную область, и пятая – онтология, относящаяся к производству, а не закупке сырья. Подход, основанный на данных, сравнил документ, описывающий процесс закупки сырья, и построенные онтологии. Использовались методы точности и полноты для оценки покрытия онтологией предметной области, а также расстояние Левенштейна для определения того, насколько концепты онтологии совпадают с терминами документа. В подходе, основанном на применении, онтологии были включены в агентную систему, и исследовались результаты переговоров между агентами. Только онтология, построенная с высокой точностью и покрывающая предметную область, способна дать результат от ее применения. Затем был произведен сравнительный анализ с золотым стандартом и проведена оценка качества онтологии многокритериальным способом принятия решения. Сравнение с золотым сечением было произведено теми же методами, что и в подходе, основанном на данных. В многокритериальном подходе использовался алгоритм – аналитический иерархический процесс. Во всех результатах четвертая онтология была оценена как наиболее подходящая для поставленной задачи.