

HOW TO IMPROVE EXPRESSING DOMAIN SEMANTICS IN ONTOLOGIES

LOVRENCIC, S.

Abstract: *New ontology engineers have difficulties in expressing large part of semantics of the domain they try to describe. In previous research those difficulties were identified according to a set of proposed criteria for assessment of domain semantics captured in ontology, so-called semantic criteria. This paper describes adjustments in teaching process that have been introduced according to proposals based on difficulties identified in preliminary research. It also presents improvements in ontologies developed after those adjustments have been implemented. The need for more detailed results comparison influenced on proposal of additional semantic criteria and results obtained from their inclusion into preliminary and additional research are also presented.*

Key words: *ontology development; domain semantics; semantic criteria; ontology evaluation*



Authors' data: Assist.Prof. Dr.Sc. **Lovrencic**, S[andra]; University of Zagreb, Faculty of Organization and Informatics, Pavlinska 2, HR-42000, Varazdin, Croatia, sandra.lovrencic@foi.hr

This Publication has to be referred as: Lovrencic, S[andra] (2011). How to Improve Expressing Domain Semantics in Ontologies, Chapter 25 in DAAAM International Scientific Book 2011, 25. 309-316, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-901509-84-1, ISSN 1726-9687, Vienna, Austria DOI: 10.2507/daaam.scibook.2011.25

1. Introduction

Ontologies nowadays, as modern knowledge bases, serve as a knowledge representation technology that has been used to describe large number of domains. Since their first definition as “an explicit specification of a conceptualization” almost two decades ago (Gruber, 1993), they are now recognized as an ideal integral part of Semantic Web technologies (Hitzler et al., 2009). Many ontology development methodologies have been proposed as well as various languages and tools for their practical use (Gómez Pérez et al., 2004).

Important part of every development process is evaluation. Various evaluation methods of ontology taxonomy and content were developed (Gómez Pérez, 2004), some independent and some considering features of languages and tools used as evaluation factors (Sure et al., 2004). Several other evaluation methods suggest inclusion of users as a control point in the formal evaluation process (Noy et al., 2005) and combination of several evaluation methods based on evaluation tasks (Lovrencic & Cubrilo, 2008).

Many ontology development tutorials for specific languages and tools and according to specific methodologies were being proposed. But, identification of potential difficulties that new ontology engineers can come across when trying to represent knowledge of a certain domain is usually not taken into consideration during development process. Similarly, as programmers accustomed to procedural programming have problems when introduced to declarative programming, if new ontology engineers are accustomed to other knowledge representation methods, they may not describe domain semantics well or not use some representation possibilities of ontologies appropriately. In previous work (Lovrencic & Malekovic, 2010) some of those ontology development problems were already identified. This paper describes next step of this research and shows results obtained after adjustments in teaching process. Additional semantic criteria are also proposed.

The paper is organized as follows: in second section preliminary research from previous work and its results are briefly described; afterwards, research process and results of teaching process adjustments are provided; analysis and suggestions for new semantic criteria are given; conclusion and future work are in the final section.

2. Preliminary Research

Ontology engineering is regular part of modern information systems curricula. At Faculty of organization and informatics in Varaždin, Croatia, it is taught at two levels:

- Undergraduate students learn simple ontology development with frame systems that are used for “describing hierarchies of classes with slots” (Gruber, 1993), as one of several knowledge representation methods;
- Graduate students learn ontology engineering as a part of Semantic Web efforts, using modern languages (OWL) and formal logic (description logics).

Since “the most important aspect of the ontological representation is its capacity of expressing domain semantics” (d'Amato et al., 2008), a preliminary research had

been conducted among undergraduate students (spring semester of year 2009/2010) with the purpose to detect whether they describe chosen domains with ontologies using their full potential. They had no prior experience with ontologies, but were familiar with knowledge representation methods for knowledge management in general. This research had been driven by observations that students with no prior experience in ontology development often create too small hierarchies, use very few attributes for class description or mix classes with attributes and instances.

For research purpose, several so-called semantic criteria (criteria for assessment of domain semantics captured in ontology) were taken into consideration (Lovrencic & Malekovic, 2010):

- Total number of hierarchies – main hierarchy or hierarchies (directly describe chosen domain) plus support hierarchy or hierarchies, if any (help to better describe domain and should be connected to main hierarchies with attributes);
- Number of support hierarchies;
- Depth of main and support (where applicable) hierarchies – suggests level of detail when describing domain concepts;
- Total number of classes, number of classes in main and support hierarchies (where applicable) – combined with depth suggests how well are domain concepts described;
- Total number of attributes – as they are class descriptors, their number shows how much of domain semantics is expressed;
- Number of attributes in main and support (where applicable) hierarchies;
- Number of connecting attributes – they connect classes (and hierarchies) together (primarily meaning that the attribute value of one class is the instance of the other) and ensure expression of more domain semantics;
- Number of simple and complex attributes – simple: any, boolean, float, integer and string; complex: class, instance and symbol.

Ontology development learning part of laboratory exercises consisted of four sessions that included lessons about ontology development step by step through an example in the Protégé-Frames tool (Stanford Center for Biomedical Informatics Research, 2011), development of ontology hierarchies for a chosen domain, creation of appropriate class attributes as well as population of ontology with instances (Lovrencic & Malekovic, 2010).

50 randomly chosen ontologies were analysed according to aforementioned semantic criteria using several statistical measures: arithmetic mean, median, mode, standard deviation and skewness. Results showed that:

- about 60% of users did not use support hierarchies in representation of semantic information;
- created support hierarchies had mostly only one class and few simple attributes;
- very small number of complex attributes was used.

Detailed information about preliminary research and obtained results can be found in previous work (Lovrencic & Malekovic, 2010). After the research several

suggestions had been made for further work, one of them being adjustment in the teaching process.

3. Improvements in Expressing Domain Semantics

3.1 Teaching Process Adjustments

According to results of preliminary research, adjustments in teaching process were made during laboratory exercises in spring semester of year 2010/2011:

- Session 1 – Several examples were explained to student groups (and individually) with more emphasis on difference between main and support hierarchy and on complex attributes for connecting classes.
- Session 2 – Hierarchy part of examples from first session was explained again to all students at the beginning of the session. Appropriateness of students chosen domains for ontology development was discussed (with changes where needed – all domains had to be different). Main session task was to develop the hierarchy (or hierarchies), but no direct suggestions for support hierarchies were made. Instead, several general possibilities to better express domain semantics were discussed with each student.
- Session 3 – Attributes part of examples from first session was explained again to all students at the beginning of the session. Main session task was to create class attributes. Various possibilities for complex attributes and connection of classes were discussed with each student according to their specific domain.
- Session 4 – Students created some parts of ontology at home (frames for instance entry window, population of ontology with instances, several queries, ontology visualization). They completed ontologies and presented their work.

3.2 Comparison of Results

From 50 randomly chosen ontologies data were analysed according to semantic criteria using the same statistical measures as in preliminary research.

Semantic criteria	Values											
	0		1		2		3		4		5	
	A	B	A	B	A	B	A	B	A	B	A	B
Number of hierarchies	-	-	30	16	14	15	1	7	5	7	0	5
Number of main hierarchies	-	-	49	50	1	0	0	0	0	0	0	0
Number of support hierarchies	30	16	14	15	2	7	4	7	0	5	0	0
Depth of main hierarchy	2	5	5	13	21	16	16	10	6	6	0	0
Depth of support hierarchy	14	15	4	18	2	1	0	0	0	0	0	0

Tab. 1. Comparison of Selected Semantic Criteria Values

Data in Table 1 under columns **A** are taken from preliminary research and data under columns **B** are obtained from a research after teaching process adjustment. It can be seen that number of ontologies with support hierarchies was increased from 20 (40%) to 34 (68%). Students ontologies had one main hierarchy, but a number of

support hierarchies greater than one was increased from 6 (30% of 20) to 21 (61,76% of 34). Main hierarchy depth of 2 is increased, but an average depth is actually decreased from 2,38 to 1,98 (-16,8%), unlike the depth of support hierarchies that has an average increase of 90% (from 0,31 to 0,59). Large differences in results show that teaching process adjustments positively influenced on the use of support hierarchies and their better description in the ontology development process. Slight decrease in depth of main hierarchy might be due to normal deviation or because of more focus on description of support hierarchies.

Table 2 shows results comparison of several statistical measures according to semantic criteria related to classes and hierarchies. There were no cases of ontologies that have very large number of classes (according to average) after teaching process adjustments. Standard deviation and skewness confirm that distribution of number of hierarchies in those ontologies was more uniform. When ontologies were graded it was noted that there was smaller number of cases where students mistakenly represented instances or attributes as classes, which can explain less classes per ontology. Skewness results also show that asymmetry in distribution was smaller both in number of hierarchies and their depth. Usage of support hierarchies and a number of classes in them was increased, but standard deviation and skewness were not changed, meaning that the asymmetry was the same (although not large).

Semantic criteria		No of hierarch.	No of support hierarch.	Depth of main hierarch.	Depth of support hierarch.	No of classes	No of classes in main hierarch.	No of classes in support hierarch.	
Raw values	Min	A	1	0	0	6	1	1	
		B	1	0	0	5	1	1	
	Max	A	4	3	4	110	105	12	
		B	5	4	4	56	48	17	
Statistical measures	Av.	A	1,62	0,6	2,38	0,31	21,76	22,05	3,23
		B	2,40	1,4	1,98	0,5882	16,18	9,91	4,9412
	Me d.	A	1	0	2	0	16,5	16	2
		B	2	1	2	1	14,5	7,5	4
	Mo de	A	1	0	2	0	14	16	1
		B	1	0	2	1	11	1	1
	St. dev.	A	0,9452	0,9035	0,9666	0,6806	17,8309	22,9858	3,4296
		B	1,3401	1,3401	1,1691	0,5569	8,310	9,4528	3,7331
	Ske w.	A	1,6023	1,5910	-0,2809	1,5139	3,1115	2,6314	1,3704
		B	0,6466	0,6466	0,1199	0,1896	1,9788	2,3442	1,3524

Tab. 2. Hierarchy Analyses Comparison

Comparison of attribute analyses is presented in Table 3. Number of attributes was generally reduced after teaching process adjustments, but a number of those used to describe classes in support hierarchies was increased. According to standard

deviation and skewness results, attribute distribution was more even, except in the case of complex attributes (and connecting attributes, but most of complex attributes are also connecting). Students created more support hierarchies and more connecting attributes for them, with slightly increased asymmetry. Greater average in number of attributes in support hierarchies in relation to average number of connecting attributes means that not almost all attributes in support hierarchies are those used to connect them to main hierarchies (which was the case in preliminary research).

		Semantic criteria	No of attributes	No of attr. in main hierarchies	No of attr. in support hierarchies	No of connecting attributes	No of simple attributes	No of complex attributes
Raw values	Min	A	4	2	1	1	2	0
		B	5	4	1	0	3	0
	Max	A	60	54	17	8	52	10
		B	40	22	25	10	37	13
Statistical measures	Av.	A	12,18	10,13	5,38	3,4	9,4	2,78
		B	15,42	9,65	7,68	3,00	12,14	2,86
	Med.	A	9	7,5	4	2	7,5	2
		B	13,5	8	6	2	10	2
	Mode	A	8	8	2	2	8	0
		B	8	6	3	1	8	1
	St. dev.	A	9,1377	11,2339	3,8580	2,4902	8,2293	2,7575
		B	7,9336	5,0144	6,0288	2,7080	6,9371	2,8927
	Skewness	A	3,3203	3,4564	1,9580	1,0398	3,4407	0,9809
		B	0,8333	0,9157	1,0591	1,0735	0,5972	1,3467

Tab. 3. Attribute Analyses Comparison

4. New Semantic Criteria Inclusion

Results described above can not be completely interpreted due to significant changes in number of support hierarchies and maximal number of classes and attributes. It can not be concluded with certainty whether ontologies developed after teaching process adjustments did express more of domain semantics. Therefore, several new semantic criteria were included into research:

- Number of attributes per class;
- Number of attributes in main hierarchies per class in main hierarchies;
- Number of attributes in support hierarchies per class in support hierarchies.

New semantic criteria are compared according to average, standard deviation and skewness, because those measures proved in preliminary research to be the most important ones. Results presented in Table 4 show that after teaching process adjustments there was in average a greater number of attributes per class according to all three semantic criteria, although there is the smallest difference in support hierarchies. The same interpretation can be drawn from skewness – asymmetry is

smaller in distribution of attributes per class in general and in main hierarchies whereas is almost the same in support hierarchies. Therefore, it can be concluded that after teaching process adjustments there was more of domain semantics expressed in ontologies developed.

Semanti c criteria		Number of attributes per class	Number of attributes in main hierarchies per class in main hierarchies	Number of attributes in support hierarchies per class in support hierarchies	
Raw values	Mi n	A	0,1052	0,0571	0,3637
		B	0,2083	0,1042	0,1429
	Ma x	A	4,1667	13	5
		B	4	9	6
Statistical measures	Av.	A	0,7571	1,4727	1,7806
		B	1,2522	1,9138	1,8945
	St. de v.	A	0,6819	2,8940	1,2185
		B	0,9151	1,8818	1,1998
	Sk ew.	A	3,0722	3,6837	1,2966
		B	1,0036	2,1037	1,3124

Tab. 4. New Semantic Criteria Analyses Comparison

5. Conclusion and Further Research

According to presented results, new ontology engineers had significant improvements in expressing domain semantics with ontologies. They were more aware of importance of support hierarchies (70% increase), of importance of both simple and complex attributes for class description and of positive influence of the usage of connecting attributes on expressing domain semantics. In general, standard deviation and skewness results were lesser in research results after teaching process adjustments, meaning that results interpretation is more accurate.

Although presented research improved insights into difficulties of new ontology developers and better showed improvements after teaching process adjustments, it was restricted only to one university and its curricula. The opportunity to conduct the same research on other universities and to compare results would provide broader view on this subject and analysis of best teaching practices. In this way semantic criteria adequacy can be tested more comprehensively, as well as compared to other ontology evaluation methods, not specifically used for analysis of ontologies developed by new engineers.

Conduction of the research on frame-based ontology development process and not on more and more popular Semantic Web technologies is another restriction. Keeping this in mind, results showed general difficulties and guidelines for their improvements, but next step must include ontology development using OWL language and description logics. Detailed analysis of semantic criteria suitability and their possible adjustments for this knowledge representation formalism will be carefully carried out prior to research.

Since ontologies are acknowledged knowledge representation technology and integral part of Semantic Web efforts, improvements in ontology engineers' capabilities to express domain semantics are necessary. This can be achieved with adjustments in ontology development tutorials and with establishment of appropriate set of semantic criteria for ontology evaluation. The aim of work described above is to give contribution in this field of research.

6. References

- d'Amato, C.; Staab, S. & Fanizzi, N. (2008). On the influence of Description Logics Ontologies on Conceptual Similarity, *Knowledge Engineering: Practice and Patterns: Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2008)*, Gangemi, A. & Euzenat, J. (Eds.), pp. 48-63., ISBN 978-3-540-87695-3, Acitrezza, Italy, September– October 2008, Springer, Berlin
- Gómez Pérez, A. (2004). Ontology Evaluation, In: *Handbook on Ontologies*, Staab, S. & Studer, R., (Eds.), pp. 251-273, Springer-Verlag, ISBN 3-540-40834-7, Berlin, Germany
- Gómez Pérez, A.; Fernández-López, V. & Corcho, O. (2004). *Ontological Engineering*, Springer-Verlag, ISBN 1-85233-551-3, London, UK
- Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications, *Knowledge Acquisition*, Vol. 5, No. 2, (June 1993), pp. 199-220, ISSN 1042-8143
- Hitzler, P.; Krötzsch, M. & Rudolph, S. (2009). *Foundations of Semantic Web Technologies*, CRC Press, ISBN 978-1-4200-9050-5, Boca Raton
- Lovrencic, S. & Cubrilo, M. (2008). Ontology Evaluation – Comprising Verification and Validation, *Proceedings of the 19th Central European Conference on Information and Intelligent Systems (CECIIS 2008)*, Aurer, B., Baca, M. & Rabuzin, K. (Eds.), pp. 657-663, ISBN 978-953-6071-04-3, Varaždin, Croatia, September 2008, FOI, Varaždin
- Lovrencic, S. & Malekovic, M. (2010). Identifying Beginner Problems in Expressing Domain Semantics when Developing Ontologies, *Proceedings of The Fourth International Conference on Advances in Semantic Processing (SEMAPRO 2010)*, Popescu, M. & Stewart, D.L. (Eds.), pp. 31-36, ISBN 978-1-61208-000-0, Florence, Italy, October 2010, IARIA, Wilmington
- Noy, N.F.; Guha, R. & Musen, M.A. (2005). User ratings of ontologies: Who will rate the raters?, *Available from:* <http://www.stanford.edu/~natalya/papers/SS505NoyN.pdf> *Accessed:* 2011-06-03
- Sure, Y.; Gómez Pérez, A.; Daelemans, W.; Reinberger, M.-L.; Guarino, N.; Noy, N.F. (2004). Why Evaluate Ontology Technologies? Because it Works!. *IEEE Intelligent Systems*, Vol. 19, No. 4, (July-August 2004), pp. 74-81, ISSN 1541-1672
- Stanford Center for Biomedical Informatics Research (2011). what is protégé-frames?, *Available from:* <http://protege.stanford.edu/overview/protege-frames.html> *Accessed:* 2011-03-15