ENHANCED METHODOLOGY FOR ONTOLOGY DEVELOPMENT

Petr RAPANT[1], Daniela SZTURCOVA[2], Lucie HRUBA[3]

[1]Institute of Geoinformatics, Faculty of Mining and Geology, VSB – Technical University of Ostrava, 17. listopadu 15, 708 33, Ostrava – Poruba, Czech Republic,
petr.rapant@vsb.cz

[2]Institute of Geoinformatics, Faculty of Mining and Geology, VSB – Technical University of Ostrava, 17. listopadu 15, 708 33, Ostrava – Poruba, Czech Republic,
daniela.durakova@vsb.cz

[3]Department of Economy and Marketing, Regional Development Agency Ostrava, Na Jizdarne 7, 702 00, Ostrava, Czech Republic
hruba@arr.cz

Abstract

The presented paper deals with building of an initial glossary of terms, which is a preparatory phase of domain ontology building. Existing methodologies suppose to develop the initial glossary of terms by an analysis of existing documents, or using expert knowledge. There is no structured process defined for this step of the ontology building. We propose to embrace experience from object oriented analysis: to describe structure, behaviour and rules valid for a domain of interest. These three descriptions represent complex and systematic view to the domain and ensure that created glossary of terms will be exhaustive. We demonstrate our method on the domain of road transport. The method is developed as an extension of well known METHONTOLOGY, but it is general enough to be acceptable for other methodologies for the ontology building too.

Keywords: ontology design, ontology design methodology, transportation

INTRODUCTION

The research project Logic and Artificial Intelligence for Multi-agent Systems (MASS) project was carried at the Laboratory of Intelligent Systems, VSB – Technical University of Ostrava, Czech Republic, between 2004 – 2008. One of the objectives of the project was to find a solution for the communication between agents under the situated Multi-agent system (MASS), i.e. the system, in which individual agents are situated in space and spatial processes are implemented. Road transport was chosen as a test domain. The individual vehicles were employed as the MASS agents. In addressing the communication of these agents, it has proven essential to use a set of well-defined concepts with clear semantics and syntax, allowing description of both the road infrastructure itself as well as the movement of vehicles on it, including the rules which are followed in this movement. It meant to use an appropriate ontology, which describes well the domain studied and also allows clear-cut communication of the agents using clear semantics of the concepts.

DEVELOPMENT OF METHODOLOGY FOR ONTOLOGY DESIGN

Development of ontology methodologies has been addressed by many research teams. They made attempts to define the sequences of steps leading to the formation of ontology. A very good overview of different methodologies is provided at e.g. (1) work.

One of the first methodologies was presented in 1995. It was based on the experience drawn from the Enterprise Ontology and TOVE (TOronto Virtual Enterprise) draft (2), which was obtained in the field of enterprise modelling. One year later, the KACTUS project was presented at the European Conference on Artificial Intelligence, which focused on developing the ontology for the field of electric power mains (3). In the context of this project the methodologies for ontology development were also presented. The methodology used to design an ontology proposal within the SENSUS project (4) is also now firmly established. From our perspective, we consider the METHONTOLOGY to be the most interesting and well organized (5). It is dealt with in detail below.

The common feature of most of the above noted methodologies used for the formation of ontologies is that they do not systematically describe all stages of ontology development. What still needs to be processed is often the creation of the glossary of terms. However, this stage is considered, for the formation of ontology, crucial.

METHONTOLOGY METHODOLOGY

As stated above, we sought an appropriate methodology, which would allow for the systematic ontology development. After examining the various present methodologies, we observed that none of them meet our requirements in full. Therefore, we eventually decided to use the METHONTOLOGY to start with, and to develop its enhancement for the methods used for the *specification phase*, which will result in the building of the *glossary of terms*.

We have demonstrated the basic ontological types that form the content of ontology. It's a case of attributes, taxonomies, definitions of concepts, etc. METHONTOLOGY comes from these basic ontological types. METHONTOLOGY was developed in the Laboratory of Artificial Intelligence at the Polytechnic University of Madrid (5). This is largely based on IEEE Std 1074-1995. We decided to follow the METHONTOLOGY because of its transparent logical structure and integrity of its steps, which reflect the process of ontology development. This includes the following tasks:

- ontology specification,
- build glossary of terms,
- build concept taxonomies,
- build relation diagrams,
- build concept dictionary,
- describe:
  - relations,
  - attributes,
  - constants,
- describe:
  - formal axioms,
  - rules,
  - instances.

The METHONTOLOGY thus describes the sequence of steps that will, in its final stage, bring us the basic ontological types. Like most of these current methodologies, this one is no different in providing no detailed guidance on how to achieve adequacy of real ontological concepts, i.e. how to select appropriate concepts with regard to the given domain and its tasks. It only contains a brief description of how to proceed in the specification phase, i.e. during the *ontology specification*.

SPECIFICATION OF ONTOLOGY

The objective of this stage is to create a specification document written in natural language, which should include at least the following information (5):

- the purpose of ontology, including the intended use, use case scenarios, specifications of end users, etc.
- the level of formality of the implemented ontology,
- the range, including a set of concepts to be represented in the ontology, its characteristics and granularity.

It is recommended to use a middle-out approach, which allows obtain a set of terms that should be included in the ontology, without us knowing of their significance at that time. It is also recommended to group the terms into the classification trees. All this allows us not only to verify the relevance of ontology terms and seek out missing terms at an early stage of development, but also search for synonyms, or even omit the terms that are redundant. Another advantage of this approach is that it allows you to search for terms that should be the core of future ontology. Consequently, they can be generalized or, on the contrary, specialized, if needed and only to the necessary extent. The resulting set of terms is much more stable, and will require far fewer changes in the future (5).

METHONTOLOGY also recommends a whole range of techniques, how to acquire the knowledge from the area of the subject, such as different interviews, document analyses, etc. Based on this knowledge a glossary of terms is to be built later. However, this methodology provides no structured process to obtain a glossary of terms. Basically, an intuitive approach is herein assumed. We have attempted to resolve this issue.

PROPOSED ENHANCEMENT OF METHONTOLOGY METHODOLOGY

Building a basic glossary of terms, which so far has not taken into account the various integrity constraints, etc., must be preceded by a number of steps. As shown in our research, the steps leading to the creation of the glossary may be (at least in general) specified and systematised. Our proposal is aimed at such a specification. We are going to describe the basic principles, whose compliance will improve the relevance of the final ontology. We will, in particular, focus on the creation of ontology for the purposes of artificial intelligence and knowledge-based systems.

At the beginning of ontology development it is necessary to specify *the purpose of ontology*. The next step is to find and collect the terms used to describe a given domain. When taking this step, we are trying to maintain the shareability of the created ontology. Identified terms are evaluated with regard to the clarified *purpose* of our ontology. Subsequently, based on the selected terms, there will be the *glossary of terms* generated, which will include a simple list of terms from the given domain. Needless to say, the process preceding the creation of the glossary is iterative.

It's a good idea to realize at the beginning of the ontology development, that the domain analyzed to meet the needs of future knowledge-based system has its static and dynamic aspects, as well as behaviour rules. Static aspects of the system include the concepts of typical objects in the given domain, i.e. they describe *its static structure*. Dynamic aspects include a description of *the system behaviour* which must respect certain *rules of behaviour*.

The analysis of systems from different perspectives is dealt with inter alia using object-oriented analysis. There are a number of different methodologies to carry out such an analysis. However, those have some general features in common. They always attempt to describe the following elements:

- structure of the system
- behaviour of the system
- rules of the system behaviour.

Our proposal is based on the same approach. We have, therefore, designed the following sequence of steps that lead to the creation of the quality glossary of terms:

1) To search the initial set of terms there is a good idea to start with building up three descriptions of the analyzed domain:
    a. description of the structure
    b. description of the behaviour
    c. description of the behaviour rules

   Description of the structure is relatively simple, and should contain all the elements of the analyzed domain, which will affect or rather will implement behaviour of the domain.

Description of the behaviour may be more complicated. It should include a description of specific issues and tasks resolved within the domain, for example in the form of use case scenarios.

Description of the behaviour rules should involve all the rules that must be respected by the system elements in the task implementation resolved within the domain.

2) The next step is focused on extracting the set of terms from each of the descriptions (both, single and multi-word terms; it will be mostly the nouns and the verbs) shown in the descriptions and bearing the information related to the domain. This way we will obtain three independent sets of terms that do not entirely agree with each other at first glance. For further processing it is important to arrange them in alphabetical order. We will maintain any duplicity between the lists in this step.

3) Consequently, we will make a comparison of these three sets of terms, such as using a table with three columns. Each column will show one list. The entire table will be then arranged alphabetically, so that each line would contain the identical terms used in multiple description. If there is no identical term for a given term in another description (i.e. in another column), we will leave this column line blank. The final table shows that if we selected only one description for the terms used, the result would never be sufficiently representative.

4) We will proceed to search the individual basic terms which form the multiple word terms, and group them into individual hierarchies. If a multiple word term is considered unique, it may be introduced as one term (e.g. *the intended direction of continuation after the intersection*). Further on, the terms that are not included in the original description, but logically belong on the list are added. The rules of addition are based on analogy (e.g. the term *to the right* can be supplemented with the term *to the left*).

5) Eventually, we will search the synonyms and homonyms and group them together. The synonyms are given in succession on one line, separated by commas. In the first place, we will enlist the term that is preferred. Homonyms are mentioned on separate lines.

CASE STUDY

Road transport system was chosen for our case study. We have followed proposed sequence of steps. It is easy to demonstrate that our approach leads to the creation of the quality glossary of terms.

We have developed three different descriptions of transport systems. As we can see in Tab. 1, anyone of the three descriptions (if they are used separately) leads to very different set of terms, so it could lead to different ontologies, which would never be sufficiently representative. But in the case of contemporaneous use of all three descriptions much comprehensive set of terms can be generated. So combined use of all these descriptions generates synergetic effect.

Table 1: Comparison of the sets of terms.

Set of terms describing the static side	Set of terms describing the dynamic side	Set of terms describing the rules
	appropriate distance	
		at most one
		before
car	car	car
	catch up	
	continue on driving	
	cross	
	cross into the left lane	
	cross safely	
	determine	
direction of travel		direction of travel
direction of turning		
	drive	drive

Set of terms describing the static side	Set of terms describing the dynamic side	Set of terms describing the rules
dividing and/or merging section		
	endanger	
		enforced
		exceed
	give way	give way
	increase speed	
	indicate intention to cross the road	
	intended direction of continuation after an intersection	intended direction of continuation after an intersection
	intention	
intersection		intersection
junction		
		just one
	left lane	
	limit	
	main road	
		maximum car speed-limit currently permitted
		maximum speed limit
	merge into the right lane	
	minor road	
		motion
		move into the traffic lane
		move
		must not
one-way		
		only on
	overtake	
	pavement	
pedestrian crossing	pedestrian crossing	pedestrian crossing
pedestrian	pedestrian	pedestrian
place		place
	reduce speed	
		respect
		right-most traffic lane
road infrastructure		
road section		
	safe distance	
	section	
	side lane	
	slower moving car	
		specified
	speed	
	stop vehicle	
	straight section	
	"T"-shaped intersection	
traffic lane		traffic lane
	turn blinker off	try
	turn on a left blinker	
	turn on a right blinker	
	turn right	
	turn safely	
	turn	
		turning

Next step is searching for basic terms. We search for the individual basic terms, which serve to form the multiple-word terms. If there is a unique multiple-word collocation, it can be mentioned as one term (e.g. *the intended direction of continuation after the intersection*).

Furthermore, the terms not included in the original description, but logically belonging on the list are added. Those are added based on the analogy rule (e.g. the *„to the right“* term is to be complemented with the *„to the left“* term).

Now we identify synonyms and homonyms. Identified synonyms and homonyms are grouped together; other terms are also included in the table. In the last step, synonyms are given in succession on one line, separated by commas (see e.g. "*car*" / "*vehicle*"). The first term noted is the one preferred. Homonyms are given on separate lines (see e.g. "*turn*" / "*turn*"). The other terms are given on separate lines (see e.g. "*be located*" and "*place*").

Table 2. Grouping of synonyms and homonyms (shortened).

Term			
be located			
car, *vehicle*			
cross			
	- safely		
place			
speed	- maximum	- permitted	- road - car
	- reduce, *slow down* - increase, *speed up*		
turn			
	- safely - to the right - to the left		
turn			
	- blinker	- on - off	
...			

The result is a glossary of terms (see Table 2), which may already serve as a sufficiently representative input into METHONTOLOGY.

CONCLUSION

Currently, there are a number of methodologies aimed at the development of ontologies. Their common disadvantage is that they do not deal with the systematic generation of the initial glossary of terms. This paper suggests the ways to extend the METHONTOLOGY methodology so that there is a clearly defined set of steps at the beginning of the ontology development, resulting in obtaining a representative set of terms which then may be already processed by traditional methods of the METHONTOLOGY methodology.

We demonstrated the usability of our proposed method on the domain of road transport. The resulting glossary of terms is more complex, systematic and therefore more exhaustive than to search the terms in professional documents or to address experts.

In addition, from the proposed set of steps, it is evident that proposed approach is very general and thus applicable within other methodologies, which, at their beginning, result from the glossary of terms.

ACKNOWLEDGEMENTS

This research has been supported by the program "Information Society" of the Czech Academy of Sciences, project No. 1ET101940420 "Logic and Artificial Intelligence for multi-agent systems."

REFERENCES

1. Gómez-Pérez, A., Fernández-López, M., Corcho, O. (2004) Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. 2nd printing, Springer-Verlag, London. ISBN 1-85233-551-3.
2. TOVE Ontology Project. Enterprise Integration Laboratory, University of Toronto. http://www.eil.utoronto.ca/enterprise-modelling/tove/, 10.3.2010.
3. Wielinga, B., Schreiber, A. Th., Jansweijer, W., Anjewierden, A., van Harmelen, F. (1994) Framework and formalism for expressing ontologies. ESPRIT Project 8145 KACTUS, Free University of Amsterdam.
4. Ontology Creation and Use: SENSUS. http://www.isi.edu/natural-language/resources/sensus.html (10.3.2010)
5. Fernández, M., Gómez-Pérez, A., Juristo, N. (1997) METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Spring Symposium Series, Stanford, pp. 33-40.