Methodology of Ontology Building

Martina Číhalová[1] , Daniela Ďuráková[2], Lucie Hrubá[3], Petr Rapant[4]

[1] Katedra informatiky, Fakulta elektrotechniky a informatiky, Vysoká škola báňská–Technická univerzita Ostrava, 17. Listopadu 15,
708 33, Ostrava – Poruba, Česká republika
m.tina.cihal@gmail.com
[2] Katedra informatiky, Fakulta elektrotechniky a informatiky, Vysoká škola báňská–Technická univerzita Ostrava, 17. Listopadu 15,
708 33, Ostrava – Poruba, Česká republika
daniela.durakova@gmail.com
[3] Institut geoinformatiky, Fakulta hornicko-geologická, Vysoká škola báňská – Technická univerzita Ostrava, 17. Listopadu 15,
708 33, Ostrava – Poruba, Česká republika
lucie.hruba.st1@vsb.cz
[4] Institut geoinformatiky, Fakulta hornicko-geologická, Vysoká škola báňská – Technická univerzita Ostrava, 17. Listopadu 15,
708 33, Ostrava – Poruba, Česká republika
petr.rapant@vsb.cz

Abstrakt. Mnoho výzkumných skupin se zabývalo či zabývá právě problematikou ontologií a snaží se stanovit sérii kroků vedoucích k tvorbě ontologie. Většina z nich však nepopisuje všechny fáze tvorby ontologie, a to zejména fázi předcházející tvorbě glosáře pojmů, kterou vidíme pro budování ontologie jako zásadní. Hlavním cílem článku je snaha podat právě popis této fáze, od které jiné metodiky teprve začínají. Vycházíme ze základních ontologických typů obsahu ontologie a snažíme se ukázat, jakým způsobem vést úvahy při budování ontologie, aby její obsah tvořily opravdu adekvátní pojmy. Náš příspěvek se nezaměřuje na implementaci, ale především na konceptuální a logické aspekty ontologií.

Klíčová slova: Ontologie, metodika, Methontology.

Abstract. Many research groups have been dealing especially with the problems of ontology building and trying to set a series of steps leading to the formation of ontology. Most of them however do not describe all stages of ontology formation, and especially the stage which precedes formation of glossary of terms, which we see as principal for ontology building. The main purpose of this paper is an effort to present description especially of this stage from which other methodologies only then start from. Our starting point is the classification of the basic ontological types and we try to present how to think during ontology building so that its content would be formed by adequate terms. Our contribution is not aimed at implementation but principally at conceptual and logical aspects of ontologies.

Keywords: Ontology, methodology, Methontology.

1 Introduction

Problems related to the methodology of ontology building are an independent significant area of ontological engineering. The purpose of this paper is to point out some of the insufficiencies of present methodologies and to show their convenient extension. This extension lies in the proposing of at least a general approach how to secure adequacy of factual content of ontology being built. In the first chapter we are engaged in detail in ontology content as we can show that its structure is directly related to methodology of ontology building. We rather aim at the area of ontologies utilizable in the area of artificial intelligence, and rules are considered as a part of ontology from this reason as well. For better understanding of this paper and to demonstrate appropriateness of the suggested proposals a row of practical examples is presented in a specific case study from the area of road traffic. The paper is aimed at conceptual and logical aspects of ontologies, not their implementation.

2 Ontology content

We will engage in detail in ontology content in this chapter to show that its structure is directly related to methodology of ontology building, which is then covered in the following chapter. We have showed in the paper [1] that ontology content will depend not only on a specific observed domain, but also on a viewpoint angle of its creator. This is given by the above mentioned purpose of application and problems which are to be solved within a given domain. This problem is closely related to the level of granularity of respective ontology.

2.1 The granularity level of ontology

The conceptual separation of certain aspects of reality is strong dependent on user's needs. These needs may have a significant influence on consideration which concepts will be in a respective domain primitive, i.e. concepts that are not further refined, therefore which we don't define further by other concepts and which we will consider as complex. If we are for example interested only in solving a task of searching the closest road in a graph a *road* in ontology of such system will represent a primitive concept.

In our case study from the area of road traffic we however need to solve such tasks, such as for example a correct ordering of agents before crossroads, etc. So it is necessary to define the infrastructure model in a way so the elements being relevant from the viewpoint of agent's decision making were captured. So that the environment in which the agent is moving was for him "understandable" and unambiguously graspable by logic and his decision making. For example it is necessary to add information to each traffic lane of this lane's orientation – direction. If we didn't do it this way, the agent would have insufficient information for his decision making, e.g. he wouldn't know when he is positioned in the opposite direction, and couldn't therefore follow the rule that on a road outside of town when facing front it is driven in the lane closest to the right if no special circumstances are preventing it. Therefore the primitive concept will be here the *traffic lane* and the concept *road* will then be defined by it (the concept *traffic lane* is in view of the *road* in a part-whole relation).

So we can see that different variants of conceptualization may then differ just by the level of its granularity which the creator chooses at the beginning. The level of granularity depends on the set of primitive concepts we have decided to include into our ontology. Ontology content is therefore significantly dependent already on a level of what we intend to consider as primitive concepts. Despite of this we may at certain level of generalization present an overview of what ontology should as a structure of concepts contain.

2.2 The basic ontological types

In principle we may say that the result of ontological analysis is a conceptual hierarchy capturing the most important entities of a given domain, their attributes and generally valid relations among them. The most important relations being observed are a relation of necessary implication between properties and a part-whole relation between entities which fall under the given general concepts. Ontology content may therefore be divided into several basic parts which will be elaborated in close detail below. Ontology should contain the specification of the stable part of the system. The article [3] was for us a source for expressing the content structure, however the resulting content structure of ontology is adjusted. Ontology from our viewpoint contains:

1. Conceptual (terminological) dictionary which contains:

   a) primitive concepts,
   b) complex concepts - these are captured by primitive concepts in the part of integrity limitation of the analytical type,
   c) descriptive attributes.[5]

5 Apart from the descriptive attributes (such as e.g. *registration number of a car*) the so called *reational atributes* are often presented as well, which determine the part-whole relation. We present these as a part of the integral limitation of the empirical type.

2. Integrity limitations

   a) empirical (part-whole relations),
   b) analytical (ISA hierarchy, definitions).

3. Rules

   a) expressing nomic neccessities,
   b) expressing common necessities.

The first step in building ontology should therefore be the determination of primitive concepts which will be used for definition of complex concepts. We will set by this a level of ontology granularity which comes from tasks being solved in the application. As a primitive concept we will consider for example the already mentioned traffic lane concept by which we will then define complex concepts such as road, intersections etc. (see Table 1. for details). From the illustrative example presented at the end of the previous chapter it is obvious that the attributes are bound to the primitive and complex concepts. Its amount and a particular form are again given by the application purpose. In our case the traffic lane is going to have the attribute *direction*, and the road its *type* (e.g. road for motor vehicles, freeway etc.).

Not until defining the basic terms we may express the important statements related to relations among them. We call these relations as integrity limitations. We differentiate them into the so called analytical and empirical necessities. We consider ISA hierarchies and definitions of individual complex terms as analytical necessities. The ISA relations are sometimes mistakenly interchanged with partial relations (part-whole), which however fall into the empirical integrity limitations. Differentiating differences among them is especially important in the case of the so called inheritance, see further.

For a more detailed description of ISA relations and part-whole relations we come from [2]. ISA relations are defined by specialization and generalization. *Specialization* is used for definition of possible roles of elements of a given type (e.g. a person may be a road traffic participant). Specializations may overlap each other. E.g. a person may be a student as well as an employee. The hierarchy of road signaling may be considered as an example in our case study – parent type are *road signs*, which we further divide into *horizontal and vertical*, and *vertical* further into *order/command*, *prohibitory*, *information* etc. (Similar division is also on the side of *horizontal*). A whole hierarchy of ISA relations is being formed by this sequential specialization up to the most specialized type of sign (e.g. *direction to be followed to the right*).

On the other hand while utilizing *generalization* for various previously given types common features are searched on a very general level so that a new parent type is formed. Usually it is required that the types being generalized are mutually disjunctive and its union covers a new parent type. (For example by generalization of the terms *Traffic signs*, *Traffic Lights* and *Traffic Facilities* a new term *Traffic Signals* is formed).

ISA relations in this way represent a hierarchy among the group of concepts on whose top we will find concepts with the highest degree of generalization within the given application. ISA relations "parent-child" type is usually extensionally defined as a set inclusion. (A set of *vehicles* for example form a subset of a set of *objects in a traffic infrastructure*). By this a so called principle of inheritance is set: the lowest type (more specialized) inherits all attributes from the higher one (more general). This principle can't be then applied on the part-whole relation whose example is e.g. our relation between the terms *road* and *intersection* to the term *traffic infrastructure* as a whole, which is formed just from these two parts.

While building ontology for example for a knowledge base, it is convenient to capture also the so called *nomic necessities*, by which we understand the explicit specification of various conventions and rules we consider as obligatory and truthful empirically, not analytically. For example the laws of nature are a nomic necessity. A typical example in our case study is a rule which presents that two vehicles can't be on a given location at the same time. This rule has a character of a physical law which will be in the hierarchy of "hardness" of rules higher than for example the already mentioned rule of driving in the traffic lane closest to the right (outside of town). It is obvious that this rule is not valid without an exception and it is sort of a convention (if we are for example in the Great Britain we will be obviously driving on the left). It is therefore a so called rule describing some *common necessity*. Even though we consider only cases of countries in which it is driven on the right, there will always exist cases for which this rule may be broken. If a vehicle for example gets to a situation that could lead to breaking the "harder" rule expressing a *nomic necessity*, such decision must be carried out so that this situation would be prevented despite the price of breaking the rule expressing a *common* *necessity* – if for example there is a lane on the left of the car, the car will carry out an action of overtaking, because an event raised which allows breaking of priority of driving in the right traffic lane. On this example we see that it is really useful to differentiate the rules for the knowledge base into these basic types, because these types explicitly capture their various priorities.

3 Methodology of building ontology

The objective of methodology of building ontology is to describe a procedure, how to acquire the above mentioned structure of its contents with respect to a given domain of interest. Many research groups currently deal with problems of methodology and try to specify a step sequence leading to ontology creation. For an arbitrary scientific discipline the creation of metodics is considered as a certain achievement of maturity of the given discipline. The methodology has to be broadly acceptable, what necessarily does not mean that it must be the only one. From that point of view the ontology engineering is still a very young discipline, because although quite a number of methodologies exists for building ontology at present, any of them cannot be considered as versatile and widely acceptable. In previous years several significant methodologies occurred, the most important of which we mention below.

3.1 The overview of the current methodologies

A very good overview of methodologies is provided in the work [5], from which we proceeded in the paper. One of the first methodologies was presented in the year 1995 based on experience acquired by designing Enterprise Ontology and TOVE (TOronto Virtual Enterprise), which both were generated in the enterprise modelling domain. A year later at the European Conference on Artificial Intelligence the KAKTUS project were presented that has been building ontology for the field of electric power mains. In the context of the project methods of building ontology were presented as well. A non-ignorable methodology is also the one having been used for ontology design of the SENSUS project. As the most interesting and clearly arranged from our point of view we consider METHONTOLOGY [5], we dwell on in more detail below.

METHONTOLOGY has been created in the Laboratory of Artificial Intelligence at the Polytechnic University of Madrid. To a great extent it is based on IEEE Std 1074-1995. We decided to proceed according to *METHONTOLOGY,* because of its transparent logical structure and completeness of its steps, which express the process of building ontology. It includes the following tasks (Figure 1.):

- build glossary of terms,
- build concept taxonomies,
- build ad hoc binary relation diagrams,
- build concept dictionary,
- describe ad hoc binary relations, instance attributes, class attributes and constants,
- describe formal axioms and rules,

and finally describe instances.

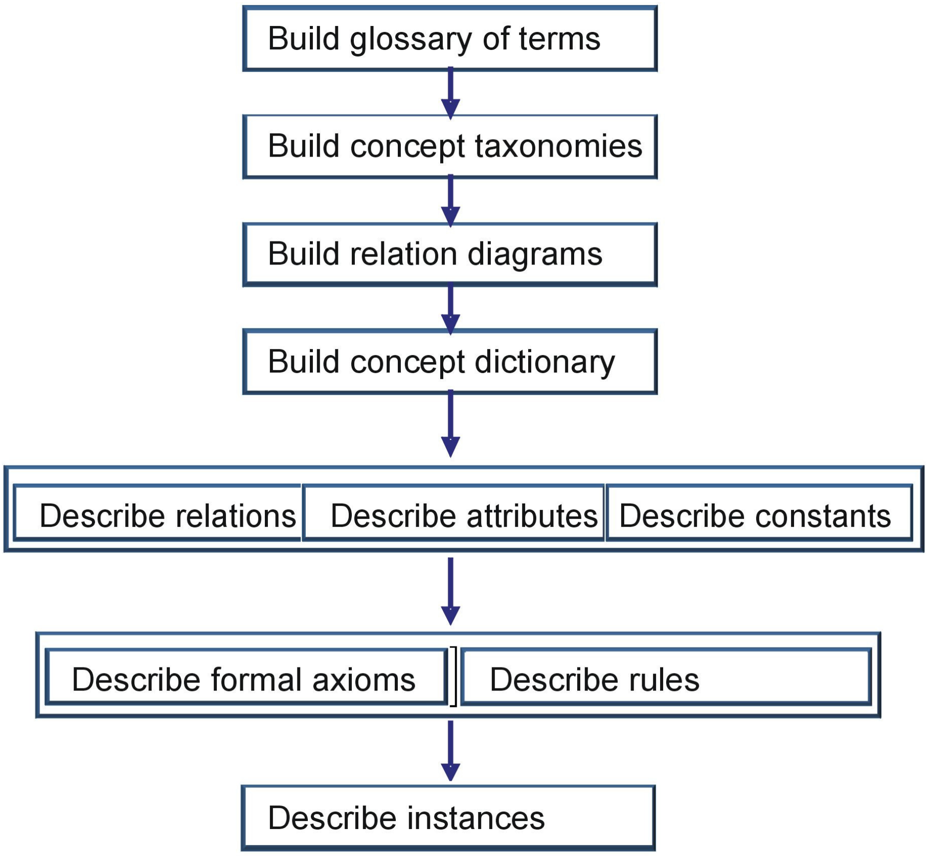

Figure 1: Tasks of conceptualization activity according to METHONTOLOGY [5].

METHONTOLOGY describes individual steps, at whose end we will obtain basic ontological types. As the majority of current methodologies it is rather focused on what ontological types should be in its contents and does not provide a more detailed guide, how to achieve adequacy of real terms of ontology, i.e. how to select convenient concepts with respect to a given domain and its tasks. Generally it is stated that the principle means for support of adequacy of factual contents of ontology are basic ontologies and content (design) samples [6]. However, we would anticipate from methodology especially that it will provide the user with a guide, what questions for instance to ask at the beginning of building ontology to be able to create effectively a glossary of terms that is a corner-stone of each building ontology.

3.2 The methodology supplement of the steps foregoing the creation of glossary of terms

We know indeed, how the procedure of building ontology should look like and in what sequence to form its different parts, but we are not sure, *in what way* to select particular concepts to the glossary. However, building the basic glossary of terms, in which we do not take into account various integrity restrictions yet, quite a number of steps must precede, which have not been directly mentioned so far by any specified methodology. So the selection process of suitable concepts is frequently a very intuitive matter. However, steps leading to creation of glossary can be specified and systematized at least generally. Our design is aimed just at such specification. So we try now to outline basic principles that have to be tracked, in order to become aware based on them, how to build the resulting ontology. We focus particularly on creation of ontology for purposes of artificial intelligence and knowledge systems.

At the beginning it is necessary to specify a purpose of the system and define so a basic system specification, which we will proceed from while building ontology. A further step is to search and gather already existing concepts in a given domain or in a field of interest. By that step we try to achieve observing the aim of shareability of ontology being built. However, the found concepts we already assess with respect to the clarified purpose of our ontology and awareness of differences between the static and dynamic parts of the given system, see below. Based on the concepts selected later so-called glossary of terms will occur that will contain basic concepts of the given domain and their definitions (Figure 2). It is necessary to say that the process preceding creation of glossary of terms is iteration.

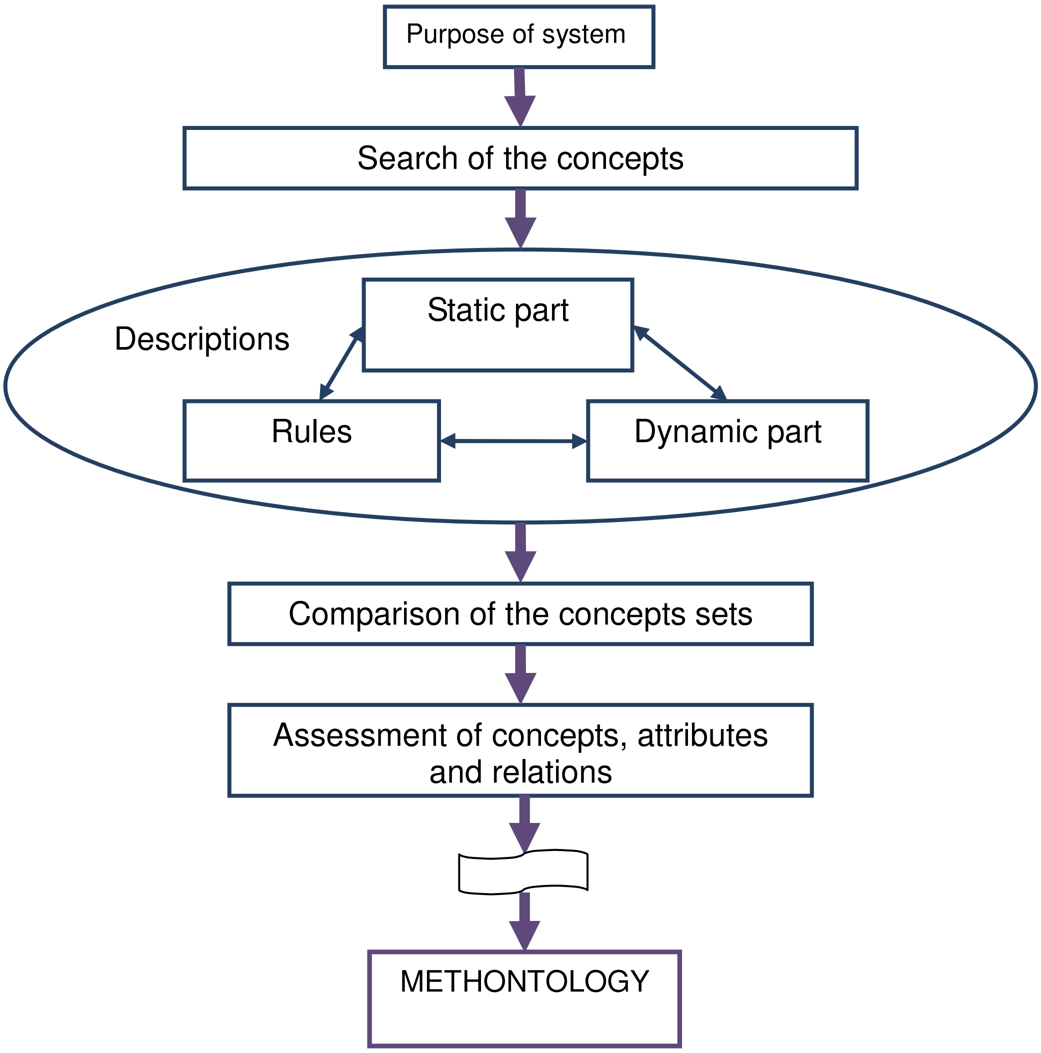

Figure 2: The proposed extension of methodology of ontology building

It is good to realize just at the very beginning of building ontology that the domain analyzed of the future knowledge system can be divided into its static and dynamic parts. This division results from the following general system definition: *Each system is a set of elements and relations between them that are purposefully defined, thus they fulfill their objective.* The static part of the system involves concepts of typical objects of the given domain. And just these concepts we should search just at the beginning of building ontology. A way out can be different sources. For this purpose, we should use not only various articles dealing with the given domain, but we especially tried to follow standards used in the field of the respective domain. Perhaps it is suitable to proceed from existing basic ontologies. A basic criterion for selection of concepts should be especially complying with *common sense* of criterion. [6] For instance, in case of a tracked domain of intelligent transport systems, a basic input was GDF standard, which is available as ISO 14825 in the valid version 4.0. The ISO standard is the result of consensus of quite a wide professional group; therefore it can be considered a useful source of basic terms and their definition for ontology. Our focus on GDF standard follows from the effort to interconnect the data modelling real infrastructure in the GIS environment with a multi-agent system.

6 For ontologies to be shared, methodology of their creation must be unified. The main method for building ontology is above all the effort to adapt created concepts to "reality" as much as possible. This “reality” is for us the reality as traditionally understood by a human entity, i.e. the reality as we understand it by our *common sense*.

The static part so involve for instance the typical concepts as follows: *vehicle, roads, crossroads* etc. The resulting set of concepts concerning the static part will capture in principle basic classes that occur in the respective domain. Consequently then it will be necessary to specify among them basic taxonomic (ISA relations) and partonomic (part-whole) relations that can exist among these classes, or perhaps even structures of dependencies. Taxonomy creates in principle an ontology skeleton.

The concepts concerning the dynamic part of the system we disclose so that we consider typical situations that can occur in the system and the problems to be solved by the system. So we obtain important attributes and relations, and in case of ontologies concerning the context of artificial intelligence, especially rules. For instance, each agent - vehicle has to solve the problem of overtaking. After analyzing the task of overtaking we obtain, for instance, concepts as *speed*, relation *be an obstacle*, etc. A rule related to the "overtaking" problems is for instance the rule of inability of taking up the equal place with a fixed entity in the same time, on whose basis we create rules for different road priorities etc. However, we often obtain by the procedure even further new classes. Let us show two examples: One of them describes implementation of a new class *junction* to ontology and the other shows how involvement of the dynamic part allows identifying other concepts.

Example 1: A simple change of the number of lanes may incur on the road in a certain place. This phenomenon must be captured in our case study because the process of decision-making of an agent on the possibility or impossibility of a lane change must necessarily follow. The agent's ability to respond to this situation becomes the key one e.g. in cases, when the lane, in which the agent is moving, ends. To be able to continue driving, i.e. to fulfil its goal, the agent must cross to another lane, which is a continuing lane in the direction of its target. Similarly, in case of lining in front of an intersection it is necessary for the agent to line correctly with regard to its target. If we had not conceptually capture the aspect of change of lanes for the agent's seeing, the agent would not be able to fulfil any of the basic tasks, which is e.g. passage through the intersection according to the given itinerary. It is evident that our ontology must conceptually capture even the phenomenon, when a simple change of lanes on the way incurs. As mentioned before, in the search for a suitable term, we adhered to the basic method of building ontology, which is the effort for shareability. We particularly tried to go through available literature, whether the phenomenon searched by us has a specific term allocated or not. GDF standard was again the input for us, specifically the term *junction*.

Example 2: Every newly captured concept understandably often brings about the need to terminologically introduce other concepts related to it. If we earmark and consequently name a place in the infrastructure, where a simple change of lanes incurs, it is, in fact, a divide, a connection between two different elements of the road. The first element has the original number of lanes, the second element the changed one (higher or lower). To finally define the term of *junction*, it is suitable to capture the concept of *road element*, which is a section of the road, where the number of lanes is not changed. Every lane has a flag, which is negligible for a driver – agent with regard to his decision-making process. We considered:

- the information on the direction in the sense of the driving direction or the opposite direction.
- the information on the fact which lane or lanes continue (a change of the number of lanes, lining up in front of an intersection).

These important terms must be included in the ontology as the lane attributes.

In conclusion of the stage of gathering concepts it is necessary to perform comparison of the acquired concepts and descriptions, on whose basis we remove redundancies, unify the description, perform corrections of the concepts. The description should represent semantics for the given purpose. We perform classification for the selected concepts - concepts, attributes, relations. Further, we express in a separate part rules that use already defined concepts. A result of the procedure is the mentioned above glossary of terms and record of rules.

4 Conclusion

General methodologies provide detailed overviews on how the ontology structure should look like with respect to individual ontological types. However, often the principal issue when building ontology tends to be just the selection of convenient concepts and description of their semantics with respect to the above mentioned tension. In this paper we tried to solve the issue and provide at least a general common guide what questions should be asked when building ontology in the area of artificial intelligence and how to proceed in order not only its formally logical correctness to be ensured, but also adequacy of its factual contents and possibility of its sharing. For better understanding we illustrated the proposed procedures on practical examples from the sphere of road transport.
Our main benefit we see in that we have tried to describe steps foregoing the creation of glossary of terms, from which other methodologies only starts, so that the term selection is systematic.

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. Číhalová, M. Ďurákova, D., Hrubá, L., Rapant, P.: *Building of Traffic Infrastructure Ontology for MAS*. In: Pešková, Kateřina (ed.). Symposium GIS Ostrava 2008, Ostrava, VSB-TU Ostrava, 2008.
2. Duží, M.: *Konceptuální modelování.* Datový model HIT. Slezská univerzita v Opavě 2000.
3. Duží, M., Materna, P.: *Concepts and Ontologies.* In Symposium EJC 2008. Ed. Yasushi Kiyoki, Takehiro Tokuda, Tsukuba Japan:Waki Print Pia, Kanagawa, Japan, 2008, Vol. 18, 45-64.
4. GDF – Geographic Data Files – 4.0, ISO/TC 204, Final Version, únor 2001
5. GÓMEZ-PÉREZ, A., FERNÁNDEZ-LÓPEZ, M., CORCHO, O. Ontological Engineering: *with exaples from the areas of Knowledge Management, e-Commerce nd the Semantic Web*. 2nd printing, Springer-Verlag, London, 2004. ISBN 1-85233-551-3.
6. Svátek, V., Vacura, M.: *Ontologické inženýrství*. In: Datakon 2007, Masarykova univerzita 2007, p. 60-91, ISBN 978-80-7355-076-9.