ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Ontologies and Knowledge Base

Vladimír SMATANÍK, Boris BUČKO
University of Žilina, Faculty of Management Science and Informatics

Abstract

People use native language to describe the world which surrounds them. Using native language is the most intuitive way to describe things. Possibility of describing the object with the combination of known phrases is the main advantage of the native language. If we examine deeper this form of world's description we could clearly see that it suffers from ambiguity because of synonyms, homonyms, pronouns or concealed words.

To extract the information from written or other recorded sources we need to know not only the meaning (semantics) of the individual expressions, but the context as well. We need to identify the contents on the basis of relationships between expressions. To do so we build up some sort of an intelligent dictionary of expressions in our brains, which we can call knowledge base.

Keywords: ontology, knowledge base, description logic, open world assumption

Description of the world

The most intuitive form of description of the world for people is natural language. Situations, relationships, real or abstract objects are assigned to expressions (words, phrases), by which they distinguish between them. The advantage of natural language is the ability to describe the object for which we do not know the exact phrase, by combining familiar words. On the other hand, this form of description of the world suffers from ambiguity because of synonyms, homonyms, confusion, or concealed expressions.

In order to properly extract information from hearing or reading the data, we need to know not only the meaning (semantics) of words, but also the context. This means identifying the contents on the basis of relations between terms. Such a method of data handling is quite natural for people, because throughout life we learn the semantics of words, while we create the context (relations) between them. This way, we build intelligent dictionary of terms in your brain, which we might call knowledge base. Using the adopted logical system we can derive new knowledge from the existing remembered facts, which have not been explicitly expressed.

How to capture knowledge in computer

Human approach towards knowledge is quite abstract. In most cases we do not realize that the world, as we perceive it in our minds, is just a subjective model of the real world. And the concept is understood as a model of reality with neglected facts, unnecessary for the studied problem.

Automated work with knowledge requires a similar approach. It is necessary to create a model of the world, capturing only necessary facts and describe it in a form computer could understand. Therefore, the proposal of an accurate model requires cooperation of not only an expert from the studied domain, but also of the knowledge engineer. The domain expert provides knowledge of the area and knowledge engineer must transform it into a suitable form. This process is not routine work. It requires experience and understanding of the principles of knowledge.

Essentially it can be argued that knowledge in a structured form consist of two interconnected components, data and metadata. The data are used for quantitative or qualitative description of the modelled phenomenon and characteristics of the observed object. Metadata are additional information that serve to describe the significance of the data themselves and the relationships between them.

The actual data can be stored in different structures, however the most commonly used for this purpose are relational database (base of data). However, current database systems largely support the storage of data in its "raw" state, with no additional information which are necessary to preserve the semantics of the concept.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

This fact lead to the introduction of the knowledge bases that allow interconnection of the data together with metadata in a common concept. During the inferring process the knowledge base remains unaltered, though the content changes (increases and is modified) during the life cycle of the system in which it is used. The changes of data are again carried by the knowledge engineer with the help of a domain expert.

The principle of knowledge bases requires a description of the knowledge by means of representation language. Representation languages – formalisms, allow using a set of signs and symbols to capture the world model from the human mind into the world model described, for example, by mathematical formulation. Possibilities and correctness of stored and derived knowledge are largely determined by the possibilities of chosen formalism.

There are several formalisms that can be used for the purpose of storing knowledge. Among the most widespread we include descriptive logic, predicate logic, production rules, semantic (associative) networks, procedural methods and frames. In order not to talk about the knowledge base only in theory, it is appropriate to describe its specific realization. Therefore, from this point on, we will focus on the more detailed description and use of descriptive logic (from now on DL), which already has built-in support in some languages.

Descriptive logic

In DL the model of the world can be understood as a set of concepts and a set of roles (relationships between concepts). All concepts and relations forms a set, called the domain. Each concept is a subset of the domain and semantic relationships define relations in this domain. Among the concepts there are always included: the universal concept ($\top$) and non-existent concept ($\bot$). Each concept is more specific than universal concept and more general than non-existent concept. Description of objects in this way is largely simplified, and therefore it is called the model.

Formally is DL based on predicate logic, with the modified language, which is sufficient for the purposes of modelling and also has good computational properties, such as decidability, or clarity. It is important to note that DL is widely recognized and mathematically correct approach.

Knowledge representation through DL is divided into two basic, interrelated components: [1]

1. **TBox** (Terminology Box) - contains terminology used in a given domain. It introduces all the concepts, defining their characteristics and role hierarchy between them. Principally, TBox can be compared to a database schema.
2. **ABox** (Assertions Box) - contains statements about individuals with the terminology of TBOX and there are assigned to concepts. Individuality is a specific instance of the concept. Concepts, therefore, describe the role of the individual and the roles describe relations between individualities. ABox can be compared to a database instance with the appropriate data.

In view of the presumption of validity of the missing information in the description of model, there are two different general approaches. One is closed-world assumption and the other is assumption of an open world. Open world assumption is an important starting point for an understanding of DL, but so far closed world assumption was widely spread.

Closed-world assumption (CWA - Closed World Assumption) is essentially the assumption that anything that is not currently known is not valid. CWA envisages that all relevant arguments are contained in formal knowledge system, so any claim that is true in the modelled world, can be formally deduced from the formally stored facts. This approach is currently the most often applied one in various information systems. Theory (and not only the scientific theory) is however rarely complete. This is way there is an alternative approach - **open world assumption** (OWA - Open World Assumption). OWA assumes that from the missing data are not automatically invalid. [2]

Open world assumption is rather unusual, but for a proper understanding of the principle of DL it is necessary to acquire it. One of its properties is the presumption of overlapping concepts. From the statement "Peter is a man" it cannot be automatically implied that Peter is not a woman. In order to be able to say that Peter is not a woman, it is necessary to specify that Man and Woman are disjoint concepts.

How to store knowledge in the ontology?

On the standard information system, it is possible to look from the perspective of layers: a data, logic and presentation layer. Using formalized

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

means such as DL, it is possible to define the ontological layer that adds semantics to the data in the data layer and combines them in a common concept. Its purpose is making the most accurate explicit definition of the intended model with respect to modelled reality. This means that the language and implementation independent concepts of the selected domain that form an abstract domain structure (fixed frame) have to be expressed by user in the chosen formalism. Ontology is in this sense the most general fundamental knowledge base in the studied domain.

Descriptive logic allows us to mathematically formulate the area of interest, but itself it is not a computer language. The terms and means of the conceptual model can be represented through a range of different languages. There are several ontology languages (OWL, RDF, CYCL, OBO, SCL, LOOM, OWL / RDF), which are used for different types of ontologies. [3]

OWL language, just like RDF, is designed to provide the means for defining classes, their properties and relations. RDF however provides only the most basic of those resources. OWL is in some versions equipped with the means of descriptive logic and thus allows the expression of restrictions on classes and provides other means for defining classes and reasoning based on a logical basis. The basic domain, which generated ontology concerns, forms classes arranged in hierarchical structures (taxonomy). Each element is representing the object of the real world belongs to a universal class (top concept) owl:Thing. The concept-oriented representation, which has the character of a multidimensional hierarchy, has also the empty class owl:Nothing, which represents an absurd concept.

Language OWL uses a standardized XML format, which because of the general prevalence allows easy sharing and transferring of knowledge between different platforms (operating systems, software tools, programming languages, etc.).[4]

Knowledge Discovery Ontology

Working with knowledge is not only about the storage and representation, but also about the access. It is very closely linked to the issue of mapping, which can be seen at several levels:

- At the level of modelling it is a mapping of ER model into the conceptual model.
- at a formal transcript level it can be seen as the conversion between relational data models and formal ontology,
- at the implementation level it is mapping of a relational database schema into RDF schema or OWL,
- and it can be also seen as mapping of the individual database instances into the instances of OWL and RDF.

To access the ontological knowledge, SPARQL language can be used. It is partly similar to database query language SQL. If the ontology is based on the DL, it is possible to use a language SPARQL-DL. There are several tools designed to create documents defining OWL ontologies for specfici domains. One of them is the Protege tool developed at the Stanford University School of Medicine, which besides the main functionality (ontology editor) includes the possibility to visualize the ontology, to gain additional knowledge by means of automatic reasoning of ontologies, and questions over created OWL using the SPARQL language. [3]

Conclusion

The issue of storing knowledge is very broad and open. Our proposed approach of using ontologies based on the principle of DL is one of the solutions based on current options. Given the growing need for data and information processing in the context it is very likely that the future will bring completely different and effective approaches.

Acknowledgements

This paper is supported by the following project: University Science Park of the University of Zilina (ITMS: 26220220184) supported by the Research&Development Operational Program funded by the European Regional Development Fund.

References

[1] M. Obitko, „Introduction to Ontologies and Semantic Web,“ 2006-2007. [Online]. Available: http://www.obitko.com/tutorials/ontologies-semantic-web/description-logics.html). [Cit. 14 10 2013].

[2] H. H. T. Z. V. M. Lukasová A., Formální reprezentace znalostí, Ostrava: Ostravská univerzita v Ostravě, 2010.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

[3] S. D. Maniraj V., „Ontology Languages – A Review," International Journal of Computer Theory and Engineering, zv. Vol.2, %1. vyd.No.6, 2010.

[4] „Semantic Web," W3C, [Online]. Available: http://www.w3.org/standards/semanticweb/. [Cit. 14 10 2013].

[5] „protégé," Stanford Center for Biomedical Informatics Research, 2013. [Online]. Available: http://protege.stanford.edu/. [Cit. 14 10 2013].J. D. Poston and W. D. Horne, "Discontiguous OFDM considerations for dynamic spectrum access in idel TV channels," in Proc. IEEE DySPAN, 2005.