International Journal of Information Technology and Business Management
29th Aug 2014. Vol.28 No.1

ISSN 2304-0777 www.jitbm.com

INTEGRATION AND LANGUAGES OF ONTOLOGY

Michal Joštiak, Marian Švalec
Department of Informatics, Faculty of Management Science and Informatics, University of Žilina, Slovakia
{michal.jostiak, marian.svalec}@uniza.sk

Abstract

From the article, we can get an overview of what ontologies are and how we most often meet with them - through the ontological languages. All examples are practically shown in e-mail communication that is easy to understand and demonstration of the basic features of described notation of ontology.

Keywords: ontology, database, languages, XML, RDF, OWL

INTRODCUTION

Current trends in storage and processing of data are adapting to handle large amount of data in order that the result does not produce a pure product of data selection or data filtering. Then on the resulting information is put criteria defining knowledge, i.e. information that is re-usable and creates a competitive advantage in the process of data processing.

The traditional method of data processing that is creating as the final product – information is trivial and in principle the simplest case of data processing. If the data do not produce information, it is necessary to define a strategy of their future use; otherwise, they are significant costs and increase the overall complexity of information systems.

Ontology penetrates through the different classifications of data and their interpretations, and finds in them a new order, a new categorization, a new meanings and context. In themselves, ontologies carry great potential for mass deployment in the field of natural sciences such as biology, medicine, but also in the commercial sector where we can find the large databases.

ONTOLOGY

In computer science ontology is a set of terms and relations between data - dictionary, which allows you to differentiate semantically different and associate semantically similar terms, and this dictionary is readable for the computer. Most often it is the data, that is large respectively it has a complex structure. The dictionary formally unifies terminology for entities and their attributes, and defines the relationships between entities. A key feature of the use of such data is to know their meaning. Therefore, for the correct interpretation it is necessary to know the semantic accuracy.

Ontology is also defined as explicit specification of a conceptualization [1]. In this sense, the conceptualization is a system of terms that describe the real system. We can say that this is a way how to describe a real system which is equally clear to us and to the computer. The method of its creation is by developing conceptual database schema. Therefore, the ontology gets the role of the extension over primary (structured, e.g. relational-database) resources to ensure the conceptual abstraction necessary for conceptual querying [2]. Ontology should be application-independent.

METHODS OF CREATING ONTOLOGIES FROM DATABASE

Existing relational databases, which contain enormous amount of data, can be used to fill ontology with knowledge. There are numerous methods how to use existing databases to fill ontology using rules or using the so-called middle model [3].

ISSN 2304-0777 www.jitbm.com

Methods of using a relational database to create ontologies:

- A creation of a new database schema based on the analysis of the existing relational schema and mapping this scheme to ontology through reverse engineering. A new data model is described as so-called middle model; it is a common model over the individual relational models.
- Implementation of instance of relational ontology through using of the external resources. It uses a declarative interface between ontology and XML data sources.
- The creation of ontology from a conceptual database schema; the aim is to create RDF (S) ontologies. However, this method has unclear semantics and do not dispose with inference model that is required for automatic inference.
- The creation of ontology from the conceptual database schema; the aim is to create OWL ontologies.

ONTOLOGY LANGUAGES

Because it is obvious that the re-use of knowledge is a critical part, the question is how to save once created information so that it can be re-used many times. Such a requirement is necessary for the definition of the advanced data processing system, regardless of the type of data that will be processed by this system. It is clear, that way-out is data representation.

The methods of data representation are based on the current data storage possibilities that are given by particular existing information and communication means for storing large amounts of data. The basic way is to store data in the form of so-called flat files, which are stored or managed by local or shared storage. In addition to the frequent characteristic formats for spreadsheets are clearly preferred formats of data storage in a structured form with defined semantics (meaning). The most common case is XML and the related formats such as SGML or RDF.

The need to exchange the large amount of data together with their description gradually led to the creation of a number of self-describing language standards. Because of its simplicity, XML is now considered as standard markup language. XML stands for eXtensible Markup Language. Language has been developed by the W3C and it is based on previous experience with markup languages (e.g. SGML). XML is now primarily recommended generalized language for describing data.

XML is the ideal format for storing structured and partially structured text that is intended for distribution and use in the various applications in the different locations. The actual XML document contains specific instructions called tags, elements and entities. The resulting document is a self-describing; therefore, it is possible to use it to define data and their meaning (semantics).

XML itself does not provide any functionality. The strength of XML-based solutions is the result of an XML document with other technologies which use a formalized description of the language data.

Declarative languages such as RDF are intended to create the formalized valid arguments about modeled world. In formation of all these arguments, it is used exclusively precisely specified meaning of words, which are defined by its position in the RDF - dictionaries / RDFS - dictionaries.

Language OWL, as well as RDFS, is designed to provide the means for defining classes, their properties and relations. The difference is primarily that the RDFS provides only most basic from these means; OWL is equipped with descriptive logic, which allows expression constraints the classes (concepts) and provide other means to define classes and the derivation means based on the logical basis. These means of descriptive logic language give the possibility to express the concept or property as a logical composition of other concepts or features. These options are not included in RDFS. In addition, if ontology is created for some domain, the formal semantics of OWL allows the derivation of the logic consequences, i.e. the other statements that are not directly included in the ontology.

ONTOLOGY IN PRACTICE

The dictionary of e-mail ontology is a relatively small set of concepts and relationships, which is known to the most of the internet users. Therefore, it will serve as the best educational base.

International Journal of Information Technology and Business Management
29th May 2013. Vol.13 No.1

ISSN 2304-0777 www.jitbm.com

Language RDF creates so-called triples that describe the attributes of an entity; it describes subject - source, predicate - object and property – value which are identified through URI [4].

Subject	Predicate	Property
Mail	has recipient	mail@exa.com
Mail	has sender	exa@mail.com
Mail	has subject	Application form
mail@exa.com	has server	exa.com

Tab.1. Triples in RDF

The main objects of each email are a sender, a recipient, a subject, a message, time of departure, time of receipt and the like. These terms can be further defined by their broad features, the recipient and the sender are the same term that can be defined as a person, respectively an e-mail address. Thus, we created the first relationships between the objects: sender is_a person, recipient is_a person. That way, with defining objects and relationships, we create ontology which can be understood by a man and a computer too. In case of the computer, the form of entry is an ontology language which can be parsed; in case of the man, the appropriate form is representation as graph.

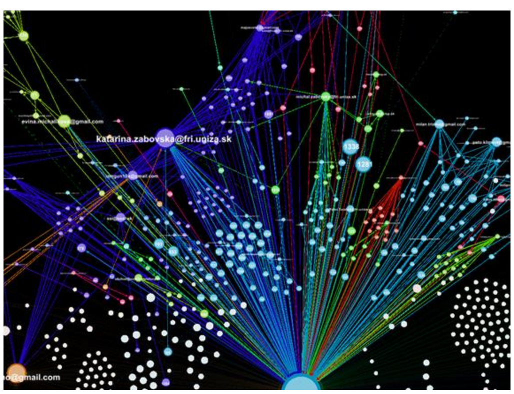

Fig.1. Graphical representation of e-mail communication in Gephi

However, the process remains unresolved either automatically or semi-automatically converting existing data into ontological forms in which we manually create or name relations - semantic precision and terms. This part is very difficult to create a fully automatically, as often nor the creators have sometimes no idea of the importance of all ontology terms. Therefore, an expert in the field should not miss in creating ontologies.

The OBO Foundry is a collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain [5]. In addition, they are trying to create a new single ontology order they have their own format for writing ontologies – OBO, and a tool for the creating and editing of OBO documents.

OBO format is based on OWL with the several extensions that are not supported by mark up languages. Language is much more readable and shorter to write. It consists of the head which defines basic things like in OWL. The biggest change is in the main attribute; it does not consist of classes but terms. In this attribute, the tags id and name are required and a number of the optional attributes that provide a declaration of properties and hierarchy of terms.

```
[Term]
id: MO:0000026
name: mail receiver
namespace: mail addresses
def: "Person who receive mail"
synonym: "recipient"
is_a: MO:0000034 ! email format
relationship: part_of MO:000092
```

Fig.2. Term Receiver in email ontology in format OBO

CONCLUSION

Design of the database schema, which uses an existing ontology (ontologies), provides a qualitatively higher level of modeling. It allows you to expand data independence (independence for database applications) on knowledge independence. Knowledge is a part of the database and it is usable for the different applications that operate over data. Proper knowledge can be created centrally over ontology and the individual applications can benefit from this knowledge.

International Journal of Information Technology and Business Management
29th May 2013. Vol.13 No.1

ISSN 2304-0777 www.jitbm.com

ACKNOWLEDGEMENT

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

REFERENCES

1. GRUBER, T. A translation approach to portable ontologies. Knowledge Acquisition, 5(2):199-220, 1993.
2. PROREC SK. Ontológia a moderná informatika. Definície ontológie [online]. 2009, [cit. 2013-03-22]. Retrieved from: <http://www.prorecsk.sk/index.php/citacka-noviniek/items/ontologia-a-moderna-informatika.html>
3. BISKUP, J., Achievements of relational database schema design Tudory revisited. Semantics in database. LCNS 1358, Springer Verlag, 1998
4. BOUDA, T., Sémantický web, standardy RDF a Topic Maps v Kurzu Digitální dokumenty (OUC, Oslo/NO, ERASMUS, podzim 2009). Inflow: information journal [online]. 2010, roč. 3, č. 2 [cit. 2013-03-22]. Retrieved from: <http://www.inflow.cz/semanticky-web-standardy-rdf-topic-maps-v-kurzu-digitalni-dokumenty-ouc-oslono-erasmus-podzim-2009>. ISSN 1802-9736.
5. THE OPEN BIOLOGICAL AND BIOMEDICAL ONTOLOGIES, [online]. [cit. 2013-03-22]. Retrieved from: <http://www.obofoundry.org/>