AN INFORMATION MODEL FOR ONTOLOGY SEARCHING AND REUSE

Tatyana Ivanova
TU-Sofia, CEE – Botevgrad
e-mail: tiv72@abv.bg
Bulgaria

Abstract: A grand number of ontologies, covering almost all domains are freely available in the internet, but many potential users prefer to design his ontology from scratch. The main reasons are related to the difficulties in searching and reusing ontologies. In fact, ontologies combine two major kinds of knowledge: domain and task knowledge (related to the tacks of it possible usage), but task knowledge is not explicitly presented, and its absence makes it difficult to reuse ontologies. In this paper the information model for supporting ontology search and reuse, bridging the gap between domain and task knowledge is presented and discussed.

Key words: ontology reuse, ontology metadata, ontology searching, ontology recommendation

1. INTRODUCTION

There are a grand number of freely available ontologies in the internet. According to the Semantic Web search engine Swoogle (http://swoogle.umbc.edu/), the number of ontologies publicly available is estimated at approximately 10 000. Some of them, such as FOAF, DOLCE or SUMO are widely used. In almost all domains, there are a number of qualitative heterogeneous ontologies that are used only in its native projects. These ontologies are hardly being used beyond the boundaries of their originating context, because of it tack dependence and difficulties in evaluating its relevance to other tacks, related to it’s domain.

There are two main approaches for reusing ontologically represented knowledge: extracting needed information by queering linked ontologies in repositories (Martin 2010), and choosing the whole ontologies for particular purpose. This work is devoted to the second approach and explores possibilities for choosing or recommending appropriate ontologies for the given tack.
Recommending ontologies is the main tack of “ontology search engines” such as Swoogle (Ding, 2004), OntoKhoj (Patel, 2003), OntoMetric (Lozano-Tello, 2005), or AkTiveRank (Alani, 2006). They crawl ontology repositories, search ontologies and rank them usually according to the searched concept density within ontologies, or popularity and relatedness. These search engines have several drawbacks:

- The searching process is performed mainly on the base of keywords, user interface doesn’t propose sufficient tools for easily specification of term’s surrounding context;
- No (or not sufficient) information about the searching persons is used in searching and ranking process.
- The searching and ranking process is not oriented to the main tack, for which the needed ontology will be used.

Many use cases in ontology evaluation and reuse, related to various domains, show that ontologies should be evaluated or ranked in the context of a particular tack (Paslaru, 2005), (Simperl, 2009). Ontologies, that are good for one tack, may be inappropriate for other, and vice versa. For example, for Semantic annotation task, labels in natural language should present concepts while for interagent communication, this is not necessary, but standardization and clear logical structure are of great importance. Many of good ontologies (for example SNOMED-CT, GALEN, FMA) are tack-independent, but are too large to effectively used and maintained. In the contrast, there are relatively small reusable ontologies in almost every domain, but they are too general, and some tack-specific extension before using them is needed. And license agreements of better ontologies may be unacceptable for many users. That is who ranking and recommendation of ontologies should be based on sufficient semantic information, describing it characteristics as well as its potential users and application tacks. This paper proposes semantic (ontology-based) information model, describing ontologies closely with its usage in typical contexts by concrete users. The model should integrate various metadata describing ontologies, users and usage tacks, sufficient for recommendation of better ontologies for given tack. To achieve that, we will first analyze main results of the latest projects, related to usage of metadata for describing ontologies, and use cases in successful ontology usage in various domains. Then we will present and discuss motivated and based on our conclusions layered ontological model of needed metadata for solving the ontology searching and recommendation problem. Our aim is to present semantically rich knowledge-based model that explicitly presents the tack-related and domain-specific characteristics of ontologies, needed for supporting or recommending its reuse.

2. RELATED WORK

There are several works attended in supporting users to choose ontologies for a new project or tack, and there are no well working methodologies that quantify the
suitability of the ontology for the given task or system. Reusing of ontologies is tightly related to it evaluation and description by usage of metadata. This election problem is based on the usage of metrics that estimate each one of the candidates (ontologies), how appropriate they are for a new system. We browse many scientific papers, investigate results from grand number of described use cases of ontology reuse, and try to compare ontology searching systems. Our main aim is to find what information is valuable in the process of searching, ranking and recommendation of ontologies, and how to organize the needed information to make it a good basis for ontology ranking and recommendation systems.

The ONTOMETRIC method (Lozano-Tello, 2005) presents taxonomy of 160 characteristics, called also multilevel framework of characteristics that provides the outline to be able to choose and to compare existing ontologies. They are organized in multilevel tree of characteristics. Content, Language, Methodology, Tool and Costs are on the upper level. ONTOMETRIC also present the set of processes that the user should carry out to obtain and compare several metrics for ontology evaluation.

(Hartmann, 2005) presents a Standard for ontology metadata (metadata model), a vocabulary of terms and definitions describing ontologies, called OMV (Ontology Metadata Vocabulary) standard. OMV specifies reusability, enhancing ontology features for human and machine processing purposes. The metadata schema contains information about Persons or Organization(s), developing ontologies, engineering process the ontology originally resulted from (OntologyEngineeringMethodology, OntologyEngineeringTool and the attributes version, status, creationDate and modificationDate). The usage history of the ontology is modelled by classes such as the OntologyTask and LicenceModel. Used ontology languages are represented by classes OntologySyntax, OntologyLanguage and Knowledge Representation Paradigm. It also contains information about ontology type (application, domain, core, task and upper-level ontologies), level of formality and types of Knowledge Representation (KR). The domain the ontology describes is represented by the class OntologyDomain that can be mapped to some topic hierarchy such as the Yahoo or DMOZ hierarchy. The metadata scheme includes also graph topology metrics represented as integer values of the DatatypeProperties numberOfClasses, numberOfProperties, numberOfAxioms, numberOfIndividuals. A framework for development and deployment of ontology metadata (DEMO) is presented in (Hartmann, 2006). This framework comprises an inventory of methods to collaboratively extend OMV in accordance to the requirements of an emerging community of industrial and academia users. DEMO proposes methodologies and tools for evolution and extension of OMV.

For ranking ontologies according to some criteria, it should be first evaluated and compared on the base of clear metrics. (Gangemi, 2006) identifies three main types of measures for ontology evaluation:

- Structural measures, that evaluate and compare ontologies as graphs;
• Functional measures, that are related to the intended use of an ontology and of its components;
• Usability-profiling measures, that depends on the level of annotation of the considered ontology.

Various **structural dimension measures** have been defined and tested on the base of the number of graph nodes, root nodes, leaf nodes, sibling nodes, levels, paths, topological properties such as depth, breadth, tangledness, and fan-outness (related to the ‘dispersion’ of graph nodes).

The **functional dimension** is coincident with the main purpose of ontology, i.e. the area of usage. They include competence adequacy (adequacy to the intended task, specificity of this task, and topic specificity); NLP adequacy (compliance with lexical distinctions), and functional modularity (ontology stratification, or granularity). Functional dimension in some cases depends from structural characteristics and it should be described by usage of standardized metadata.

**Usability-profiling measures** focus on the set of ontology annotations: the metadata about ontology and its elements, and the metadata structure. Presence, amount, completeness, and reliability are the usability measures of annotations. Annotations should contain information about structural, functional, or user-oriented properties of an ontology, as well as lifecycle-oriented properties, e.g. license, price, authorship, versioning, interfacing, etc.

As a main conclusion from the literature review, we note that ontologies are very complicated information objects that can be described by using of grand variety of parameters. Its evaluation or recommendation is of tight dependence of the intended task for it usage. Clear and standardized, well-structured and extensible model, including all the parameters, describing ontologies is needed to support ontology annotation with semantically rich metadata, needed in the process of ontology evaluation and recommendation. Our main goal in this work is to present a well working information model, giving structured representation of all the ontology parameters, shown in the literature as valuable for selecting the appropriate ontology. We believe that our model will be used as a information ground for ontology evaluation, recommendation or selection of better ontology in various contexts.

### 3. THE OVERALL ARCHITECTURE OF THE INFORMATION MODEL

#### 3.1. Requirements to the model

The information model should contain all the information, necessary for recommending ontologies in every domain for every task. It is difficult to include all this information from the beginning and consequently, the framework, based on the model should be easily extensible and modular. The Information should be easily usable for both humans and computers. That is why it schema should be based on the
standards for machine-readable information representation (XML, RDF, RDFs, OWL), and should include all the knowledge, needed for syntactic, semantic, functional and usability evaluation and comparison of ontological documents. The model also should support reliable and efficient working of ontology annotation, evaluation and recommendation systems. It should represent valuable for these tasks knowledge, free from semantic errors, contradictions or vagueness.

3.2. The layered model

We present a metadata model, consisting of layered system of mapped ontologies (fig.1). On the first (upper) layer it include the Ontology Metadata Ontology (OMO), Upper Tack Ontology (UTO), and User Model Ontology (UMO), and Additional ontologies are on the lower layers (fig. 1). Upper layer ontologies present information structures, usable in many domains, related to ontologies, tacks and users, and in lower layers specializations of upper level classes and domain-specific information is structured.

3.3. Upper ontologies

The OMO contains all the terminology, needed for comparison of ontologies in the context of specific tack (described outside the ontology). We collect the needed parameters as a result of exploring many scientific papers and projects deliberative, related to ontology metadata and ontology evaluation and reuse (for example (Lozano-Tello, 2005), (Paslaru, 2005), (Hartmann, 2006), (Montiel-Ponsoda, 2007), (Simperl, 2009), (Ding, 2004), (Patel, 2003), (Hartmann, 2005), (Alani, 2006), (Gangemi, 2006)). We also have used OMV ontology (Montiel-Ponsoda, 2007), (Hartmann, 2005) as a basis for developing our metadata ontology.
We also have used the OMV extension to capture multilingualism (lexomv OWL ontology from http://sourceforge.net/projects/omv2/files/). We have made the following main changes in the extended by multilingual information OMV model:

- Presenting almost all instances as classes to ensure mapping between OMO and low level ontologies;
- Adding new subclasses to represent valuable ontology subtypes, as core ontology, upper ontology, tack ontology, domain ontology, application ontology, lexical ontology, formal ontology, etc…
- Adding target audience class for mapping the metadata ontology to user description ontologies as FOAF or various user profile ontologies
- We also have added two ObjectProperties, related to applicability of various ontology types for given tasks.
- We have added properties for defining the typical usage of described entities (as Boolean values or as elements of calculable metrics)
- We added properties synonymy, misspelling, abbreviations for effective support of searching

![Fig. 2. The ontology metadata ontology in Protege](image)

The Upper Tack Ontology is needed for structured, machine-processable representation of terminology, related to description of all the main tacks, in which ontologies are used. Task knowledge describes a recursive decomposition of top-level tasks of ontology usage in sub-tasks, the control-flow of these sub-tasks, and the
description of the roles for the domain knowledge that is used or produced by the sub-tasks (Martins, 2008). The results of our investigation about ontology usage show that ontologies have been used in almost all application domains as an structured information representation technologies both in relatively general tacks (as searching, retrieval, information integration) and strictly domain specific tacks (as test generation in e-learning, business process modeling and prediction, etc…). The UTO represents the terminology, related to general ontology application tacks. (JASPER, 1999 identify four main categories of ontology applications:

1) Neutral authoring;
2) Ontology as specification;
3) Common access to information;
4) Ontology-based search.

For each category, it indicates their intended purpose (for example, for web searching or searching in databases), the role of the ontology (for example to support indexing, query refinement or ranking results), the supporting technologies (for example to what search machines it is recommended), who the principal actors are and what they do. In each category, we analyze specific ontology application scenarios.

The semantic of the more of terms, describing ontology usage tacks is unclear, borders of different application areas are not clearly defined, and it is difficult to present strict and consistent classification of all the terminology, related to ontology usage tacks. At the beginning, we will specify some important guidelines:

- Upper Tack Ontology should include possible synonyms of the terms and description of the context, in which the synonymy exists;
- Upper Tack ontology should describe only global (interdomain) tacks. Specific tacks of every domain should be described on the lover levels by using domain specific tacks ontologies;
- Upper Tack ontology should be terminological, and at the same time formal, supporting reasoning, free of contradictions;
- Upper Tack ontology should be multilingual, as many applications use terms in at least two languages.

Our Upper Tack Ontology has 6 upper tack classes (on the higher level):

- Knowledge representation;
- As a vocabulary (for communication and collaboration);
- Searching and information retrieval;
- Information extraction;
- Information integration;
- Annotation.

Every upper tack class has his subclasses, properties and relations to some of the other classes. For example, Knowledge representation has two subclasses: Knowledge modeling and Knowledge management. Upper tack ontology classes are mapped to higher-level classes of the imported domain –specific tack ontologies.
The People Ontology is needed not only for personalizing the recommendation of ontologies. It should be also used in the situations, when ontologies are not annotated with needed metadata, but are popular, and information, related to it successful usage can be found in internet. It should be include schema for describing web users (for example FOAF ontology), authorship information, such as provided by Dublin Core for text documents, licensing, etc.

3.4. Low layer ontologies

Additional ontologies are mainly ontologies on the lower levels. These are small domain ontologies, presented tack-oriented domain terminology, and mapped to them corresponding tack ontologies. Its presence in the model is optional (only small number of such ontologies is imported into the system at the same time). An additional domain and tack ontologies should be imported only when ontologies, intended to specific domain tacks will be searched and ranked. The system should include all the needed additional ontologies for domains or tacks, or presents strategies for searching or developing needed additional ontologies. Ontology management tools and methodologies ontologies may also be included to support recommendations, related to possibly needed modifications of returned ontologies.

Mappings between domain and tack ontologies can be specified by semantic relations of type “usable for” or “preferable ontology type”, or by using axioms. The axiomatic approach is more flexible. In the domain part of the model, some axioms can be included, saying what types of ontologies are usable or preferable for different tacks. For example:

- Ontology, that is not mapped to FOAF, not recommendable for collaboration;
- If ontology is not lexicalized, it is not appropriate for learning;
- Simple hierarchy is not suitable for learning mathematics;

Other axioms, stating recommendations about needed preliminary customization of ontology before use may be included. For example:

- If the ontology is very big domain ontology, modularization may be necessary before using it;
- If the ontology is classified as reusable, it tends to be over generalized and relevant domain knowledge possibly is omitted, thus it may require considerable modifications before they can be reutilized for specific purposes

In the tack part of our model, we can specify axioms, saying preferable requirements of some type of tacks. For example:

- For Semantic annotation task concepts should be denominated in natural language, and the natural language used in the ontology labels should be the same as the one used by the users and concepts should be denominated using naming conventions and in a linguistically predictable form to facilitate the automatic document annotation task.
• For Information retrieval task ontology should be formal to enable automatic reasoning, concepts should be denominated in natural language and should be specified to enable ontology-based query formulation and analysis. The ontology should provide a rich semantic representation (including synonyms and abbreviations) of the domain to refine the retrieval algorithm.

4. DISCUSSION

The proposed model can incorporate the grand number of ontologies. The upper level ontologies (tack ontology, ontology metadata ontology and people ontology) should be developed, formalized and tested by professionals. These ontologies represent the overall framework of the model and are permanent components of the model. Domain and specific tack ontologies are stored in the system library and may be imported on demand. Only some relatively small domain ontologies are usually needed at the same time in memory. In addition, ontology learning or collaborative ontology development methods and tools can be used for semi-automatic (interactive) ontology enrichment or evolution. As tack-related domain modeling is very specific modeling, full automation or reuse is not recommendable. Humans should check (and align) every ontology before its addition to the model.

We have validated the multilevel framework of linked ontologies by means of questionnaires. During evaluation, we have used OMO, UTO, e-learning domain, tack ontologies and computer science domain ontology. We have fixed some conceptual errors and mistakes in the Hierarchies and list of characteristics. Mappings between OMO and UTO and axioms, stating relations between tacks and ontology types should be carefully making and testing, as they include important information for domain-task relatedness and type of ontology to task relatedness. After first importing of new domain ontology in the system, the related to this domain tack ontology should be chosen or developed, and relations between them and axioms should be included.

People ontology is useful for both personalizing recommendations and usage of voluntary Web 2.0 annotations of ontologies by usage of tagging systems, folksonomies, and information, extracted from blogs. In many cases, these annotations may contain valuable metadata about ontologies.

5. CONCLUSION AND FUTURE WORK

This paper presents an information model, giving structured representation of the ontology metadata, needed for searching, recommendation and selection of the appropriate ontology both for direct usage and for future customization or evolution. The proposed model include structural representation of all types of needed in the
process of searching, evaluation and ranking metadata (including domain-related, ontology metadata, ontology usage tacks metadata, collaboration and user’s metadata). It is modular and flexible. Some initial versions of basic upper ontologies, outlining the main framework of the model, are also developed and tested. As a future work, we will develop an architecture and prototype of ontology recommendation system, using this model. We believe that our model will be used as a information ground for ontology evaluation, recommendation or selection of better ontology in various contexts.

REFERENCES