Iterative Ontology Selection Guided by User for Building Domain Ontologies

Asma Minyaoui, Faiez Gargouri
Miracl Laboratory, University Of Sfax
Technology Park, PB: 242, Sakiet Ezzit
Sfax, Tunisia 3021

Abstract

In this paper we present a new method for ontology selection in a reuse context. The novel feature of this method is the iterative selection of the reused ontologies. Ontology selection is guided by the user according to his requirements and his perception to the target domain. Starting from a first selected ontology, the concepts with the weakest density are identified then the ontology developer is enabled to choose among them the ones to be refined in order to cover a specific scope of the domain.

Introduction

Domain ontologies are considered as an instrument for knowledge representation, sharing, reuse and interoperability (Tankeleviciene and Damasevicius 2009). Their developing is a time-consuming and expensive process which makes the reuse approach increasingly adopted in order to reduce development costs and to enhance interoperability. Domain ontologies aim to model a specific domain which represents a part of the world. Such modelisation depends strongly on the personal perspective of the ontology developer. In fact, different ontologies for the same domain can arise due to the different perception of the domain based on cultural background, education, ideology, etc. Besides, each domain has a large expand with boundaries that can change from a developer to another. Usually, these points are treated while building domain ontologies from scratch. In the present research we aim to take into account these points in a reuse context, mainly during the selection of the ontologies to be reused. The main advantage of the proposed ontology selection method is the iterative selection process where each iteration is guided by the user according to his perception to the domain. In result, the final ontology is built incrementally; each increment aims to cover a specific scope of the domain according to the user requirements.

Ontology Reuse

Ontology re-use is an agreed upon goal in ontology engineering. It reduces the cost of creating ontologies, improves the quality of the resulting ontologies, and eases later interaction between systems(Stecher et al. 2008). In the ontology creation context, ontology reuse can have several forms(Pinto, Gómez-Pèrez, and Martins 1999):

- Integration of ontologies when building a new ontology reusing other available ontologies: the result ontology can be built from assembling other publicly available ontologies that match the appropriate requirements (adequate levels of detail and granularity, language, etc.)
- Integration of ontologies by merging different ontologies about the same subject into a single one that ”unifies” all of them: the resulting ontology is built by merging ideas, concepts, distinctions, axioms, etc, from other existing ontologies on exactly the same subject. This ontology tries to unify concepts, terminology, definitions, constraints, etc., from all of them. So, ontologies are merged, unified into a single one.

Practices show that one single ontology is usually unable to cover a set of given terms that represent a domain. Most often query terms are spread over two or more ontologies(Vrandečić and Sure 2007). However, it is more likely to find ontology combinations that jointly cover a given domain. The analysis of several case studies of ontology reuse in(Pinto and Martins 2000)reveals a general ontology reuse process as follows:

1. Finding ontologies: candidate ontologies discovery
2. Selecting the ones to be reused: evaluating ontology quality according to a set of criteria
3. Customization: once the set of reusable ontologies has been determined, they are adapted to the user’s need.
4. Merging: ontologies covering similar domains are merged to one.
5. Integration: ontologies modelling different domains are integrated into final application ontology and to the application system.

As shows the described reuse process, most of the ontology reuse approaches start by selecting a combination of ontologies at first, then after that merging them in order to get a larger ontology that cover the given domain.

Ontology Selection

The discovery of the appropriate ontology or ontology part is a central ingredient for effective ontology reuse(Stecher et al. 2008). As it is the first phase of the reuse process, its march may shape the rest of the phases and the quality of the resulting ontology. Ontology selection is defined as the process that allows identifying one or more ontologies or ontology modules that satisfy certain criteria(Sabou et al. 2006). Some researches(Sabou et al. 2006)consider that ontology evaluation is core to ontology selection since the evaluation is the main task of a selection process. Despite the fact that ontology evaluation is a crucial task, there is no general widely accepted definition yet. In common, Ontology evaluation is considered as the task of assessing the quality of an ontology(Vrandečić and Sure 2007). The main question to be answered is: how to measure the quality of a given ontology? Many works have tried to answer this question by proposing several methods and approaches for ontology evaluation in order to select the appropriate ones. In the following, we exemplify some of these approaches:

1. Gold standard : the idea is to use similarity measures to compare an ontology with an existing ontology that is considered as a reference (Ehrig and al. 2005)
2. Application based: this approach suppose that the quality of an ontology is directly proportional to the performance of an application that uses it (Sabou, Lopez and Motta 2006).
3. Data driven: the quality of the ontology is represented by its appropriateness to cover the topic of a corpus (Brewster and al. 2004).
4. Assessment by humans to show how well the ontology meets a set of predefined criteria, standards, requirements as in OntoMetric (Lozano-Tello, Gomez-Perez and Sosa 2003) and the peer-review based approach (Supekar 2005).

Various web based ontology selection method was proposed to search and rank ontologies. In the following, we exemplify some of these methods:

- Swoogle :is a semantic web search engine that crawls, indexes, and stores all types of semantic web document in a triple store(Ding et al. 2004). It uses ranking standards that are similar to those used by Google. In fact, Swoogle adopt a PageRank-like method by analyzing links and referrals between ontologies.
- AKTiveRank: is a system for ranking ontologies by aggregating various measures to assess the structural features of concepts(Alani and Brewster 2005),such as structural density, concept centrality, and semantic similarity.
- Ontosearch: is a semantic web engine that searches and queries web ontology and their associated data sets(Pan, Thomas, and Sleeman 2006). It provides the functionality to query an ontology repository and to add additional ontologies to the index by providing the URI of an ontology.
- OntoKhoj: is a semantic web portal that crawls, classifies, ranks and search ontologies(Patel et al. 2003). It extended the functionality of the Google PageRank algorithm and developed the OntoRank algorithm that considers different types of links and additional constraint like distance.
- OntoSelect: uses dynamically organized ontology libraries(Buitelaar et al. 2004). It includes functionality for searching ontologies based on three standards, namely, coverage, structure, and connectedness.

Proposed Approach

Current ontology selection approaches are based on ontology evaluation techniques and usually give valuable results. Unfortunately these results stay not reliable in a reuse context since the evaluation process was led regardless to the intended use of the ontology and regardless to the developer specific needs. In fact, most of these approaches consider ontology selection as an independent task. Quality is not a property of something but a judgment so must be relative to some purpose(Sabou et al. 2006). In fact, ontology selection comes usually as a sub-task of an ontology reuse process. Each reuse process has its own particularities and requirements. Previous approaches do not take into consideration this point. In consequence their result may be not useful for the reuse process. From another side, most of the ontology reuse approaches start by selecting a combination of ontologies at first, then after that merging them in order to get a larger ontology that cover the given domain. In the present work we propose a novel ontology selection method which is reuse oriented. In fact, we propose an iterative method for selecting ontology for reuse. The main question that we care about is how to select the set of ontologies that cover the best a given domain according to the ontology developer requirements.In fact, we propose an iterative method for selecting ontologies for reuse where we allow user to guide the selection of the combination of ontologies that will form jointly the final one. The selection will be iterative and will take end on user demand. Each iteration have a different input given by the user and is not blindly guided by an algorithm as happens in other methods. In this work we admit two main assumptions .First, the target domain has a large expand with boundaries that can change from a developer to another according to his perception of the domain, so it is up to him to define the scope of his interest. Secondly, Ontology modelisation depends strongly on the personal perspective of the ontology developer due to the different perception of the domain based on cultural background, education, ideology,etc. Usually, these points are treated while building domain ontologies from scratch. In the present research we aim to take into account these points but in a reuse context, mainly during the selection of the ontologies to be reused. Many works admit that an iterative process can be helpful in better adjusting the results to query terms. Most of them use iterations within query expansion phase as in(Pinto, Gómez-Pèrez, and Martins 1999). At the best of our knowledge, none of them have proposed an iterative ontology selection guided by the user within the selection phase. The idea of the proposed method is as following:

1. In one hand we have a Domain D to be represented; in the other hand we have a set of candidate ontologies O1.On.
2. User types query in natural language.

3. Ontology selection phase: in order to identify the ontology that (1) covers the query terms and (2) has the best ranking after an evaluation process. At this preliminary stage of work we use Swoogle to provide the candidate ontologies.

4. If no ontology has been found, query expansion phase takes place and so on until finding the ontology Oi.

5. When Oi is identified, Poor Concepts' detection phase takes place. The object of this phase is to identify concepts in Oi that have the weakest density. In fact, when searching for a "good" representation of a specific concept, one would expect to find a certain degree of detail in the representation of the knowledge concerning that concept (Alani and Brewster 2005). So, we will consider concepts with a low number of links as concepts that must be refined. To identify these concepts we will use the Density Measure (DEM) proposed by (Alani and Brewster 2005). DEM is intended to approximate the representational-density of classes and consequently the level of knowledge detail. A concept C is considered as a "poor" concept when dem[c] ¡ DEM[o]. Where dem(c) is the Density Measure for class c and DEM[o] is the density of the whole ontology. Let$S = \{S1, S2, S3, S4\} = \{subclasses[c], superclasses[c], relations[c], siblings[c]\}$

$$dem[c] = \sum_{i=1}^{4} wi|Si|$$

$$DEM[o] = \frac{1}{n}\sum_{i=1}^{n} dem[c]$$

where wi is a weight factor set to a default value of 1, $n = E(o,T) + P(o,T)$ which is the number of matched classes in ontology o, and E(o,T) and P(o,T) are the sets of classes of ontology o that have labels that match any of the search terms t exactly or partially, respectively.

6. The output of the previous step is a set of "poor" concepts. These concepts have a low density so they are considered as underrepresented and need to be refined. The proposed system allows to the user to choose the concepts that he wants to refine. The user is invited to choose a set of three concepts that he considers as relevant to his need. In this way user is adjusting the next ontology selection to his perception of the domain and to his specific needs. Actually, he is indicating the direction in which ontology will be granulated.

7. The labels of the chosen concept will be used as the input of the Ontology selection phase, here the iterations start. We will get a new ontology Oi+1 that covers the new set of terms issuing from the label of the chosen "poor" concepts.

8. A merging phase is required now to merge Oi and Oi+1 into one single ontology O. Here the final ontology starts to be built incrementally(Merging technique is not yet addressed). The user can put an end to the process if he is satisfied with the ontology O.

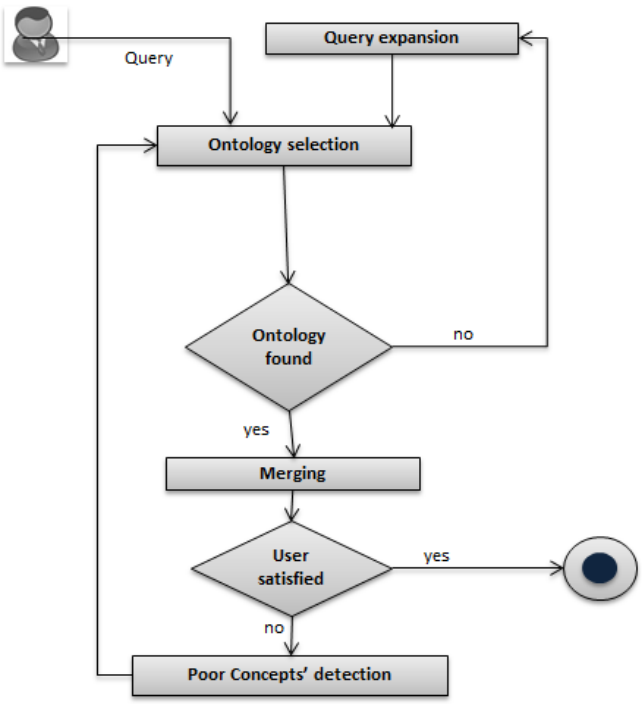

Figure 1: Iterative Ontology Selection

9. The ontology O goes as an input to the Poor Concepts' detection phase.

10. Go back to step (3).

The proposed method is illustrated in the figure 1.

Conclusion And Future Work

In this work a new method for ontology selection for reuse context have been presented. The novel feature of this method is the iterative selection of the reused ontologies. Ontology selection is guided by the user according to his requirements and his perception to the target domain. In this stage of work we focus on the iterative selection aspect of the presented approach. Next steps in our planned work will focus on the eventual ontology evaluation techniques to be used in the selection phase as well as the merging phase. We aim to start building the system described in this paper soon and experiment with it to evaluate how well such an approach may work in real life scenarios.

References

Alani, H., and Brewster, C. 2005. Ontology ranking based on the analysis of concept structures. In *K-CAP '05: Proceedings of the 3rd international conference on Knowledge capture*, 51–58. New York, NY, USA: ACM.

Buitelaar, P.; Eigner, T.; ; and Declerck, T. 2004. Ontoselect: A dynamic ontology library with support for ontology selection. in . In *Proceedings of the Demo Session at the ISWC'04, Hiroshima, Japan, Nov. 2004*.

Ding, L.; Finin, T.; Joshi, A.; Pan, R.; Cost, R. S.; Peng, Y.; Reddivari, P.; Doshi, V.; and Sachs, J. 2004. Swoogle: a search and metadata engine for the semantic web. In *CIKM '04: Proceedings of the thirteenth ACM international conference on Information and knowledge management*, 652–659. New York, NY, USA: ACM.

Pan, J. Z.; Thomas, E.; and Sleeman, D. 2006. Ontosearch2: Searching and querying web ontologies. In *In Proc. of the IADIS International Conference*, 211–218.

Patel, C.; Supekar, K.; Lee, Y.; and Park, E. 2003. Ontokhoj a semantic web portal for ontology searching, ranking, and classification. In *Proc. 5th ACM Int. Workshop on Web Information and Data Management*, 58–61.

Pinto, H. S., and Martins, J. P. 2000. Reusing ontologies. In *In AAAI 2000 Spring Symposium on Bringing Knowledge to Business Processes*, 77–84. AAAI Press.

Pinto, H. S.; Gómez-Pèrez, A.; and Martins, J. P. 1999. Some issues on ontology integration. *Praxis* 18(Borst 1997):1–12.

Sabou, M.; Lopez, V.; Motta, E.; and Uren, V. 2006. Ontology selection: Ontology evaluation on the real semantic web.

Stecher, R.; Niedere, C.; Nejdl, W.; and Bouquet, P. 2008. Adaptive ontology re-use: finding and re-using sub-ontologies. *IJWIS* 4(2):198–214.

Vrandečić, D., and Sure, Y. 2007. How to design better ontology metrics. In *Proceedings of the 4th European conference on The Semantic Web: Research and Applications*, ESWC '07, 311–325. Berlin, Heidelberg: Springer-Verlag.