Guiding the Ontology Matching Process with Reasoning in a PDMS

François-Élie Calvier and Chantal Reynaud

LRI, Univ Paris-Sud & INRIA Saclay - Île-de-France
4, rue Jacques Monod - Bât. G
91893 Orsay Cedex FRANCE
{francois.calvier,chantal.reynaud}@lri.fr
http://www.lri.fr/iasi

Abstract. This article focuses on ontology matching in a decentralized setting. The work takes place in the MediaD project. Ontologies are the description of peers data belonging to the peer data management system SomeRDFS. We show how to take advantage of query answering in order to help discovering new mappings between ontologies, either mapping shortcuts corresponding to a composition of pre-existent mappings or mappings which can not be inferred from the network but yet relevant.

Key words: ontology matching, peer-to-peer, data management systems.

1 Introduction

Our work takes place in the setting of the MediaD project[1], which aims at creating a peer-to-peer data management system (PDMS) called SomeRDFS [1] allowing the deployment of very large applications that scale up to thousands of peers. We are interested in making the generation of mappings automatically supported by query answering. We propose to use query answering to generate mapping shortcuts and to identify relations, denoted target relations, which are starting points in the mapping discovering process. These relations allow identifying relevant mapping candidates limiting in that way the matching process to a restricted set of elements. Discovered mappings can be relevant or not according to the strategy involved in the PDMS. Indeed, a peer can decide to look for new mappings whatever they are (default strategy denoted S_1) or to look for particular mappings: either (strategy denoted S_2) new mappings involving peers already logically connected to it (there exists a mapping between their two ontologies) or (strategy denoted S_3) mappings involving peers not yet logically connected to it.

The paper is organized as follows. Section 2 shows how the query answering process can be used. Section 3 focuses on the identification of mapping candidates from target relations. We conclude and outline remaining research issues in Section 4.

[1] Research project funded by France Telecom R&D

2 Using Query Answering

2.1 Mappings Shortcuts Discovery

A mapping shortcut is a composition of mappings. Mapping shortcuts consolidate PDMSs by creating direct links between indirectly connected peers.We propose a two-step automatic selection process. We first identify potentially useful mappings shortcuts exploiting query answering. In this step, the goal is to retain only mappings which would be useful in the rewriting process with regard to the queries really posed by users to the peer $\mathcal{P}$. However, all these mappings will not be systematically added to the set of mappings of $\mathcal{P}$ because the usefulness of some of them may be low. Thus, we propose then a second selection step based on filtering criteria which can be different from one peer to another.

To achieve the first step we need to distinguish the rewriting and evaluation phases of query answering. Query answering will not be a unique and global process anymore but two connected processes which can be separated if needed. This separation allows to identify the relations that are interesting according to the user, i.e. the ones whithin the obtained rewritings he has chosen to evaluate.

The second selection step is based on the strategy of the peer and potentially exploits filtering criteria defined by the administrator of this peer. The usable criteria are specific to each peer but are limited. They concern either the kind of user who posed the query which originated the mapping (user-criterion) or the kind of relation belonging to $\mathcal{P}$ involved in the mapping (relation-criterion).

2.2 Identification of Target Relations Using Query Answering

In our approach we consider that a relation is a target relation if it is an obstacle for its peer in achieving the strategy it has chosen to implement. The definition of a target relation will then be based on a counting function. That function will differ according to the strategy of the peer and also according to the method used to count. The result of the counting function will be compared to a threshold that will be fixed by the administrator of the peer. When the value of the function is lower than the threshold the relation will be a target relation.

Definition (Target Relation) $\mathcal{P}_1$:R_1 is a target relation iff f($\mathcal{P}_1$:R_1) < t, f being a counting function and t a threshold.
In [2], we precise the definition of the function f for the relation R_1 of the peer $\mathcal{P}_1$ according to the strategy chosen by the peer and according to the method, C_1 or C_2, used to count. C_1 operates with regard to the knowledge of the peer, its ontology and its mappings. C_2 is based on rewritings obtained from queries.

If the strategy of $\mathcal{P}_1$ is S_1 the result of f($\mathcal{P}_1$:R_1) is the number of distant relations specializing R_1. If the strategy of $\mathcal{P}_1$ is S_2 the result of f($\mathcal{P}_1$:R_1) is the number of distant peers involved in the set of relations more specific than R_1. If the strategy of $\mathcal{P}_1$ is S_3, R_1 will be a target relation if there is at least one peer involved in a low number of specialization statements of R_1. Thus, f($\mathcal{P}_1$:R_1) provides the minimum number of relations of a given distant peer specializing R_1.

3 Obtaining a Set of Relevant Mapping Candidates

Target relations are used to identify a restricted set of mapping candidates according to two scenarios. In the first scenario, let us consider $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ three peers with C_1, C_2 and C_3 three classes and the following mappings: $\mathcal{P}_1{:}C_1(X) \Rightarrow \mathcal{P}_2{:}C_2(X)$ and $\mathcal{P}_3{:}C_3(X) \Rightarrow \mathcal{P}_2{:}C_2(X)$, each known by the two involved peers. This scenario is represented Figure 1.

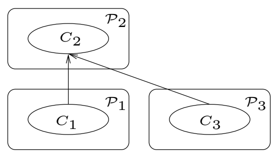

Fig. 1. Scenario 1

From the point of view of $\mathcal{P}_1$ $C_1(X)$ is a target relation . That target relation is interesting since $\mathcal{P}_1{:}C_1(X) \Rightarrow \mathcal{P}_2{:}C_2(X)$ is a mapping in $\mathcal{P}_1$, $Q_5(X) \equiv \mathcal{P}_2{:}C_2(X)$ could be a query posed to $\mathcal{P}_2$ by $\mathcal{P}_1$. The obtained rewritings would be $\mathcal{P}_1{:}C_1(X)$ and $\mathcal{P}_3{:}C_3(X)$ and looking for mappings between all the relations belonging to this set of rewritings is relevant.

In the second scenario let us consider $\mathcal{P}_1$ and $\mathcal{P}_2$ two peers, $\mathcal{P}_1{:}C_1$, $\mathcal{P}_2{:}C_2$ and $\mathcal{P}_2{:}C_3$ three classes. $\mathcal{P}_2{:}C_2(X) \Rightarrow \mathcal{P}_2{:}C_3(X)$ is a statement in $\mathcal{P}_2$. $\mathcal{P}_2{:}C_2(X) \Rightarrow \mathcal{P}_1{:}C_1(X)$ is a mapping in $\mathcal{P}_2$ and $\mathcal{P}_1$. This scenario is represented Figure 2.

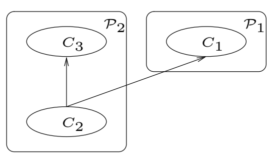

Fig. 2. Scenario 2

From the point of view of $\mathcal{P}_2$ $C_2(X)$ and $C_3(X)$ are target relations. This scenario is interesting since $\mathcal{P}_2{:}C_2(X) \Rightarrow \mathcal{P}_1{:}C_1(X)$ is a mapping in $\mathcal{P}_2$, it could be relevant to look for mappings between $C_1(X)$ and $C_3(X)$, two relations which subsume $C_2(X)$.

For each target relation we look for sets of mapping candidates, denoted MC. Our approach is based on the idea that it is relevant to look for connections between relations if they have common points. In our setting the common point that we are going to consider is a common relation, either more general or more specific. The construction of the set of mapping candidates can be achieved according to two processes, one for each scenario.

4 Conclusion

In this paper we have presented how SomeRDFS query answering can offer an automated support for discovering new mappings. In particular, we have shown that query answering in a decentralized setting can be used to select elements which are relevant to be matched when the number of elements to be matched is a priori huge and when no peer has a global view of the ontologies in the network. Our approach is based on query answering and filtering criteria. Future work will be devoted to the alignment process itself performed on each set of mapping candidates and relying on earlier work done in the group [3].

References

1. Adjiman, P., Chatalic, P., Goasdoué, F., Rousset, M.C., Simon, L.: Distributed reasoning in a peer-to-peer setting: Application to the semantic web. JAIR **25** (2006) 269–314
2. Calvier, F.E., Reynaud, C.: Guiding the ontology matching process with reasoning in a pdms. Technical Report 1495, CNRS-Université Paris Sud - LRI (June 2008)
3. Reynaud, C., Safar, B.: When usual structural alignment techniques don't apply. In: The ISWC'06 workshop on Ontology matching (OM-06). (2006)