Approximate Query Reformulation based on Hierarchical Ontology Mapping

Jun-ichi Akahani and **Kaoru Hiramatsu** and **Tetsuji Satoh**
NTT Communication Science Laboratories, NTT Corporation
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
{akahani, hiramatu, satoh}@cslab.kecl.ntt.co.jp

Abstract

This paper proposes an approximate query reformulation framework for integrating multiple ontologies. In order to achieve semantic interoperability in the Semantic Web, multiple ontologies have to be integrated. Ontology integration requires approximation mechanisms, since often no perfectly corresponding ontologies exist. However, most previous research efforts on ontology integration have not provided clear semantics for approximation. In this paper, we propose a formal framework for approximate query reformulation and provide a reformulation method based on hierarchical ontology mapping.

Introduction

In order to achieve semantic interoperability in heterogeneous information services, ontologies have been widely used. As the research and development on ontologies such as the Semantic Web have grown, several domain ontologies have been constructed. However, the de-centralized nature of the Web makes it difficult to construct or standardize a single ontology. Ontology integration is thus necessary.

When integrating ontologies, those that correspond exactly are seldom found. For example, there may be no corresponding class for Cajun restaurants in a Japanese ontology for restaurants. In such a case, one may use an approximation mechanism to replace "Cajun" with the American restaurant class in the Japanese ontology. However, most previous research efforts on ontology integration do not provide clear semantics for approximation.

In this paper, we propose a formal framework for approximate query reformulation (Akahani, Hiramatsu, & Kogure 2002). In this framework, a query represented in one ontology is reformulated approximately into a query represented in another ontology by using an ontology mapping specification. In order to characterize *closer* reformulation, we adapt and extend the notion of *maximally-contained* reformulation (Halevy 2000) in the database literature. Specifically, we introduce two kinds of reformulation: *minimally-containing* reformulation and *maximally-contained* reformulation.

In our framework, ontology mapping specifications are described as an ontology. This paper focuses on one-to-one subsumption mapping between ontologies, which we call *hierarchical ontology mapping*. We provide a reformulation method based on hierarchical ontology mapping by introducing two kinds of reformulation operators: the *most special generalizers* for minimally-containing reformulation and the *most general specializers* for maximally-contained reformulation.

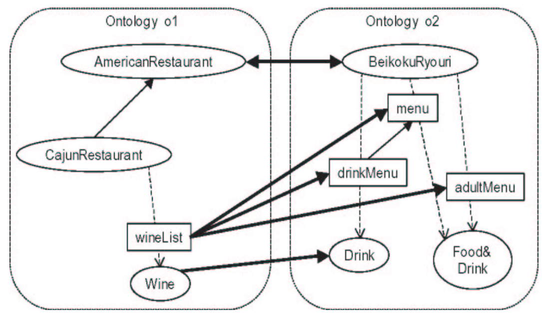

Figure 1: Ontology Mapping

In the following sections, we first propose an approximate query reformulation framework. We then provide a reformulation method based on hierarchical ontology mapping. Finally, we relate our framework to previous efforts in the field and present our conclusions.

Approximate Query Reformulation

Preliminaries: Queries over Terminological Knowledge Bases

In this section, we define queries over terminological knowledge bases extending a framework for approximate terminological queries (Stuckenschmidt & van Harmelen 2002).

We first define terminological knowledge bases labeled by ontologies.

Def. 1 (Terminological Knowledge Base in an Ontology) *Given a set $\mathcal{C}^i$ of classes and a set $\mathcal{R}^i$ of relations in an ontology O^i, and a set $\mathcal{O}$ of objects, a terminological knowledge base (KB) T^i in the ontology O^i is a set of axioms of the forms*

- $C_1^i \sqsubseteq C_2^i$ *where* $C_1^i, C_2^i \in \mathcal{C}^i$.
- $R_1^i \sqsubseteq R_2^i$ *where* $R_1^i, R_2^i \in \mathcal{R}^i$.
- $C^i(a)$ *where* $C^i \in \mathcal{C}^i$ *and* $a \in \mathcal{O}$.
- $R^i(a, b)$ *where* $R^i \in \mathcal{R}^i$ *and* $a, b \in \mathcal{O}$.

For example, the terminological KB T^1 in ontology `o1` shown in Figure 1 contains the following axiom.

$$\texttt{CajunRestaurant}^1 \sqsubseteq \texttt{AmericanRestaurant}^1$$

Similarly, the terminological KB T^2 in ontology `o2` contains the following axiom.

$$\texttt{drinkMenu}^2 \sqsubseteq \texttt{menu}^2$$

The semantics of a terminological KB is defined using an interpretation function $\mathcal{I}$ in the usual way. We define the models of KB and logical consequences as follows.

Def. 2 (Semantics) *An interpretation $\mathcal{I}$ is a model for the KB T^i if $\mathcal{I} \models A$ for every axiom $A \in T^i$ where $\models$ is defined as follows:*

- $\mathcal{I} \models C_1^i \sqsubseteq C_2^i$ *iff* $\mathcal{I}(C_1^i) \subseteq \mathcal{I}(C_2^i)$.
- $\mathcal{I} \models R_1^i \sqsubseteq R_2^i$ *iff* $\mathcal{I}(R_1^i) \subseteq \mathcal{I}(R_2^i)$.
- $\mathcal{I} \models C^i(a)$ *iff* $\mathcal{I}(a) \in \mathcal{I}(C^i)$.
- $\mathcal{I} \models R^i(a,b)$ *iff* $\mathcal{I}(\langle a,b\rangle) \in \mathcal{I}(R^i)$.

An axiom A logically follows from a set of axioms S (denoted by $S \models A$) if $\mathcal{I} \models S$ implies $\mathcal{I} \models A$ for every model $\mathcal{I}$.

We next define queries over terminological KBs. A query is a conjunction of a query about classes and a query about relations.

Def. 3 (Query) *Let $\mathcal{V}$ be a set of variables disjoint from $\mathcal{O}$. A query Q^i in ontology O^i is of the form*

$$Q_C^i \wedge Q_R^i$$

where

- Q_C^i *is a boolean combination of $C^i(x)$ where $C^i \in \mathcal{C}^i$ and $x \in \mathcal{O} \cup \mathcal{V}$,*
- Q_R^i *is a boolean combination of $R^i(x,y)$ where $R^i \in \mathcal{R}^i$ and $x, y \in \mathcal{O} \cup \mathcal{V}$.*

For example, the following denotes a query about a Cajun restaurant that has Chardonnay on its wine list in ontology `o1`.

$$\texttt{CajunRestaurant}^1(x) \wedge \texttt{wineList}^1(x, \text{`}\texttt{Chardonnay}\text{'})$$

The answer to a query Q^i can be obtained by substituting variables $v_1, \cdots, v_n$ contained in Q^i by a tuple $\langle a_1, \cdots, a_n\rangle$ of objects. We denote this substitution σ. The answer set $A(Q^i)$ is a set of tuples such that $T^i \models Q^i\sigma$. The semantic relations between different queries are defined as follows.

Def. 4 (Query Containment) *For queries Q_1 and Q_2, Q_1 is said to be contained in Q_2 (denoted by $Q_1 \sqsubseteq Q_2$) if $A(Q_1) \subseteq A(Q_2)$.*

Mapping among Multiple Ontologies

In our framework, a query represented in one ontology is reformulated approximately into a query represented in another ontology by using an ontology mapping specification. In this paper, we focus on one-to-one subsumption relations between classes and relations. We call this type of mapping *hierarchical ontology mapping*.

Def. 5 (Hierarchical Ontology Mapping) *Ontology mapping M^{ij} between ontology O^i and O^j is a set of assertions and is divided into four subsets as follows:*

$$M^{ij} = M_{cg}^{ij} \cup M_{cs}^{ij} \cup M_{rg}^{ij} \cup M_{rs}^{ij}$$

where

- $M_{cg}^{ij} = \{C^i \sqsubseteq C^j \mid C^i \in \mathcal{C}^i \text{ and } C^j \in \mathcal{C}^j\}$,
- $M_{cs}^{ij} = \{C^j \sqsubseteq C^i \mid C^i \in \mathcal{C}^i \text{ and } C^j \in \mathcal{C}^j\}$,
- $M_{rg}^{ij} = \{R^i \sqsubseteq R^j \mid R^i \in \mathcal{R}^i \text{ and } R^j \in \mathcal{R}^j\}$,
- $M_{rs}^{ij} = \{R^j \sqsubseteq R^i \mid R^i \in \mathcal{R}^i \text{ and } R^j \in \mathcal{R}^j\}$.

For example, ontology mapping M^{12} between ontology O^1 and O^2 in Figure 1 is the union of the following sets.

$$\begin{aligned}
M_{cg}^{12} &= \{\texttt{AmericanRestaurant}^1 \sqsubseteq \texttt{BeikokuRyouri}^2, \\
&\quad\ \texttt{Wine}^1 \sqsubseteq \texttt{Drink}^2\} \\
M_{cs}^{12} &= \{\texttt{BeikokuRyouri}^2 \sqsubseteq \texttt{AmericanRestaurant}^1\} \\
M_{rg}^{12} &= \{\texttt{wineList}^1 \sqsubseteq \texttt{drinkMenu}^2, \\
&\quad\ \texttt{wineList}^1 \sqsubseteq \texttt{menu}^2, \\
&\quad\ \texttt{wineList}^1 \sqsubseteq \texttt{adultMenu}^2\} \\
M_{rs}^{12} &= \phi
\end{aligned}$$

There may be many possible reformulated queries, but we prefer closer reformulation. We therefore adapt and extend the notion of *maximally-contained* reformulation (Halevy 2000) in the database literature. Specifically, we characterize two kinds of reformulation: minimally-containing reformulation and maximally-contained reformulation. In minimally-containing reformulation, the reformulated query minimally covers the original query. On the other hand, in maximally-contained reformulation, the reformulated query is maximally covered by the original query.

Assuming that ontology mapping M^{ij} is consistent with KBs T^i and T^j, we characterize approximate query reformulation using query containment in the merged KB $T^i \cup M^{ij} \cup T^j$. We extend the definition of the answer set $A(Q)$ to be a set of tuples such that $T^i \cup M^{ij} \cup T^j \models Q\sigma$. Approximate query reformulation is defined as follows.

Def. 6 (Approximate Query Reformulation) *Let Q^i and Q^j be queries in ontology O^i and O^j, respectively.*

- *Q^j is an equivalent reformulation of Q^i if $Q^j \sqsubseteq Q^i$ and $Q^i \sqsubseteq Q^j$.*
- *Q^j is a* **minimally-containing** *reformulation of Q^i if $Q^i \sqsubseteq Q^j$ and there is no other query Q_1^j in O^j such that $Q^i \sqsubseteq Q_1^j$ and $Q_1^j \sqsubseteq Q^j$.*
- *Q^j is a* **maximally-contained** *reformulation of Q^i if $Q^j \sqsubseteq Q^i$ and there is no other query Q_1^j in O^j such that $Q^j \sqsubseteq Q_1^j$ and $Q_1^j \sqsubseteq Q^i$.*

Recall the example query above. A reformulated query

$$\texttt{Beikokuryouri}^2(x) \wedge \texttt{drinkMenu}^2(x, \text{`}\texttt{Chardonnay}\text{'})$$

is not a minimally-containing reformulation, as there is a minimally-containing reformulated query as follows;

$$\begin{aligned}
&\texttt{Beikokuryouri}^2(x) \wedge \texttt{drinkMenu}^2(x, \text{`}\texttt{Chardonnay}\text{'}) \\
&\wedge \texttt{adultMenu}^2(x, \text{`}\texttt{Chardonnay}\text{'})
\end{aligned}$$

Reformulating Queries Approximately

In this section, we address how queries are reformulated based on hierarchical ontology mapping. Specifically, we present two kinds of reformulation operators: the most special generalizers for minimally-containing reformulation and the most general specializers for maximally-contained reformulation.

A reformulated query consists of classes and relations appeared in the *range* of ontology mapping. Intuitively, classes and relations in a minimally-containing reformulation should minimally subsume those in the original query. Therefore, calculation of the least upper bounds for classes and relations is necessary. Similarly, maximally-contained reformulation requires calculation of the greatest lower bounds. We first define the least upper bounds and greatest lower bounds for a class and a relation. This definition is an extended version of (Stuckenschmidt 2002).

Def. 7 (Least Upper Bounds and Greatest Lower Bounds) *Let C be a class, T be a KB, and TC be a set of classes, then the least upper bounds $LUB(C, T, TC)$ and greatest lower bounds $GLB(C, T, TC)$ are defined as follows:*

- $LUB(C, T, TC) = \{C' \mid C' \in TC, T \models C \sqsubseteq C'$ *and there is no other* $C'_1 \in TC$ *such that* $T \models C \sqsubseteq C'_1$ *and* $T \models C'_1 \sqsubseteq C'\}$.
- $GLB(C, T, TC) = \{C' \mid C' \in TC, T \models C' \sqsubseteq C$ *and there is no other* $C'_1 \in TC$ *such that* $T \models C' \sqsubseteq C'_1$ *and* $T \models C'_1 \sqsubseteq C\}$.

The least upper bounds and greatest lower bounds for a relation are defined similarly.

The mapping range $D(M^{ij})$ of ontology mapping M^{ij} is defined to be a set of classes and relations in ontology O^i that appear in M^{ij}.

Minimally-containing reformulation requires calculation of the least upper bounds of classes and relations in the original query with respect to the merged KB. As ontology mapping is divided into four subsets, we only take into consideration M^{ij}_{cg} and M^{ij}_{rg} for classes and relations, respectively. For example, the least upper bounds of a class $\texttt{CajunRestaurant}^1$ with respect to a KB $T^1 \cup M^{12}_{cg}$ in the mapping range of M^{12}_{cg} is the following set.

$$LUB(\texttt{CajunRestaurant}^1, T^1 \cup M^{12}_{cg}, D(M^{12}_{cg})) = \{\texttt{Beikokuryouri}^2\}$$

Similarly, the least upper bounds of a relation $\texttt{wineList}^1$ with respect to a KB $T^1 \cup M^{12}_{rg}$ in the mapping range of M^{12}_{rg} is the following set.

$$LUB(\texttt{wineList}^1, T^1 \cup M^{12}_{rg}, D(M^{12}_{rg})) = \{\texttt{drinkMenu}^2, \texttt{adultMenu}^2\}$$

Using the least upper bounds, we can define the most special generalizers for class queries and relation queries.

Def. 8 (Most Special Generalizers) *Let M^{ij} be an ontology mapping and T^i be a KB in ontology O^i. A most special generalizer for a class query $C^i(x)$ is defined as follows:*

$$MSG(C^i(x)) = C^j_1(x) \wedge \cdots \wedge C^j_n(x)$$

where $LUB(C^i, T^i \cup M^{ij}_{cg}, D(M^{ij}_{cg})) = \{C^j_1, \cdots, C^j_n\}$.

A most special generalizer for a relation query $R^i(x, y)$ is defined as follows:

$$MSG(R^i(x, y)) = R^j_1(x, y) \wedge \cdots \wedge R^j_n(x, y)$$

where $LUB(R^i, T^i \cup M^{ij}_{rg}, D(M^{ij}_{rg})) = \{R^j_1, \cdots, R^j_n\}$.

Based on the above examples of least upper bounds, we have the following most special generalizers.

$$MSG(\texttt{CajunRestaurant}^1(x)) = \texttt{Beikokuryouri}^2(x)$$

$$MSG(\texttt{wineList}^1(x, \text{`}\texttt{Chardonnay}')) = \texttt{drinkMenu}^2(x, \text{`}\texttt{Chardonnay}') \wedge \texttt{adultMenu}^2(x, \text{`}\texttt{Chardonnay}')$$

Applying these most special generalizers, the query in ontology `o1`

$$\texttt{CajunRestaurant}^1(x) \wedge \texttt{wineList}^1(x, \text{`}\texttt{Chardonnay}')$$

is reformulated approximately into the query in ontology `o2`

$$\texttt{Beikokuryouri}^2(x) \wedge \texttt{drinkMenu}^2(x, \text{`}\texttt{Chardonnay}') \wedge \texttt{adultMenu}^2(x, \text{`}\texttt{Chardonnay}').$$

Similarly, we can define the most general specializers for class queries and relation queries using the greatest lower bounds.

Def. 9 (Most General Specializers) *Let M^{ij} be an ontology mapping and T^i be a KB in ontology O^i. A most general specializers for a class query $C^i(x)$ is defined as follows:*

$$MGS(C^i(x)) = C^j_1(x) \vee \cdots \vee C^j_n(x)$$

where $GLB(C^i, T^i \cup M^{ij}_{cs}, D(M^{ij}_{cs})) = \{C^j_1, \cdots, C^j_n\}$.

A most special generalizer for a relation query $R^i(x, y)$ is defined as follows:

$$MGS(R^i(x, y)) = R^j_1(x, y) \vee \cdots \vee R^j_n(x, y)$$

where $GLB(R^i, T^i \cup M^{ij}_{rs}, D(M^{ij}_{rs})) = \{R^j_1, \cdots, R^j_n\}$.

The following theorem assures the correctness of our framework.

Theorem 1 *Let Q^i be a query in ontology O^i, then*

- *if Q^i is reformulated into Q^j_g in ontology O^j by the most special generalizers, then Q^j_g is a minimally-containing reformulation of Q^i,*
- *if Q^i is reformulated into Q^j_s in ontology O^j by the most general specializers, then Q^j_s is a maximally-contained reformulation of Q^i.*

Related Work

Approximate terminological query framework (Stuckenschmidt & van Harmelen 2002) provides a formal framework for query approximation. However, in this respect, query approximation is used to improve the efficiency in a single ontology. Thus, they did not provide ontology mapping. The approximate information filtering framework (Stuckenschmidt 2002) has also been proposed. However, they only dealt with class hierarchies and the maximally-contained reformulation in our framework.

Most previous research efforts on ontology integration have used ad-hoc mapping rules between ontologies (as surveyed in (Wache *et al.* 2001)). This approach allows flexibility in ontology integration, but most works do not provide semantics for the mapping rules. One exception is the Ontology Integration Framework (Calvanese, De Giacomo, & Lenzerini 2002) which provides clear semantics for ontology integration by defining *sound* and *complete* semantic conditions for each mapping rule. However, each mapping rule and its semantic conditions have to be specified by users. It is therefore difficult to ensure consistency in the mapping rules. In contrast, our framework can generate *sound* and *complete* mapping rules by specifying ontology mapping. It is relatively easy to check the consistency, since ontology mapping specifications are described as an ontology.

The OBSERVER system (Mena *et al.* 2000) provides primitives for defining relationships between ontologies, such as *synonym*, *hypernym*, and *hyponym*. Furthermore, it provides a query reformulation algorithm using these primitives However, except for *synonym*, these mapping primitives do not have formal semantics. Although reformulated queries are evaluated by a closeness metrics, they are not logically grounded. In contrast, our approximate query reformulation provides reformulation operators with clear semantics.

Conclusions

In this paper, we presented a formal framework for approximate query reformulation. In order to characterize closer reformulation, we introduced two kinds of reformulation: minimally-containing reformulation and maximally-contained reformulation. We have shown that approximate query reformulation can be defined with query containment by virtually merging source ontology, target ontology and ontology mapping.

We also provided a reformulation method based on hierarchical ontology mapping by introducing two kinds of reformulation operators. For minimally-containing reformulation, the most special generalizers reformulate a class (or relation) expression in an original query into conjunction of the least upper bounds of the class (or relation). For maximally-contained reformulation, the reformulated query by the most general specializers is disjunction of the greatest lower bounds. We showed the correctness of our framework.

The approximate query reformulation framework has been incorporated into the GeoLinkAgent system (Akahani *et al.* 2002). In the prototype system, agents coordinate regional information services provided by the GeoLink system (Hiramatsu *et al.* 2000), which is used in the Digital City Kyoto prototype (Ishida *et al.* 1999). Approximate query reformulation is required for such domains that have cross-cultural aspects, because ontologies vary from region to region due to cultural differences.

This paper focused on hierarchical ontology mapping. More complex mapping specification may be necessary for flexibility in ontology integration. However, it is not clear our theorem holds for complex mapping such as conjunction of relations. Many interesting issues require further investigation.

References

Akahani, J.; Hiramatsu, K.; Furukawa, Y.; and Kogure, K. 2002. Agent-based coordination of regional information services. In *Digital Cities II: Computational and Sociological Approaches*, LNCS 2362, 283–291. Springer-Verlag.

Akahani, J.; Hiramatsu, K.; and Kogure, K. 2002. Coordinating heterogeneous information services based on approximate ontology translation. In *Proc. of AAMAS-2002 Workshop on Agentcities: Challenges in Open Agent Systems*, 10–14.

Calvanese, D.; De Giacomo, G.; and Lenzerini, M. 2002. A framework for ontology integration. In Cruz, I.; Decker, S.; Euzenat, J.; and McGuinness, D., eds., *The Emerging Semantic Web*. IOS Press. 201–214.

Halevy, A. Y. 2000. Theory of answering queries using views. *SIGMOD Record* 29(4):40–47.

Hiramatsu, K.; Kobayashi, K.; Benjamin, B.; Ishida, T.; and Akahani, J. 2000. Map-based user interface for Digital City Kyoto. In *Proc. of INET2000 The Internet Global Summit*.

Ishida, T.; Akahani, J.; Hiramatsu, K.; Isbister, K.; Lisowski, S.; Nakanishi, H.; Okamoto, M.; Miyazaki, Y.; and Tsutsuguchi, K. 1999. Digital City Kyoto: Towards a social information infrastructure. In *Proc. of International Workshop on Cooperative Information Agents (CIA-99)*, 23–35.

Mena, E.; Illarramendi, A.; Kashyap, V.; and Sheth, A. P. 2000. OBSERVER: An approach for query processing in global information systems based on interoperation across pre-existing ontologies. *Distributed and Parallel Databases* 8(2):223–271.

Stuckenschmidt, H., and van Harmelen, F. 2002. Approximating terminological queries. In *Proc. of the 4th International Conference on Flexible Query Answering Systems (FQAS'02)*. Springer-Verlag.

Stuckenschmidt, H. 2002. Approximate information filtering with multiple classification hierarchies. *International Journal of Computational Intelligence and Applications* 2(3):295–302.

Wache, H.; Voegele, T.; Visser, U.; Stuckenschmidt, H.; Schuster, G.; Neumann, H.; and Huebner, S. 2001. Ontology-based integration of information – a survey of existing approaches. In *Proc. of IJCAI 2001 Workshop on Ontologies and Information Sharing*.