Schema Level and Data Level Mapping Composition

Md Anisur Rahman
Comouter Science Department
Khulna University, Bangladesh

Mehedi Masud
Computer Science Department
Taif University, Taif, Saudi Arabia

ABSTRACT

Schema mappings and data mappings constitute essential building blocks of data integration, data exchange and peer-to-peer data sharing systems. At present, either schema-level mappings or data-level mappings are used for data sharing purposes. In this paper we consider the semantics of bi-level mapping that combines the schema-level and data-level mappings. Tabular representation of the mappings helps to solve many mapping-related algorithmic and semantic problems, like mapping composition. Composition of mappings between sources has several computational advantages in a peer data sharing system, such as yielding more efficient query translation and pruning redundant paths. Considering the need of mapping composistion, this paper presents a mechanism for composing two bi-level mappings by using tableaux.

General Terms:

Database, Peer to Peer

Keywords:

Mappings, data Interoperability, mapping composition

1. INTRODUCTION

Integrated access to distributed and heterogeneous information sources, e.g. *data integration*, *data exchange* and *P2P data sharing*, is an important research area. In data integration systems data residing at different sources are combined to provide the users with a unified view [1]. In data exchange systems, instance of a target schema is created from data structured under a source schema [2]. In a P2P data sharing system each peer can exchange data with a set of other peers that are independently created [3]. Schema mappings is the main technique for sharing and exchanging data in all the systems resolving data heterogeinity among the sources. A schema mapping describes the relationship between two database schemas at high level. There are three approaches for specifying the schema mappings: *local-as-view* (LAV), *global-as view* (GAV), and *global-and-local-as-view* (GLAV). GLAV is a mixed approach and specifies the mappings by a set of assertions of the form $\forall \vec{x}(\exists \vec{y} \phi_{\mathcal{S}}(\vec{x}, \vec{y}) \rightsquigarrow \exists \vec{z} \phi_{\mathcal{G}}(\vec{x}, \vec{z}))$ where $\phi_{\mathcal{G}}$ is a query over $\mathcal{G}$ and $\phi_{\mathcal{S}}$ is a query over $\mathcal{S}$. Here, $\mathcal{G}$ is global schema and $\mathcal{S}$ is source schema.

GAV specifies the mappings by a set of assertions of the form $\forall \vec{x}(\phi_{\mathcal{S}}(\vec{x}) \rightsquigarrow g(\vec{x}))$ where g is an element of global schema (target schema) $\mathcal{G}$ and $\phi_{\mathcal{S}}$ is a query over source schema $\mathcal{S}$. Relations in $\mathcal{G}$ are views and queries are expressed over the views. Queries can simply be evaluated over the data satisfying the global relations. LAV approach specifies the mappings by a set of assertions of the form $\forall \vec{x}(s(\vec{x}) \rightsquigarrow \phi_{\mathcal{G}}(\vec{x}))$ where s is an element of $\mathcal{S}$ and $\phi_{\mathcal{G}}$ is a query over $\mathcal{G}$.

Peer data sharing systems use either schema-level or data-level mappings to resolve schema as well as data heterogeneity among data sources (peers). Schema-level mappings create structural relationships among different schemas. On the other hand, data-level mappings associate data values in two different sources. Creating a unique global mediated schema is impractical in a peer data sharing systems due to the volatility, peer autonomy, and scalability issues. Instead, mappings are implemented only between pairs of peers to unify the heterogeneous data sources. These mappings describe the relationship between the terms used in different peers. Schema mappings [4, 5, 6, 7, 8] and data mappings [9] are used to address the schema-level and data-level heterogeneity, respectively. Several strategies are introduced for the schema mappings between the mediated and local schemas including, global-as-view (GAV) [1], local-as-view (LAV) [8], and global-and-local-as-view (GLAV) [5]. In GAV approach, the mediated schema is described in terms of local sources. In LAV, the local sources are described in terms of the mediated schema. GLAV is the combination of GAV and LAV approaches to integrate the mediated and local schemas. Schema-level mappings are effective only when the differences between the schemas are mainly structural, i.e. attribute values represent the same information, or can be transformed to be the same. However, data-level mappings are necessary when semantically related attribute values differ. Data mappings are implemented by mapping tables [9] which are relations on the attributes being mapped. The tuples in the mapping tables show the correspondence between values in the mapped relations. These tables are treated as constraints (aka mapping constraints) on the exchange of data between peers. These two kinds of mappings are complementary to each other. However, existing peer database systems have been based solely on either one of these mappings. We show in [10, 11] that if both mappings are addressed simultaneously in a single framework, the resulting approach enhances data sharing in a way such that we can overcome the limitations of the non-combined approaches. In this paper, we consider mapping composition in a model of a peer data sharing system (PDSS) which uses bi-level mapping. Mapping composition is based on Tableaux [12] that represent schema mappings and data-level mappings in a general form. Later we consider this Tableaux for mapping composition.

1.1 Motivation

Combining two mappings into a single one is called mapping composition. Composition of mappings between data sources has several computational advantages, such as yielding more efficient query translation, pruning redundant paths, and better query execution plans. This paper considers a tableau based [12] technique for composing the bi-level mappings. Essentially, the bi-level mappings are first expressed by tableaux [12]. Then the composition is performed by manipulating those tableaux.

1.2 Objectives

Follwoing are the overall objective of this paper:
1. Analyze the problems to compose mappings of schema and data level mappings in data sharing environments.
2. Propose a framework that enables heterogeneous data sources to be shared efficiently considering data and schema level heterogeneity. Moreover, propose a model to compose mappings between two indirect data sources.

2. LITERATURE REVIEW

Schema mappings [4, 5, 6, 7, 8] and data mappings [9] are used to address the schema-level and data-level heterogeneity, respectively. Several strategies are introduced for the schema mappings between the mediated and local schemas including, global-as-view (GAV) [1], local-as-view (LAV) [8], and global-and-local-as-view (GLAV) [5]. In GAV approach, the mediated schema is described in terms of local sources. In LAV, the local sources are described in terms of the mediated schema. GLAV is the combination of GAV and LAV approaches to integrate the mediated and local schemas. Schema-level mappings are effective only when the differences between the schemas are mainly structural, i.e. attribute values represent the same information, or can be transformed to be the same. However, data-level mappings are necessary when semantically related attribute values differ. Data mappings are implemented by mapping tables [9] which are relations on the attributes being mapped. The tuples in the mapping tables show the correspondence between values in the mapped relations. These tables are treated as constraints (aka mapping constraints) on the exchange of data between peers. In the following, some related works regarding the mappings of peer database management systems are discussed.

In a peer data sharing system, each peer chooses its own database schemas, and maintains data independently of other peers. Contrary to the traditional data integration systems where a global mediated schema is required for data exchange, in a PDMS the semantic relationships exist between a pair of peers, or among a small set of peers for sharing data. The data is shared globally among the peers by traversing the transitive relationships among semantically related peers. Creating a unique global mediated schema is impractical in a PDMS due to the volatility, peer autonomy, and scalability issues. Instead, mappings are implemented only between pairs of peers to unify the heterogeneous data sources. These mappings describe the relationship between the terms used in different peers. Hyperion system [11, 14] addresses the problem of mapping data in P2P systems where different peers may use different values to identify or describe the same data. Hyperion relies on mapping tables that list pairs of corresponding values for search domains that are used in different peers. Mapping tables provide the foundation for exchanging information between peers.

The Piazza system [16] provides a solution to peer data management system where the single logical schema of data integration systems is replaced by a set of mediator schemas that are interlinked to define semantic mappings between the peer schemas. Piazza uses two data integration formalisms local-as-view (LAV) and global-as-view (GAV) for peer mappings. GAV is used to define relations of the mediator's schema over the relations in the sources and LAV is used to define relations in the sources over the mediated schema. In Piazza, a reformulation algorithm for query processing is presented that addresses both GAV and LAV mappings. However, the Piazza system considers only the schema-level heterogeneity among the peers.

The SASMINT system [17] provides a solution for supporting interoperability infrastructures that enables sharing and exchange of data among diverse sources. The system mainly finds and resolves syntactic, semantic, and structural conflicts among schemas and matches schemas automatically. This system can be very useful to discover mappings between peers automatically. The system mainly discovers mappings between two peers which are directly connected. However, in this project the mappings are discovered between two peers which are connected through other peers in the path.

3. SYSTEM MODEL FOR MAPPING COMPOSITION

In this section we present a data sharing system model Π where we conisder the composition of mappings. Before defining the system we recall different notions as presented in [10, 11].

A data source $S_i \in \mathcal{S}$ is a tuple, where $S_i = (PS_i, R_i, L_i)$. Where,

- PS_i is the source schema through which data in a source is exposed to the external world.
- R_i is the set of sources comprised of local and external sources.
- L_i is the set of GLAV local mappings which define the mappings between R_i and PS_i. Each local mapping, called *mapping assertion* (aka *tuple generating dependency*), in L_i has the form

$$\forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \rightsquigarrow \exists\vec{z}\psi(\vec{x},\vec{z}))$$

where $\varphi(\vec{x},\vec{y})$ and $\psi(\vec{x},\vec{z})$ are conjunctive queries over the relations in R_i and PS_i respectively.

Source mappings $\mathcal{M}_i \in \mathcal{M}$ is a set of mappings, called bi-level mapping or *source mappings*, that define the schema and data-level mappings between sources. The construction of mappings $M_i^j \subseteq \mathcal{M}_i$ forms an acquaintance (*i, j*) between S_i and S_j. Each mapping $m \in \mathcal{M}_i$ is a pair $< m_{j,k}^S, m_{j,k}^D >$, where:

- $m_{j,k}^S$ is a GLAV mapping (practically GAV, since $s(\vec{x})$ is always a single relation) of the form

$$\forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \rightsquigarrow s(\vec{x}))$$

where $\varphi(\vec{x},\vec{y})$ is a conjunctive query over the peer schema of a peer P_j and $s(\vec{x})$ is the k^{th} external source of S_i.
- $m_{j,k}^D$=MT=$\{mt_1, mt_2, \ldots, mt_q\} \subseteq MT_j^i$ is a set of mapping tables. MT_j^i denotes the set of mapping tables used to map data of S_j to data of S_i.
  m can alternatively be represented with the mapping assertion as follows:

$$\forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \overset{MT}{\rightsquigarrow} s(\vec{x}))$$

A data sharing system Π is defined by a pair $\langle\mathcal{S},\mathcal{M}\rangle$, where $\mathcal{S} = \{S_1, S_2, \cdots, S_n\}$ is a set of sources and $\mathcal{M} = \{\mathcal{M}_1, \cdots, \mathcal{M}_n\}$ is a set of peer mappings.

It is assumed that a curator with expertise in different domains is responsible for generating the mapping tables. Schema mappings between two sources are initially created by the corresponding source administrators when they agree to share data. Generating the mappings automatically is another research area and this paper does not address this issue.

The semantics of source mappings can be given in terms of FOL.

Let $MT = \{mt_1[\vec{P_1}, \vec{Q_1}], mt_2[\vec{P_2}, \vec{Q_2}], \ldots, mt_n[\vec{P_n}, \vec{Q_n}]\}$ be a set of mapping tables, where for any pairs of mapping tables $(mt_i[\vec{P_i}, \vec{Q_i}], mt_j[\vec{P_j}, \vec{Q_j}])$, $mt_i \neq mt_j \implies \vec{P_i} \neq \vec{P_j}$, then $MT[\vec{x}]$ denotes a set of tuples resulted from transformation of $\vec{x}$ by all the member mapping tables of MT. Formally,

The semantics of source mappings is defined below. It is already mentioned that a mapping assertion of a source mapping is of the form:

$$\forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \overset{MT}{\leadsto} s(\vec{x}))$$

Let us assume that the above assertion defines an external source s of source S_i in terms of the source schema of a source S_j. An interpretation of the schema of S_i and S_j satisfies the assertion if that interpretation satisfies the following formula

$$\forall\vec{x}\forall\vec{z}(\exists\vec{y}(\varphi(\vec{x},\vec{y}) \wedge \vec{z} \in MT[\vec{x}]) \equiv s(\vec{z}))$$

A mapping can be interpreted as a definition of how the data of the external source would be instantiated by the data of other peers. The formula also tells us that before instantiating the external source, data is converted using the corresponding mapping tables of MT. However, if there is no data-level heterogeneity, no mapping table is needed. In that case, an empty mapping table ϕ is used in the assertion. In that case, a peer mapping is represented as $\forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \overset{\phi}{\leadsto} s(\vec{x}))$ which satisfies the FOL formula $\forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \equiv s(\vec{x}))$.

A data sharing system Π is given in terms of a set of models that satisfy the local and source mappings of Π. Let a source database $\mathcal{D}$ for Π be a disjoint union of a set of local databases in each source S_i of Π. Given a source database $\mathcal{D}$ for Π, the set of models of Π relative to $\mathcal{D}$ is:

$$sem^{\mathcal{D}}(\Pi) = \{\mathcal{I} | \mathcal{I} \text{ is a finite model of all peer theories } F_i \text{ relative to } \mathcal{D}\text{, and } \mathcal{I} \text{ satisfies all peer mappings}\}$$

4. MAPPING COMPOSITION

Consider three data sources A, B, and C and the mappings $M_{A\to B}$ and $M_{B\to C}$ among them. In general, two mappings $M_{A\to B}$ and $M_{B\to C}$ can be composed when there exists a common relation between them. For creating the mapping $M_{A\to C}$ from $M_{A\to B}$ and $M_{B\to C}$, first the mappings are converted into tableaux [12] T_{SAB} and T_{SBC}. Then the tableaux are merged into a tableau T_{SAC}. Tableaux T_{SAC} contains all the information to generate $M_{A\to C}$. A homomorphism funtion θ_{AB} is used to map the elements between $M_{A\to B}$ and $M_{B\to C}$ to produce $M_{A\to C}$.

In the following we show the theory to represent bi-level mappings using tabular forms.

Let $\mathfrak{m} : \forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \overset{MT}{\leadsto} E(\vec{x}))$ be an arbitrary bi-level mapping between sources S_1 and S_2, where $\vec{x} = (x_1, \ldots, x_p)$ and $\vec{y} = (x_1, \ldots, x_q)$. Note that $\varphi(\vec{x},\vec{y})$ must have relations only from the source S_1 and $E(\vec{x})$ must be an external data source of source S_2. A MAT $\mathcal{T}^E_{\varphi_{MT}}$ for $\mathfrak{m}$ is a partitioned table whose left hand side partition is the summary-free source tableau $\mathcal{T}_{\varphi_{MT}}$ (representing the query $\{\vec{x'} | \forall\vec{x'}\exists\vec{y'}\varphi_{MT}(\vec{x'},\vec{y'})\}$) and the right hand side partition is the summary-free target tableau $\mathcal{T}_E$ (representing the query $\{\vec{x'} | \forall\vec{x'}E(\vec{x'})\}$)) Where,

```
Procedure BLM2SMT($\mathfrak{m}$)
Input : A Bi-Level mapping $\mathfrak{m} : \forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \overset{MT}{\leadsto} E(\vec{x}))$
Output: An SMT $T^{\psi}_{\varphi}$ representing $\mathfrak{m}$
begin
    Let $\sigma \equiv \forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \rightarrow E(\vec{x}))$
    $\mathcal{T}^{\psi}_{\varphi} = \langle\mathcal{T}_{\varphi}, \mathcal{T}_{\psi}\rangle = TGD2SMT(\sigma)$
    for each $mt[P,Q] \in MT$ do
        add $Q$ to $T_{\varphi}.Columns$
        add a new row to $T_{\varphi}.Rows$ with tag $mt$
        $\mathcal{T}_{\varphi}.Rows[mt][P] \leftarrow FreshNonDistinguished()$
        for each $R \in \mathcal{T}_{\varphi}.Tags$ do
            if $IsDistinguished(\mathcal{T}_{\varphi}.Rows[R][P])$ then
                $Temp = \mathcal{T}_{\varphi}.Rows[R][P]$
                $\mathcal{T}_{\varphi}.Rows[R][P] = \mathcal{T}_{\varphi}.Rows[mt][P]$
                $\mathcal{T}_{\varphi}.Rows[mt][Q] = Temp$
            endif
        endfor
    endfor
    $\mathcal{T}^{\psi}_{\varphi} \leftarrow \langle\mathcal{T}_{\varphi}, \mathcal{T}_{\psi}\rangle$
    return $\mathcal{T}^{\psi}_{\varphi}$
end
```

Fig. 1. Procedure BLM2SMT

$$\begin{aligned}
\vec{x'} &\equiv (\vec{x} - \vec{P} + \vec{Q}) \\
\vec{y'} &\equiv (\vec{y} + \vec{P}) \\
\vec{P} &\equiv \bigcup_{mt(\vec{p},\vec{q}) \in MT} \vec{p} \\
\vec{Q} &\equiv \bigcup_{mt(\vec{p},\vec{q}) \in MT} \vec{q} \\
\varphi_{MT} &\equiv \varphi(\vec{x'},\vec{y'}) \bigwedge_{mt(\vec{p},\vec{q}) \in MT} mt(\vec{p},\vec{q})
\end{aligned}$$

An algorithm for computing MAT $T^E_{\varphi_{MT}}$ for the mapping assertion $\mathfrak{m} : \forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \overset{MT}{\leadsto} E(\vec{x}))$ is shown in Figure 1. $BLM2MAT(\mathfrak{m})$ initially uses the function $TGD2SMT(\sigma)$ to create a MAT $\mathcal{T}^E_{\varphi} = \langle\mathcal{T}_{\varphi}, \mathcal{T}_E\rangle$ for the mapping assertion $\sigma : \forall\vec{x}(\exists\vec{y}\varphi(\vec{x},\vec{y}) \leadsto E(\vec{x}))$ (without considering the mapping tables). Then it augments $\mathcal{T}_{\varphi}$ by adding each mapping table $mt \in MT$ to the rows of $\mathcal{T}_{\varphi}$. When a row is added to $\mathcal{T}_{\varphi}$ for a mapping table $mt(\vec{p},\vec{q}) \in MT$, the following steps are taken:

(1) A new row is added to $\mathcal{T}_{\varphi}$ with the tag mt.

(2) New columns are added to $\mathcal{T}_{\varphi}$ for the attributes $\vec{q}$.

(3) For some row R, if $T_{\varphi}[R][\vec{p}]$ contains some distinguished variable $\vec{a_i}$, then $T_{\varphi}[R][\vec{q}]$ is assigned $\vec{a_i}$.

(4) The columns $\vec{q}$ of $\mathcal{T}_{\varphi}$ are unified by putting the same same non-distinguished variables for each row of $\mathcal{T}_{\varphi}$.

When all the modification is done to $\mathcal{T}_{\varphi}$, we call the modified $\mathcal{T}_{\varphi}$, $\mathcal{T}_{\varphi_{MT}}$. Finally, $\mathcal{T}_{\varphi_{MT}}$ is merged with $\mathcal{T}_E$ to form the MAT $\mathcal{T}^E_{\varphi_{MT}} = \langle\mathcal{T}_{\varphi_{MT}}, \mathcal{T}_E\rangle$.

EXAMPLE 1. *Consider the P2P data sharing system* $\Pi =< \mathcal{P}, \mathcal{M} >$ *with the following settings:*
$\mathcal{P} = \{P_1, P_2\}$
$\mathcal{M} = \{\mathfrak{m}\}$
$\mathfrak{m} \equiv \forall x\exists y\forall z(P2E(x,y) \wedge P2J(y,z) \overset{MT}{\leadsto} E(x,z))$
$MT = \{mt(Job\_Description, Position)\}$

When the algorithm $BLM2MAT()$ of Figure 1 is applied to m, *initially it creates the MAT $\mathcal{T}_\varphi^\psi = \langle \mathcal{T}_\varphi, \mathcal{T}_\psi \rangle$ of Figure 2(a) using the algorithm $MA2MAT()$ of Figure 1 on the mapping assertion $\sigma \equiv \forall x \exists y \forall z (P2E(x, y) \wedge P2J(y, z) \rightsquigarrow E(x, z))$. $\mathcal{T}_\varphi$ is then modified step by step for the mapping tables mt. Figure 2(b) shows the new $\mathcal{T}_\varphi^\psi$ after a new row mt and a new column $Position$ is added to $\mathcal{T}_\varphi$. Figure 2(c) shows that a fresh non-distinguished variable b_2 has been assigned to $\mathcal{T}_\varphi[mt][Job\_Description]$. Figure 2(d) depicts that distinguished variable a_2 of $\mathcal{T}_\varphi[mt][Job\_Description]$ is copied to $\mathcal{T}_\varphi[mt][Position]$. Finally, all the symbols of the column $Job\_Descrion$ of $\mathcal{T}_\varphi$ is replaced by the same variable b_2 as shown in Figure 2(e).* □

PROPOSITION 1. *Let* $\mathrm{m} \equiv \forall \vec{x}(\exists \vec{y} \varphi(\vec{x}, \vec{y}) \overset{MT}{\rightsquigarrow} E(\vec{x}))$ *be a bi-level mapping. $\mathcal{T}_{\varphi_{MT}}^E = \langle \mathcal{T}_{\varphi_{MT}}, \mathcal{T}_E \rangle$ be the schema mapping tableaux (SMT) for* m. *Let I be an instance of the relations of φ. It is valid that*

$$\forall_{I'}((I' \in \mathbb{CI}(\mathrm{m}, I)) \Rightarrow (\mathcal{T}_{\varphi_{MT}}(I \bowtie MT) \equiv \mathcal{T}_E(I')))$$

Now we describe the mapping compostion process.
Let $M_{A \to B}$ is the mapping between A and B, and $M_{B \to C}$ be the mapping between B and C. The target is to find a mapping $M_{A \to C}$ between A and C that is equivalent to the composition of the two mappings $M_{A \to B}$ and $M_{B \to C}$. The following example shows a composition of two mappings.

Consider three data sources S_1, S_2 and S_3 with the following mappings among them.

(1) $M_{3 \to 2}$: $\pi_{C_{311}, C_{322}}(R_{31} \bowtie_{C_{312}=C_{321}} R_{32}) \overset{\{mt_{32}\}}{\rightsquigarrow} R_{21}$ is the mapping between S_3 and S_2:

(2) $M_{2 \to 1}$: $R_{21} \overset{mt_{21}}{\rightsquigarrow} R_{11}$ is the mapping between S_1 and S_2:.

From the above setting, an indirect mapping $M_{3 \to 1}$ between S_3 and S_1 can be constitueted as follows:

(3) $M_{3 \to 1} : \pi_{C_{311}, C_{322}}(R_{31} \bowtie_{C_{312}=C_{321}} R_{32}) \overset{\{mt_{32} \bowtie mt_{21}\}}{\rightsquigarrow} R_{11}$ is the composed mapping between S_1 and S_2.
According to the algorithm in Figure 1, tableau T_{S21} from M_{21} and T_{S32} from M_{S21} are produced. After generating the tableau the following steps are considered to generate the tableau T_{31} from T_{S21} and T_{S32}. Finally, mapping M_{31} is generated from tableau T_{31}.

(1) Merge T_{S32} and T_{S21} and generate initial tableau T_{S31}^1. In the merging process all the Rows, Columns, Tags, and Summaries of T_{S32} and T_{S21} are transferred in T_{S31}^1.
(2) Identify a common relation R_{21} between T_{T32} and T_{S21}.
(3) Create a link between the mapping tables of S_2 to the relations of S_3. The process has following two steps:
  —For each column C in the common relation, if $(C, C') \in \theta_{32}$ where C' is a column of a arbitrary relation R of S_3, the value of $T_{S31}^1.Rows[R][C']$ is updated by the value of $T_{S31}^1.Rows[R_{21}][C]$.

  —If $T_{S31}^1.Rows[R_{21}][C]$ be a distinguished variable then $T_{S31}^1.Summary[C']$ is also updated by the value $T_{S31}^1.Rows[R_{21}][C]$. After taking similar action for every columns of R_{21} in T_{S31}^1, T_{S31}^1 takes the shape of T_{S31}^2.
(4) Eliminate the common relation R_{21} from T_{S31}^2 using the following steps:
  —The columns having non-null values only in the row with R_{21} are deleted.
  —Delete the row with tag R_{21} from T_{S31}^2.

  After eliminating the R_{21}, T_{S31}^2 becomes T_{S31}^3.
(5) Remove the redundant columns of T_{S31}^3 (i.e. the columns with the same name).

Now we describe the steps for computing of target tableau, T_{T31}, and the homomorphism, θ_{31} :

(1) Create a tableau T_{T31} and initialize with the values of T_{T21}.
(2) Create an empty homomorphism θ_{31}.
(3) For each column C of T_{T21}, do the following steps:
  —Identify a column B that is mapped to C by the homomorphism θ_{21}.
  —If B has an entry in $T_{S31}.Summary$ then (C, B) and $(T_{T21}.Summary[C]\ , T_{S31}.Summary[B])$ pairs are added to θ_{31}.
  —If B column is Null in $T_{S31}.Summary$ then a column A is identified that is mapped to B by the homomorphism θ_{32}. If A column has an entry in $T_{S31}.Summary$ then (C, A) and $(T_{T21}.Summary[C], T_{S31}.Summary[A])$ pairs are added to θ_{31}.
  —If neither B nor A can be found in $T_{S31}.Summary$ for the column C, C is deleted from T_{T31}.

Now T_{S31}, T_{T31} and θ_{31} forms a triple and are then converted as normal bi-level mapping expression. In our example, when the triple $< T_{S31}, T_{T31}, \theta_{31} >$ is converted to normal expression, the mapping becomes as follows:

$$M_{3 \to 1} : \pi_{C_{311}, C_{112}}(R_{31} \bowtie_{C_{312}=C_{321}} R_{32}) \overset{\{mt_{32}, mt_{21}\}}{\rightsquigarrow} R_{11}$$

5. DSICUSSION AND CONCLUSION

This paper presented a model of data sharing system and proposed mapping semantics, called bi-level mapping and its composition, combining the schema-level and the data-level mappings necessary for resolving heterogeneity among data sources in a data sharing system. This bi-level mappings allow sources efficient data sharing facilities that sources miss considering only schema or data level mappings. The bi-level mapping is based on the tableau representation. The paper also provided an algorithm for the composition of two bi-level mappings.
To the best of our knowledge there is no implementation of combined scheme mappings in XML and unstructured database. This study is essential to understand and answer various fundamental issues and questions in regards to the suitability of bi-level mappings. There is also need to investigate the composition of mappings considering the dynamic behavior of sources.
The development of algorithms that implement our proposals would seem to be a sensible and natural follow-on. Constructing prototypes may not straightforward and could be time consuming. Therefore, our future goal is to investigate the composition of mappings considering XML databases, the dynamic behavior of peers. Further, we are interested to evaluate the whole process in a large data sharing system.

6. REFERENCES

[1] Lenzerini, M.: Data Integration: A Theoretical Perspective. In PODS, pp 233-246, (2002).
[2] Fagin, R., Kolaitis, P.G., Miller R.J.: Data Exchange: Semantics and Query Answering. In FKMP, (2003).

	$Name$	Jid	$Job\_Description$	$Name$	$Position$	
$P2E$	a_1	b_1		a_1	a_2	E
$P2J$		b_1	a_2			

(a) Initial SMT $\mathcal{T}_{\varphi}^{\psi} = \langle \mathcal{T}_{\varphi}, \mathcal{T}_{\psi} \rangle$ excluding the mapping tables

	$Name$	Jid	$Job\_Description$	$Position$	$Name$	$Position$	
$P2E$	a_1	b_1			a_1	a_2	E
$P2J$		b_1	a_2				
mt							

(b) New rows and columns added to $\mathcal{T}_{\varphi}$

	$Name$	Jid	$Job\_Description$	$Position$	$Name$	$Position$	
$P2E$	a_1	b_1			a_1	a_2	E
$P2J$		b_1	a_2				
mt			b_2				

(c) Fresh non-distinguished symbol added to the left column of mt

	$Name$	Jid	$Job\_Description$	$Position$	$Name$	$Position$	
$P2E$	a_1	b_1			a_1	a_2	E
$P2J$		b_1	a_2				
mt			b_2	a_2			

(d) Distinguished symbol is copied to the right column of mt

	$Name$	Jid	$Job\_Description$	$Position$	$Name$	$Position$	
$P2E$	a_1	b_1			a_1	a_2	E
$P2J$		b_1	b_2				
mt			b_2	a_2			

(e) Final SMT after symbols in the left column of mt being unified

Fig. 2. An example showing the steps for converting a Bi-Level Mapping to a SMT

[3] Arenas,M., Kantere, V., Kementsietsidis, A., Kiringa, I., Miller, R.J., Mylopoulos, J.: The Hyperion Project: From Data Integration to Data Coordination. In: SIGMOD RECORD (2003)

[4] Chawathe, S., Garcia-Molina, H., Hammer,J., Ireland, K., Papakonstantinou, Y., Ullman, J. and Widom, J.: The TSIMMIS Project: Integration of Heterogeneous Information Sources. In: IPSJ (1994)

[5] Ullman, J.D.: Information Integration Using Logical Views. In: ICDT(1997)

[6] Boyd M., Kittivoravitkul, S., Lazanitis, C., McBrien, P.J., Rizopoulos, N.: AutoMed: A BAV Data Integration System for Heterogeneous Data Sources. In: CAiSE (2004)

[7] Miller, R.J., Hernndez, M., Haas, L.M., Yan, L., Howard, C.T., Fagin R., Popa, L.: The Clio Project: Managing Heterogeneity. SIGMOD Record(2001)

[8] Levy, A.Y., Rajaraman A., Ordille, J.J.: Querying Heterogeneous Information Sources Using Source Descriptions. In: VLDB (1996)

[9] Kemensietsidis, A., Arenas,M.: Mapping Data in Peer-to-Peer Systems: Semantics and Algorithmic Issues. In: ACM SIGMOD (2003)

[10] M. A. Rahman, M. M. Masud, I. Kiringa, and A. El Saddik. Bi-Level Mapping: Combining Schema and Data Level Heterogeneity in Peer Data Sharing Systems. *Proc. Alberto Mendelzon Workshop on Foundations of Data Management*, Vol. 450, 2009.

[11] M. A. Rahman, M. M. Masud, I. Kiringa, and A. El Saddik. A Peer Data Sharing System Combining Schema and Data Level Mappings. In *International Journal of Semantic Computing*, Vol 3, No. 1, pp 105-129, 2009.

[12] Aho, A.V., Sagiv, Y., Ulman, J.D.,: Efficient Optimization of a Class of Relational Expressions. ACM Transactions on Dtabase Systems, Vol. 4, No. 4, pp 435-454, (1979)

[13] M. A. Rahman, M. M. Masud, I. Kiringa, and A. El Saddik. Tableaux-based optimization of schema mappings for data integration. In *Journal of Intelligent Information Systems*, Vol 38, Issue 2, pp 533-554, 2012.

[14] Arenas,M., Kantere, V., Kementsietsidis, A., Kiringa, I., Miller, R.J., Mylopoulos, J.: The Hyperion Project: From Data Integration to Data Coordination. In: SIGMOD RECORD (2003)

[15] P. Rodriguez-Gianolli, M. Garzetti, L. Jiang, A. Kementsietsidis, I. Kiringa, M. Masud, R. Miller, and J. Mylopoulos: Data Sharing in the Hyperion Peer Database System, In Proc. Of the Intl Conf. on Very Large Data Bases (VLDB), pp. 1291-1294, 2005.

[16] Halevy, A.Y., Ives Z.G., Madhavan, J., Mork, P., Suciu, D., Tatarinov, I.: The Piazza Peer Data Management System. IEEE T.K.D.E. 16, 787–798 (2004)

[17] O. Unal and H. Afsarmanesh. Semi-automated schema integration with SASMINT. In *Journal of Knowledge and Information System Journal*, Vol. 23, No. 1, pp. 99-128, 2010.