Union and Intersection of Schema Mappings

Jorge Pérez   Reinhard Pichler
Emanuel Sallinger   Vadim Savenkov

Universidad de Chile, TU Vienna

Schema mappings are essential to perform several data management tasks.

- Schema mapping: specification that describes the relationship between schemas.

Schema mappings are essential to perform several data management tasks.

- Schema mapping: specification that describes the relationship between schemas.

CliA:

name	balance	city

CliB:

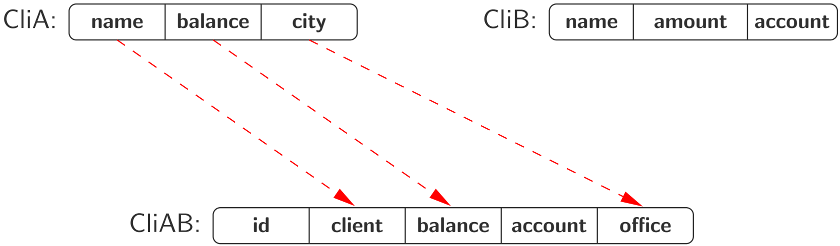
name	amount	account

CliAB:

id	client	balance	account	office

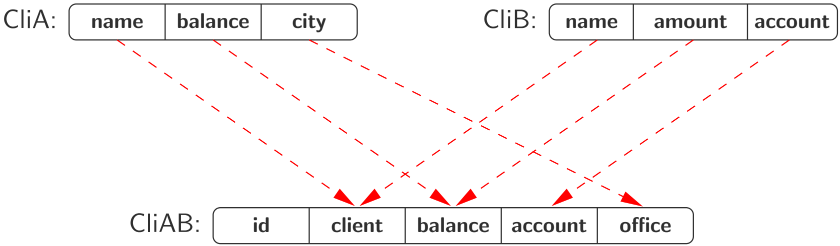
Schema mappings are essential to perform several data management tasks.

- Schema mapping: specification that describes the relationship between schemas.



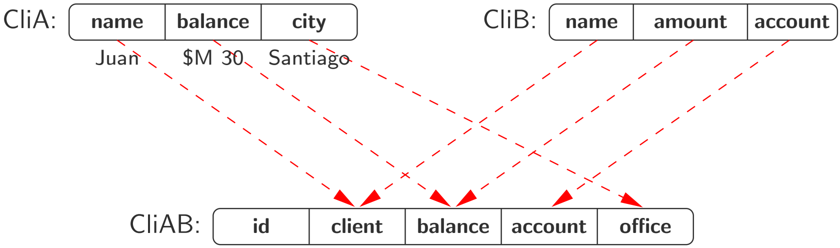
Schema mappings are essential to perform several data management tasks.

- Schema mapping: specification that describes the relationship between schemas.



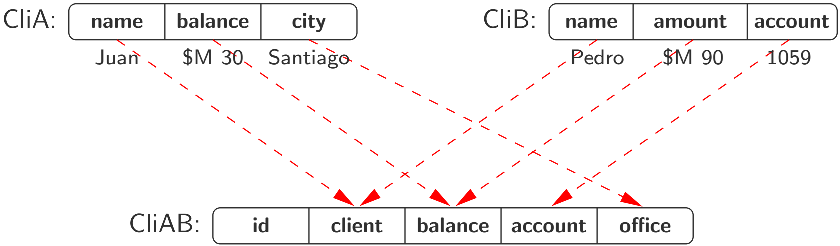
Schema mappings are essential to perform several data management tasks.

- Schema mapping: specification that describes the relationship between schemas.



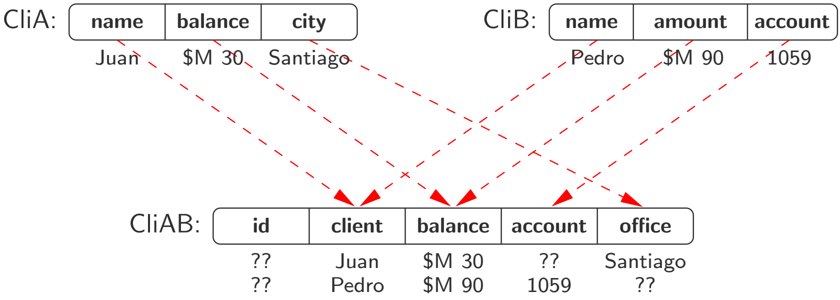
Schema mappings are essential to perform several data management tasks.

- Schema mapping: specification that describes the relationship between schemas.



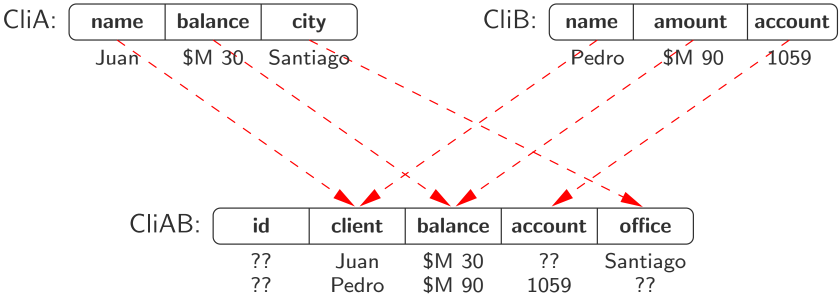
Schema mappings are essential to perform several data management tasks.

- Schema mapping: specification that describes the relationship between schemas.



Schema mappings are essential to perform several data management tasks.

- Schema mapping: specification that describes the relationship between schemas.



Schema mappings contain metadata

Schema mappings are usually given in the form of logical specifications

- Two database schemas: *source* and *target*
- schema mapping: logical specification over the schemas

Schema mappings are usually given in the form of logical specifications

- Two database schemas: *source* and *target*
- schema mapping: logical specification over the schemas

Example:

Source: $\{\text{ClientA}(\textbf{name}, \textbf{balance}, \textbf{city})\}$

Target: $\{\text{ClientAB}(\textbf{id}, \textbf{client}, \textbf{balance}, \textbf{account}, \textbf{office})\}$

$$\text{ClientA}(x, y, z) \;\rightarrow\; \exists U \exists V\ \text{ClientAB}(U, x, y, V, z)$$

Schema mappings are usually given in the form of logical specifications

- Two database schemas: *source* and *target*
- schema mapping: logical specification over the schemas

Example:

Source: {ClientA(**name**, **balance**, **city**) }

Target: {ClientAB(**id**, **client**, **balance**, **account**, **office**) }

$$\mathrm{ClientA}(x, y, z) \;\rightarrow\; \exists U \exists V \; \mathrm{ClientAB}(U, x, y, V, z)$$ *st-tgds*

Schema mappings are usually given in the form of logical specifications

- Two database schemas: *source* and *target*
- schema mapping: logical specification over the schemas

Example:

Source: {ClientA(**name**, **balance**, **city**) }

Target: {ClientAB(**id**, **client**, **balance**, **account**, **office**) }

$\text{ClientA}(x, y, z) \;\rightarrow\; \exists U \exists V\ \text{ClientAB}(U, x, y, V, z)$ *st-tgds*

- Formally a mapping is just a binary relation:
  the set of pairs (I, J) that satisfy the logical specification

$$\mathcal{M} = \{(I, J) \mid (I, J) \models \Sigma\}$$

In several scenarios we need to reuse the metadata in the schema mappings

Operators on schema mappings

- Composition [MH03, Melnik04, FKPT04]
- Inversion [Fagin06, FKPT07, APR08, APRR09, FN10]
- Extract, Merge [Melnik04, BM06, APRR10]

Plus some general notions for schema mappings:

- Information transfer [FKPT09, APRR10]
- Redundancy [APRR10]

In several scenarios we need to reuse the metadata in the schema mappings

Operators on schema mappings

- Composition [MH03, Melnik04, FKPT04]
- Inversion [Fagin06, FKPT07, APR08, APRR09, FN10]
- Extract, Merge [Melnik04, BM06, APRR10]

Plus some general notions for schema mappings:

- Information transfer [FKPT09, APRR10]
- Redundancy [APRR10]

We propose two new operators, *Union* and *Intersection*, using the *information transfer* notion.

Outline

Intuition and definition

Existence of union and intersection

Concluding remarks and future work

Outline

Intuition and definition

Existence of union and intersection

Concluding remarks and future work

Information transferred by a mapping

Source: {Emp(**name**, **lives_in**, **works_in**) }

Target_1: {Person(**ssn**, **name**) }

Information transferred by a mapping

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {Person(**ssn**, **name**) }

$$\mathcal{M}_1: \qquad \text{Emp}(x, y, z) \quad \rightarrow \quad \exists u \; \text{Person}(u, x)$$

Information transferred by a mapping

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {Person(**ssn**, **name**) }
Target$_2$: {WorksIn(**name**, **place**) }

$$\mathcal{M}_1: \qquad \text{Emp}(x, y, z) \quad \rightarrow \quad \exists u \ \text{Person}(u, x)$$

Information transferred by a mapping

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {Person(**ssn**, **name**) }
Target$_2$: {WorksIn(**name**, **place**) }

$$
\begin{array}{llcl}
\mathcal{M}_1: & \text{Emp}(x, y, z) & \rightarrow & \exists u \ \text{Person}(u, x) \\
\mathcal{M}_2: & \text{Emp}(x, y, z) & \rightarrow & \text{WorksIn}(x, z)
\end{array}
$$

Information transferred by a mapping

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {Person(**ssn**, **name**) }
Target$_2$: {WorksIn(**name**, **place**) }

$$\begin{array}{lccc} \mathcal{M}_1: & \text{Emp}(x,y,z) & \rightarrow & \exists u \ \text{Person}(u,x) \\ \mathcal{M}_2: & \text{Emp}(x,y,z) & \rightarrow & \text{WorksIn}(x,z) \end{array}$$

Intuitively:

$\mathcal{M}_2$ is *more informative* than $\mathcal{M}_1$

Information transferred by a mapping

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {Person(**ssn**, **name**) }
Target$_2$: {WorksIn(**name**, **place**) }
Target$_3$: {LivesIn(**name**, **place**) }

$$\mathcal{M}_1: \quad \text{Emp}(x, y, z) \rightarrow \exists u \ \text{Person}(u, x)$$
$$\mathcal{M}_2: \quad \text{Emp}(x, y, z) \rightarrow \text{WorksIn}(x, z)$$

Intuitively:

$\mathcal{M}_2$ is *more informative* than $\mathcal{M}_1$

Information transferred by a mapping

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {Person(**ssn**, **name**) }
Target$_2$: {WorksIn(**name**, **place**) }
Target$_3$: {LivesIn(**name**, **place**) }

$$\begin{array}{llcl} \mathcal{M}_1: & \text{Emp}(x,y,z) & \rightarrow & \exists u \ \text{Person}(u,x) \\ \mathcal{M}_2: & \text{Emp}(x,y,z) & \rightarrow & \text{WorksIn}(x,z) \\ \mathcal{M}_3: & \text{Emp}(x,y,z) & \rightarrow & \text{LivesIn}(x,y) \end{array}$$

Intuitively:

$\mathcal{M}_2$ is *more informative* than $\mathcal{M}_1$

Information transferred by a mapping

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {Person(**ssn**, **name**) }
Target$_2$: {WorksIn(**name**, **place**) }
Target$_3$: {LivesIn(**name**, **place**) }

$$
\begin{array}{llcl}
\mathcal{M}_1: & \text{Emp}(x, y, z) & \rightarrow & \exists u \ \text{Person}(u, x) \\
\mathcal{M}_2: & \text{Emp}(x, y, z) & \rightarrow & \text{WorksIn}(x, z) \\
\mathcal{M}_3: & \text{Emp}(x, y, z) & \rightarrow & \text{LivesIn}(x, y)
\end{array}
$$

Intuitively:

$\mathcal{M}_2$ is *more informative* than $\mathcal{M}_1$
$\mathcal{M}_3$ is *more informative* than $\mathcal{M}_1$

Information transferred by a mapping

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {Person(**ssn**, **name**) }
Target$_2$: {WorksIn(**name**, **place**) }
Target$_3$: {LivesIn(**name**, **place**) }

$$\begin{array}{llcl}
\mathcal{M}_1: & \text{Emp}(x, y, z) & \rightarrow & \exists u \ \text{Person}(u, x) \\
\mathcal{M}_2: & \text{Emp}(x, y, z) & \rightarrow & \text{WorksIn}(x, z) \\
\mathcal{M}_3: & \text{Emp}(x, y, z) & \rightarrow & \text{LivesIn}(x, y)
\end{array}$$

Intuitively:

$\mathcal{M}_2$ is *more informative* than $\mathcal{M}_1$
$\mathcal{M}_3$ is *more informative* than $\mathcal{M}_1$
$\mathcal{M}_2$ and $\mathcal{M}_3$ are *incomparable*

Information transferred by a mapping

Assume that $\mathcal{M}_1$ and $\mathcal{M}_2$ share the source schema.

Definition (APRR10)

$\mathcal{M}_2$ is *more (or equally) informative than* $\mathcal{M}_1$, denoted by

$$\mathcal{M}_1 \preceq \mathcal{M}_2,$$

Information transferred by a mapping

Assume that $\mathcal{M}_1$ and $\mathcal{M}_2$ share the source schema.

Definition (APRR10)

$\mathcal{M}_2$ is *more (or equally) informative than* $\mathcal{M}_1$, denoted by

$$\mathcal{M}_1 \preceq \mathcal{M}_2,$$

if there exists a mapping $\mathcal{M}'$ such that $\mathcal{M}_2 \circ \mathcal{M}' = \mathcal{M}_1$.

Information transferred by a mapping

Assume that $\mathcal{M}_1$ and $\mathcal{M}_2$ share the source schema.

Definition (APRR10)

$\mathcal{M}_2$ is *more (or equally) informative than* $\mathcal{M}_1$, denoted by

$$\mathcal{M}_1 \preceq \mathcal{M}_2,$$

if there exists a mapping $\mathcal{M}'$ such that $\mathcal{M}_2 \circ \mathcal{M}' = \mathcal{M}_1$.

Idea: $\mathcal{M}_2$ transfers information enough to *reconstruct* $\mathcal{M}_1$

Information transferred by a mapping

Assume that $\mathcal{M}_1$ and $\mathcal{M}_2$ share the source schema.

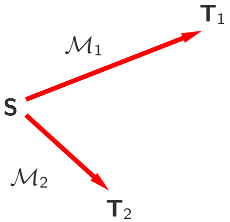
Definition (APRR10)

$\mathcal{M}_2$ is *more (or equally) informative than* $\mathcal{M}_1$, denoted by

$$\mathcal{M}_1 \preceq \mathcal{M}_2,$$

if there exists a mapping $\mathcal{M}'$ such that $\mathcal{M}_2 \circ \mathcal{M}' = \mathcal{M}_1$.

Idea: $\mathcal{M}_2$ transfers information enough to *reconstruct* $\mathcal{M}_1$



Information transferred by a mapping

Assume that $\mathcal{M}_1$ and $\mathcal{M}_2$ share the source schema.

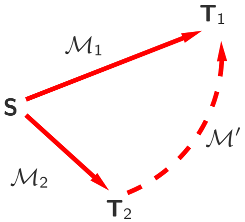
Definition (APRR10)

$\mathcal{M}_2$ is *more (or equally) informative than* $\mathcal{M}_1$, denoted by

$$\mathcal{M}_1 \preceq \mathcal{M}_2,$$

if there exists a mapping $\mathcal{M}'$ such that $\mathcal{M}_2 \circ \mathcal{M}' = \mathcal{M}_1$.

Idea: $\mathcal{M}_2$ transfers information enough to *reconstruct* $\mathcal{M}_1$



Information transferred by a mapping

Assume that $\mathcal{M}_1$ and $\mathcal{M}_2$ share the source schema.

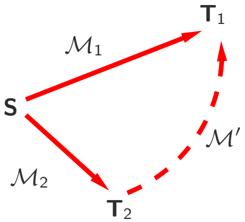
Definition (APRR10)

$\mathcal{M}_2$ is *more (or equally) informative than* $\mathcal{M}_1$, denoted by

$$\mathcal{M}_1 \preceq \mathcal{M}_2,$$

if there exists a mapping $\mathcal{M}'$ such that $\mathcal{M}_2 \circ \mathcal{M}' = \mathcal{M}_1$.

Idea: $\mathcal{M}_2$ transfers information enough to *reconstruct* $\mathcal{M}_1$



$\mathcal{M}_1 : \quad \text{Emp}(x, y, z) \rightarrow \exists u \ \text{Person}(u, x)$

Information transferred by a mapping

Assume that $\mathcal{M}_1$ and $\mathcal{M}_2$ share the source schema.

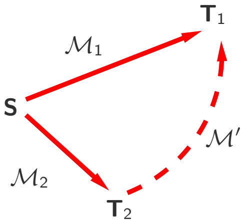
Definition (APRR10)

$\mathcal{M}_2$ is *more (or equally) informative than* $\mathcal{M}_1$, denoted by

$$\mathcal{M}_1 \preceq \mathcal{M}_2,$$

if there exists a mapping $\mathcal{M}'$ such that $\mathcal{M}_2 \circ \mathcal{M}' = \mathcal{M}_1$.

Idea: $\mathcal{M}_2$ transfers information enough to *reconstruct* $\mathcal{M}_1$



$\mathcal{M}_1$: $\text{Emp}(x, y, z) \rightarrow \exists u \ \text{Person}(u, x)$

$\mathcal{M}_2$: $\text{Emp}(x, y, z) \rightarrow \text{WorksIn}(x, z)$

Information transferred by a mapping

Assume that $\mathcal{M}_1$ and $\mathcal{M}_2$ share the source schema.

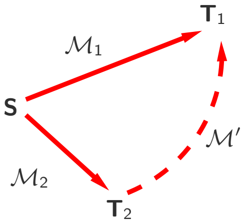
Definition (APRR10)

$\mathcal{M}_2$ is *more (or equally) informative than* $\mathcal{M}_1$, denoted by

$$\mathcal{M}_1 \preceq \mathcal{M}_2,$$

if there exists a mapping $\mathcal{M}'$ such that $\mathcal{M}_2 \circ \mathcal{M}' = \mathcal{M}_1$.

Idea: $\mathcal{M}_2$ transfers information enough to *reconstruct* $\mathcal{M}_1$



$\mathcal{M}_1 : \quad \text{Emp}(x, y, z) \rightarrow \exists u \ \text{Person}(u, x)$

$\mathcal{M}_2 : \quad \text{Emp}(x, y, z) \rightarrow \text{WorksIn}(x, z)$

$\mathcal{M}' : \quad \text{WorksIn}(x, z) \rightarrow \exists u \ \text{Person}(u, x)$

Information transferred by a mapping

Assume that $\mathcal{M}_1$ and $\mathcal{M}_2$ share the source schema.

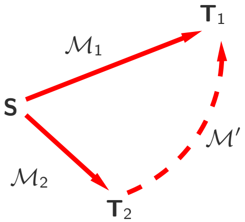
Definition (APRR10)

$\mathcal{M}_2$ is *more (or equally) informative than* $\mathcal{M}_1$, denoted by

$$\mathcal{M}_1 \preceq \mathcal{M}_2,$$

if there exists a mapping $\mathcal{M}'$ such that $\mathcal{M}_2 \circ \mathcal{M}' = \mathcal{M}_1$.

Idea: $\mathcal{M}_2$ transfers information enough to *reconstruct* $\mathcal{M}_1$



$$\mathcal{M}_1 : \quad \text{Emp}(x, y, z) \rightarrow \exists u \ \text{Person}(u, x)$$
$$\mathcal{M}_2 : \quad \text{Emp}(x, y, z) \rightarrow \text{WorksIn}(x, z)$$

$$\mathcal{M}' : \quad \text{WorksIn}(x, z) \rightarrow \exists u \ \text{Person}(u, x)$$

$$\mathcal{M}_2 \circ \mathcal{M}' = \mathcal{M}_1 \quad \Longrightarrow \quad \mathcal{M}_1 \preceq \mathcal{M}_2$$

Intuition of Intersection

Consider two mappings generated by mapping-generating tools [FKPT09] propose to chose the *more informative*

Intuition of Intersection

Consider two mappings generated by mapping-generating tools [FKPT09] propose to chose the *more informative*

What if they are incomparable in terms of information transfer?

$$\begin{array}{llcl} \mathcal{M}_2: & \mathsf{Emp}(x,y,z) & \rightarrow & \mathsf{WorksIn}(x,z) \\ \mathcal{M}_3: & \mathsf{Emp}(x,y,z) & \rightarrow & \mathsf{LivesIn}(x,y) \end{array}$$

Conservative approach:

- synthesize a new mapping that transfers only the *shared* information that is being transferred by both $\mathcal{M}_2$ and $\mathcal{M}_3$

Intuition of Intersection

Consider two mappings generated by mapping-generating tools [FKPT09] propose to chose the *more informative*

What if they are incomparable in terms of information transfer?

$$\mathcal{M}_2: \quad \mathsf{Emp}(x,y,z) \rightarrow \mathsf{WorksIn}(x,z)$$
$$\mathcal{M}_3: \quad \mathsf{Emp}(x,y,z) \rightarrow \mathsf{LivesIn}(x,y)$$

Conservative approach:

- synthesize a new mapping that transfers only the *shared* information that is being transferred by both $\mathcal{M}_2$ and $\mathcal{M}_3$

An intersection mapping

Intuition of Intersection

Consider two mappings generated by mapping-generating tools [FKPT09] propose to chose the *more informative*

What if they are incomparable in terms of information transfer?

$$\mathcal{M}_2: \quad \text{Emp}(x, y, z) \rightarrow \text{WorksIn}(x, z)$$
$$\mathcal{M}_3: \quad \text{Emp}(x, y, z) \rightarrow \text{LivesIn}(x, y)$$

Conservative approach:

- synthesize a new mapping that transfers only the *shared* information that is being transferred by both $\mathcal{M}_2$ and $\mathcal{M}_3$

An intersection mapping

$$\mathcal{M}: \quad \text{Emp}(x, y, z) \rightarrow \text{ENames}(x)$$

Defining intersection

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *intersection* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

Defining intersection

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *intersection* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

1. $\mathcal{M} \preceq \mathcal{M}_1$ and $\mathcal{M} \preceq \mathcal{M}_2$

Defining intersection

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *intersection* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

1. $\mathcal{M} \preceq \mathcal{M}_1$ and $\mathcal{M} \preceq \mathcal{M}_2$
2. if $\mathcal{N} \in \mathcal{C}$, with $\mathcal{N} \preceq \mathcal{M}_1$ and $\mathcal{N} \preceq \mathcal{M}_2$, then $\mathcal{N} \preceq \mathcal{M}$

Defining intersection

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *intersection* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

1. $\mathcal{M} \preceq \mathcal{M}_1$ and $\mathcal{M} \preceq \mathcal{M}_2$
2. if $\mathcal{N} \in \mathcal{C}$, with $\mathcal{N} \preceq \mathcal{M}_1$ and $\mathcal{N} \preceq \mathcal{M}_2$, then $\mathcal{N} \preceq \mathcal{M}$

i.e. the *greatest lower bound*

Defining intersection

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *intersection* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

1. $\mathcal{M} \preceq \mathcal{M}_1$ and $\mathcal{M} \preceq \mathcal{M}_2$
2. if $\mathcal{N} \in \mathcal{C}$, with $\mathcal{N} \preceq \mathcal{M}_1$ and $\mathcal{N} \preceq \mathcal{M}_2$, then $\mathcal{N} \preceq \mathcal{M}$

i.e. the *greatest lower bound*

We denote the intersection by $\mathcal{M}_1 \sqcap_{\mathcal{C}} \mathcal{M}_2$

Defining intersection

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *intersection* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

1. $\mathcal{M} \preceq \mathcal{M}_1$ and $\mathcal{M} \preceq \mathcal{M}_2$
2. if $\mathcal{N} \in \mathcal{C}$, with $\mathcal{N} \preceq \mathcal{M}_1$ and $\mathcal{N} \preceq \mathcal{M}_2$, then $\mathcal{N} \preceq \mathcal{M}$

i.e. the *greatest lower bound*

We denote the intersection by $\mathcal{M}_1 \sqcap_{\mathcal{C}} \mathcal{M}_2$

Research questions:

- does an intersection always exist?
- what is the expressiveness needed to specify it?
- can we effectively compute an intersection?

Union, the dual operator

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *union* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

Union, the dual operator

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *union* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

1. $\mathcal{M}_1 \preceq \mathcal{M}$ and $\mathcal{M}_2 \preceq \mathcal{M}$

Union, the dual operator

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *union* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

1. $\mathcal{M}_1 \preceq \mathcal{M}$ and $\mathcal{M}_2 \preceq \mathcal{M}$
2. if $\mathcal{N} \in \mathcal{C}$, with $\mathcal{M}_1 \preceq \mathcal{N}$ and $\mathcal{M}_2 \preceq \mathcal{N}$, then $\mathcal{M} \preceq \mathcal{N}$

Union, the dual operator

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *union* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

1. $\mathcal{M}_1 \preceq \mathcal{M}$ and $\mathcal{M}_2 \preceq \mathcal{M}$
2. if $\mathcal{N} \in \mathcal{C}$, with $\mathcal{M}_1 \preceq \mathcal{N}$ and $\mathcal{M}_2 \preceq \mathcal{N}$, then $\mathcal{M} \preceq \mathcal{N}$

i.e. *the least upper bound*

Union, the dual operator

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *union* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

1. $\mathcal{M}_1 \preceq \mathcal{M}$ and $\mathcal{M}_2 \preceq \mathcal{M}$
2. if $\mathcal{N} \in \mathcal{C}$, with $\mathcal{M}_1 \preceq \mathcal{N}$ and $\mathcal{M}_2 \preceq \mathcal{N}$, then $\mathcal{M} \preceq \mathcal{N}$

i.e. *the least upper bound*

We denote union by $\mathcal{M}_1 \sqcup_{\mathcal{C}} \mathcal{M}_2$

Union, the dual operator

Let $\mathcal{C}$ be a class of mappings, and $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}$.

Definition

The *union* of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ s.t.

1. $\mathcal{M}_1 \preceq \mathcal{M}$ and $\mathcal{M}_2 \preceq \mathcal{M}$
2. if $\mathcal{N} \in \mathcal{C}$, with $\mathcal{M}_1 \preceq \mathcal{N}$ and $\mathcal{M}_2 \preceq \mathcal{N}$, then $\mathcal{M} \preceq \mathcal{N}$

i.e. *the least upper bound*

We denote union by $\mathcal{M}_1 \sqcup_{\mathcal{C}} \mathcal{M}_2$

Intuition: the optimistic approach

Union transfers the *sum* of the info transferred by $\mathcal{M}_1$ and $\mathcal{M}_2$

for union, the same research questions apply

Outline

Intuition and definition

Existence of union and intersection

Concluding remarks and future work

$\preceq$, $\sqcap$, $\sqcup$ form a *lattice* of mappings

Theorem

There is a class of mappings $\mathcal{R}$ (that contains all the st-tgds) s.t.

$$\sqcap_{\mathcal{R}} \text{ and } \sqcup_{\mathcal{R}} \text{ always exists}$$

$\preceq$, $\sqcap$, $\sqcup$ form a *lattice* of mappings

Theorem

There is a class of mappings $\mathcal{R}$ (that contains all the st-tgds) s.t.

$$\sqcap_{\mathcal{R}} \text{ and } \sqcup_{\mathcal{R}} \text{ always exists}$$

Reformulating

$\preceq$, $\sqcap$, $\sqcup$ form a *lattice* of mappings

Theorem

There is a class of mappings $\mathcal{R}$ (that contains all the st-tgds) s.t.

$$\sqcap_{\mathcal{R}} \text{ and } \sqcup_{\mathcal{R}} \text{ always exists}$$

Reformulating

Theorem

There is a class of mappings $\mathcal{R}$ (that contains all the st-tgds) s.t.

$$(\mathcal{R}, \preceq) \text{ forms a lattice}$$

$\preceq$, $\sqcap$, $\sqcup$ form a *lattice* of mappings

Theorem

There is a class of mappings $\mathcal{R}$ (that contains all the st-tgds) s.t.

$$\sqcap_{\mathcal{R}} \text{ and } \sqcup_{\mathcal{R}} \text{ always exists}$$

Reformulating

Theorem

There is a class of mappings $\mathcal{R}$ (that contains all the st-tgds) s.t.

$$(\mathcal{R}, \preceq) \text{ forms a lattice}$$

$\mathcal{R}$ is the class REC of *recoverable mappings*: mappings that admit a *maximum recovery* (a notion of inverse mapping [APR08])

Another intuition based on query answering

Mapping $\mathcal{M}$, *target query* $Q_{\mathbf{T}}$, source instance I:

$$\underline{\text{certain}}_{\mathcal{M}}(Q_{\mathbf{T}}, I) = \bigcap_{(I,J) \in \mathcal{M}} Q_{\mathbf{T}}(J)$$

Another intuition based on query answering

Mapping $\mathcal{M}$, *target query* $Q_{\mathbf{T}}$, source instance I:

$$\underline{\text{certain}}_{\mathcal{M}}(Q_{\mathbf{T}}, I) = \bigcap_{(I,J)\in\mathcal{M}} Q_{\mathbf{T}}(J)$$

A source query $Q_{\mathbf{S}}$ is *target rewritable under* $\mathcal{M}$ if there exists a target query $Q_{\mathbf{T}}$ such that

$$Q_{\mathbf{S}}(I) = \underline{\text{certain}}_{\mathcal{M}}(Q_{\mathbf{T}}, I)$$

for every source instance I.

Another intuition based on query answering

Mapping $\mathcal{M}$, *target query* $Q_{\mathbf{T}}$, source instance I:

$$\underline{\text{certain}}_{\mathcal{M}}(Q_{\mathbf{T}}, I) = \bigcap_{(I,J)\in\mathcal{M}} Q_{\mathbf{T}}(J)$$

A source query $Q_{\mathbf{S}}$ is *target rewritable under* $\mathcal{M}$ if there exists a target query $Q_{\mathbf{T}}$ such that

$$Q_{\mathbf{S}}(I) = \underline{\text{certain}}_{\mathcal{M}}(Q_{\mathbf{T}}, I)$$

for every source instance I.

- Intuitively: if $Q_{\mathbf{S}}$ is target rewritable under $\mathcal{M}$, then $\mathcal{M}$ *transfers all the source data* retrieved by $Q_{\mathbf{S}}$.

Another intuition based on query answering

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be mappings in $\textsc{Rec}$

Another intuition based on query answering

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be mappings in $\textsc{Rec}$

Proposition

Q is target rewritable under $\mathcal{M}_1$ and $\mathcal{M}_2$
if and only if Q is target rewritable under $\mathcal{M}_1 \sqcap_{\textsc{Rec}} \mathcal{M}_2$

Another intuition based on query answering

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be mappings in $\textsc{Rec}$

Proposition

Q is target rewritable under $\mathcal{M}_1$ and *$\mathcal{M}_2$*
if and only if Q is target rewritable under $\mathcal{M}_1 \sqcap_{\textsc{Rec}} \mathcal{M}_2$

Proposition

Q is target rewritable under $\mathcal{M}_1$ or *$\mathcal{M}_2$*
if and only if Q is target rewritable under $\mathcal{M}_1 \sqcup_{\textsc{Rec}} \mathcal{M}_2$

Another intuition based on query answering

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be mappings in $\textsc{Rec}$

Proposition

Q is target rewritable under $\mathcal{M}_1$ and $\mathcal{M}_2$
if and only if Q is target rewritable under $\mathcal{M}_1 \sqcap_{\textsc{Rec}} \mathcal{M}_2$

Proposition

Q is target rewritable under $\mathcal{M}_1$ or $\mathcal{M}_2$
if and only if Q is target rewritable under $\mathcal{M}_1 \sqcup_{\textsc{Rec}} \mathcal{M}_2$

Open question

can we characterize $\sqcap_{\textsc{Rec}}$ and $\sqcup_{\textsc{Rec}}$ in terms of target rewritability?

Existence of Union

We know that $\sqcup_{\text{REC}}$ always exists, but what about st-tgds?

$$\begin{array}{llll} \mathcal{M}_2\text{:} & \mathsf{Emp}(x, y, z) & \rightarrow & \mathsf{WorksIn}(x, z) \\ \mathcal{M}_3\text{:} & \mathsf{Emp}(x, y, z) & \rightarrow & \mathsf{LivesIn}(x, y) \end{array}$$

Union:

$$\mathcal{M}\text{:} \quad \mathsf{Emp}(x, y, z) \rightarrow \mathsf{WorksIn}(x, z) \wedge \mathsf{LivesIn}(x, y)$$

Existence of Union

We know that $\sqcup_{\text{REC}}$ always exists, but what about st-tgds?

$$\begin{array}{llcl} \mathcal{M}_2: & \mathsf{Emp}(x,y,z) & \rightarrow & \mathsf{WorksIn}(x,z) \\ \mathcal{M}_3: & \mathsf{Emp}(x,y,z) & \rightarrow & \mathsf{LivesIn}(x,y) \end{array}$$

Union:

$$\mathcal{M}: \quad \mathsf{Emp}(x,y,z) \rightarrow \mathsf{WorksIn}(x,z) \wedge \mathsf{LivesIn}(x,y)$$

This simple idea always work for st-tgds

Proposition

The union $\sqcup_{\textit{st-tgds}}$ always exists

Existence of Union

We know that $\sqcup_{\mathrm{REC}}$ always exists, but what about st-tgds?

$$\begin{array}{lccc} \mathcal{M}_2: & \mathsf{Emp}(x,y,z) & \rightarrow & \mathsf{WorksIn}(x,z) \\ \mathcal{M}_3: & \mathsf{Emp}(x,y,z) & \rightarrow & \mathsf{LivesIn}(x,y) \end{array}$$

Union:

$$\mathcal{M}: \quad \mathsf{Emp}(x,y,z) \rightarrow \mathsf{WorksIn}(x,z) \wedge \mathsf{LivesIn}(x,y)$$

This simple idea always work for st-tgds

Proposition

The union $\sqcup_{\textit{st-tgds}}$ always exists

That is, union of st-tgds can be specified as st-tgds

Intersection of st-tgds cannot always be specified in FO

$$\begin{array}{llll} \mathcal{M}_1: & A(x,u) \wedge B(u,y) & \rightarrow & T_1(x,y) \\ \mathcal{M}_2: & B(y,v) \wedge A(v,x) & \rightarrow & T_2(y,x) \end{array}$$

Intersection of st-tgds cannot always be specified in FO

$$
\begin{array}{llcl}
\mathcal{M}_1: & A(x,u) \wedge B(u,y) & \rightarrow & T_1(x,y) \\
\mathcal{M}_2: & B(y,v) \wedge A(v,x) & \rightarrow & T_2(y,x)
\end{array}
$$

The following mapping $\mathcal{N}_n$ is less informative than $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$
A(u_1, v_1) \wedge B(v_1, u_1) \quad \rightarrow \quad \exists w \, P_1(w)
$$

Intersection of st-tgds cannot always be specified in FO

$$\begin{array}{llll} \mathcal{M}_1: & A(x,u) \wedge B(u,y) & \rightarrow & T_1(x,y) \\ \mathcal{M}_2: & B(y,v) \wedge A(v,x) & \rightarrow & T_2(y,x) \end{array}$$

The following mapping $\mathcal{N}_n$ is less informative than $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$\begin{array}{rcl} A(u_1,v_1) \wedge B(v_1,u_1) & \rightarrow & \exists w\ P_1(w) \\ A(u_1,v_1) \wedge B(v_1,u_2) \wedge A(u_2,v_2) \wedge B(v_2,u_1) & \rightarrow & \exists w\ P_2(w) \end{array}$$

Intersection of st-tgds cannot always be specified in FO

$$\begin{array}{lrcl} \mathcal{M}_1: & A(x,u) \wedge B(u,y) & \rightarrow & T_1(x,y) \\ \mathcal{M}_2: & B(y,v) \wedge A(v,x) & \rightarrow & T_2(y,x) \end{array}$$

The following mapping $\mathcal{N}_n$ is less informative than $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$\begin{array}{rcl} A(u_1,v_1) \wedge B(v_1,u_1) & \rightarrow & \exists w\ P_1(w) \\ A(u_1,v_1) \wedge B(v_1,u_2) \wedge A(u_2,v_2) \wedge B(v_2,u_1) & \rightarrow & \exists w\ P_2(w) \\ A(u_1,v_1) \wedge B(v_1,u_2) \wedge \cdots \wedge A(u_3,v_3) \wedge B(v_3,u_1) & \rightarrow & \exists w\ P_3(w) \end{array}$$

Intersection of st-tgds cannot always be specified in FO

$$\begin{array}{lrcl} \mathcal{M}_1: & A(x,u) \wedge B(u,y) & \rightarrow & T_1(x,y) \\ \mathcal{M}_2: & B(y,v) \wedge A(v,x) & \rightarrow & T_2(y,x) \end{array}$$

The following mapping $\mathcal{N}_n$ is less informative than $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$\begin{array}{rcl} A(u_1,v_1) \wedge B(v_1,u_1) & \rightarrow & \exists w\ P_1(w) \\ A(u_1,v_1) \wedge B(v_1,u_2) \wedge A(u_2,v_2) \wedge B(v_2,u_1) & \rightarrow & \exists w\ P_2(w) \\ A(u_1,v_1) \wedge B(v_1,u_2) \wedge \cdots \wedge A(u_3,v_3) \wedge B(v_3,u_1) & \rightarrow & \exists w\ P_3(w) \\ & \vdots & \\ \textit{if there exists an AB cycle of length } n & \rightarrow & \exists w\ P_n(w) \end{array}$$

Intersection of st-tgds cannot always be specified in FO

$$\begin{array}{llcl} \mathcal{M}_1: & A(x,u) \wedge B(u,y) & \rightarrow & T_1(x,y) \\ \mathcal{M}_2: & B(y,v) \wedge A(v,x) & \rightarrow & T_2(y,x) \end{array}$$

The following mapping $\mathcal{N}_n$ is less informative than $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$\begin{array}{rcl} A(u_1,v_1) \wedge B(v_1,u_1) & \rightarrow & \exists w\ P_1(w) \\ A(u_1,v_1) \wedge B(v_1,u_2) \wedge A(u_2,v_2) \wedge B(v_2,u_1) & \rightarrow & \exists w\ P_2(w) \\ A(u_1,v_1) \wedge B(v_1,u_2) \wedge \cdots \wedge A(u_3,v_3) \wedge B(v_3,u_1) & \rightarrow & \exists w\ P_3(w) \\ & \vdots & \\ \textit{if there exists an } AB \textit{ cycle of length } n & \rightarrow & \exists w\ P_n(w) \end{array}$$

Theorem

There is no mapping $\mathcal{M}'$ expressible in FO such that for every n

$$\mathcal{N}_n \preceq \mathcal{M}' \preceq \mathcal{M}_1 \quad \textit{and} \quad \mathcal{N}_n \preceq \mathcal{M}' \preceq \mathcal{M}_2$$

Existence of the intersection

We know that $\sqcap_{\mathrm{REC}}$ always exists but

Corollary

There are st-tgds mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ such that

$$\mathcal{M}_1 \sqcap_{\mathrm{REC}} \mathcal{M}_2 \textit{ cannot be specified in FO}$$

Existence of the intersection

We know that $\sqcap_{\mathrm{REC}}$ always exists but

Corollary

There are st-tgds mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ such that

$$\mathcal{M}_1 \sqcap_{\mathrm{REC}} \mathcal{M}_2 \text{ cannot be specified in FO}$$

Moreover

Corollary

The intersection $\sqcap_{\textit{st-tgds}}$ not always exists

Existence of the intersection

We know that $\sqcap_{\mathrm{REC}}$ always exists but

Corollary

There are st-tgds mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ such that

$$\mathcal{M}_1 \sqcap_{\mathrm{REC}} \mathcal{M}_2 \textit{ cannot be specified in FO}$$

Moreover

Corollary

The intersection $\sqcap_{\textit{st-tgds}}$ not always exists

What is the expressiveness needed
to specify the intersection of st-tgds?

Expressing the intersection of st-tgds

Observation:

- for every $\mathcal{M} \in \textsc{Rec}$ there exists an $\mathcal{M}'$ which is the *maximum recovery* of $\mathcal{M}$ (a notion of *inverse* of $\mathcal{M}$).

Expressing the intersection of st-tgds

Observation:

- for every $\mathcal{M} \in \textsc{Rec}$ there exists an $\mathcal{M}'$ which is the *maximum recovery* of $\mathcal{M}$ (a notion of *inverse* of $\mathcal{M}$).

Theorem

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be mappings in $\textsc{Rec}$, $\mathcal{M}_1'$ and $\mathcal{M}_2'$ be their maximum recoveries, and let $\mathcal{N}$ be:

$$\mathcal{N} \; = \; \big[(\mathcal{M}_1 \circ \mathcal{M}_1') \cup (\mathcal{M}_2 \circ \mathcal{M}_2')\big]$$

Expressing the intersection of st-tgds

Observation:

- for every $\mathcal{M} \in \textsc{Rec}$ there exists an $\mathcal{M}'$ which is the *maximum recovery* of $\mathcal{M}$ (a notion of *inverse* of $\mathcal{M}$).

Theorem

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be mappings in $\textsc{Rec}$, $\mathcal{M}_1'$ and $\mathcal{M}_2'$ be their maximum recoveries, and let $\mathcal{N}$ be:

$$\mathcal{N} \;=\; \big[(\mathcal{M}_1 \circ \mathcal{M}_1') \cup (\mathcal{M}_2 \circ \mathcal{M}_2')\big]$$

Then, the intersection $\mathcal{M}_1 \sqcap_{\textsc{Rec}} \mathcal{M}_2$ is the infinite composition

$$\mathcal{N} \circ \mathcal{N} \circ \mathcal{N} \circ \cdots$$

Intersection of st-tgds is expressible in SO

Theorem

Given st-tgds $\mathcal{M}_1$ and $\mathcal{M}_2$, the intersection $\mathcal{M}_1 \sqcap_{\mathrm{REC}} \mathcal{M}_2$ can be specified in Second Order logic

Intersection of st-tgds is expressible in SO

Theorem

Given st-tgds $\mathcal{M}_1$ and $\mathcal{M}_2$, the intersection $\mathcal{M}_1 \sqcap_{\mathrm{REC}} \mathcal{M}_2$ can be specified in Second Order logic

Proof idea

We show that:

Intersection of st-tgds is expressible in SO

Theorem

Given st-tgds $\mathcal{M}_1$ and $\mathcal{M}_2$, the intersection $\mathcal{M}_1 \sqcap_{\text{REC}} \mathcal{M}_2$ can be specified in Second Order logic

Proof idea

We show that:

- for an st-tgd mapping $\mathcal{M}$ the composition of $\mathcal{M}$, with its maximum recovery can be specified in FO

Intersection of st-tgds is expressible in SO

Theorem

Given st-tgds $\mathcal{M}_1$ and $\mathcal{M}_2$, the intersection $\mathcal{M}_1 \sqcap_{\mathrm{REC}} \mathcal{M}_2$ can be specified in Second Order logic

Proof idea

We show that:

- for an st-tgd mapping $\mathcal{M}$ the composition of $\mathcal{M}$, with its maximum recovery can be specified in FO
- the infinite composition of mappings specified in FO, can be specified in SO

Outline

Intuition and definition

Existence of union and intersection

Concluding remarks and future work

Union and Intersection of schema mappings

We have proposed two new operators $\sqcup$ and $\sqcap$

- based on information transfer
- applications in mapping generation and query answering
- can be used to define new operators: e.g. *difference*

Union and Intersection of schema mappings

We have proposed two new operators $\sqcup$ and $\sqcap$

- based on information transfer
- applications in mapping generation and query answering
- can be used to define new operators: e.g. *difference*

Future work include

- more on expressiveness of intersection:
  - do we really need SO? can we use less expressiveness?
  - when the intersection is expressible in FO or as st-tgds?
- complexity
  - is it decidable if $\mathcal{M}$ is the intersection of $\mathcal{M}_1$ and $\mathcal{M}_2$?

Union and Intersection of Schema Mappings

Jorge Pérez Reinhard Pichler
Emanuel Sallinger Vadim Savenkov

Universidad de Chile, TU Vienna

Outline

Intuition and definition

Existence of union and intersection

Concluding remarks and future work