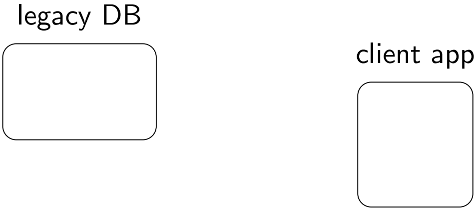
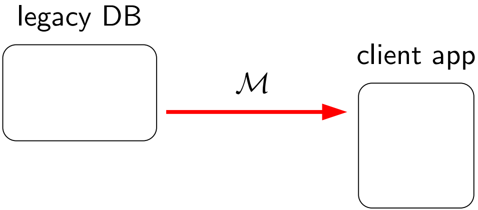
Information and *redundancy*: fundamental concepts in schema mapping management

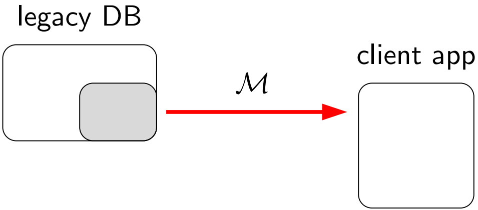
Information and *redundancy*: fundamental concepts in schema mapping management



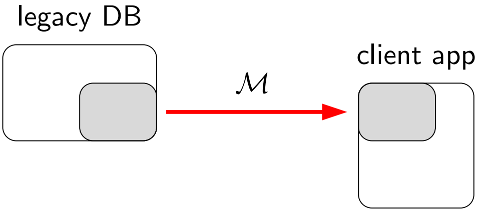
Information and *redundancy*:
fundamental concepts in schema mapping management



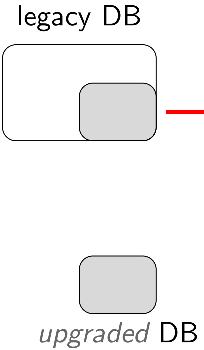
Information and *redundancy*: fundamental concepts in schema mapping management



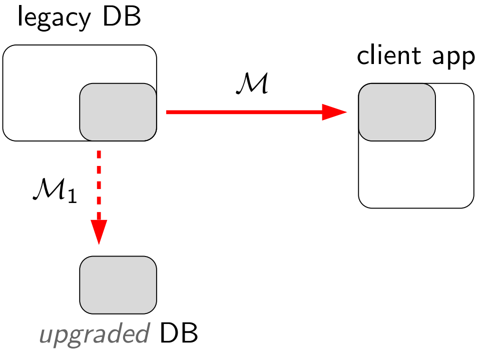
Information and *redundancy*: fundamental concepts in schema mapping management



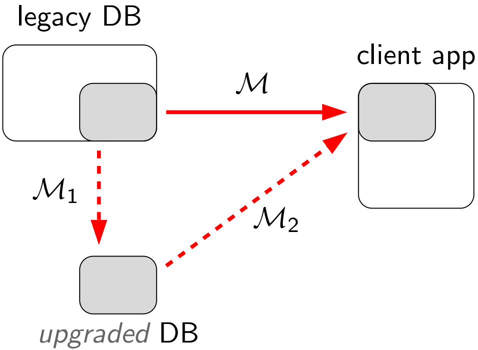
Information and *redundancy*: fundamental concepts in schema mapping management



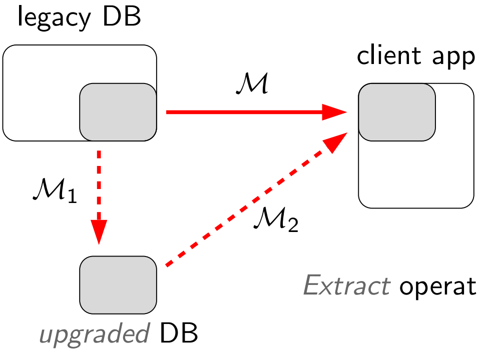
Information and *redundancy*:
fundamental concepts in schema mapping management



Information and *redundancy*: fundamental concepts in schema mapping management

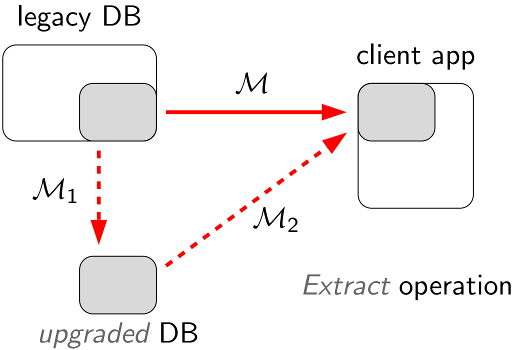


Information and *redundancy*: fundamental concepts in schema mapping management



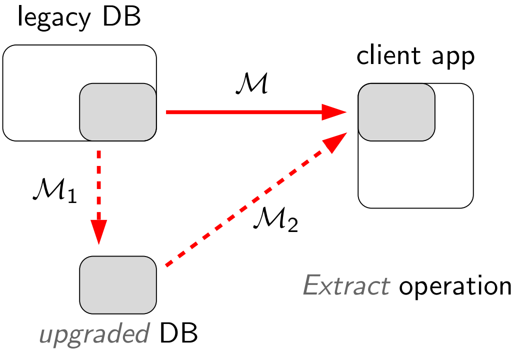
Extract operation

Information and *redundancy*: fundamental concepts in schema mapping management



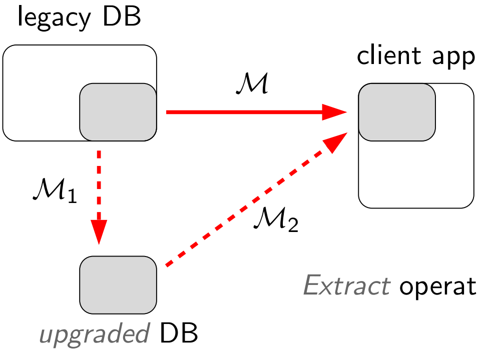
The new DB should only contain the information transferred by $\mathcal{M}$.

Information and *redundancy*: fundamental concepts in schema mapping management



The new DB should only contain the information transferred by $\mathcal{M}$.

Information and *redundancy*: fundamental concepts in schema mapping management



Extract operation

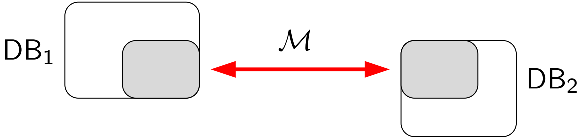
The new DB should only contain the information transferred by $\mathcal{M}$.

Information and *redundancy*: fundamental concepts in schema mapping management

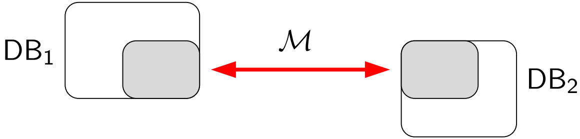
Information and *redundancy*: fundamental concepts in schema mapping management

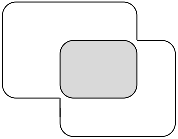
Information and *redundancy*: fundamental concepts in schema mapping management

Information and *redundancy*:
fundamental concepts in schema mapping management

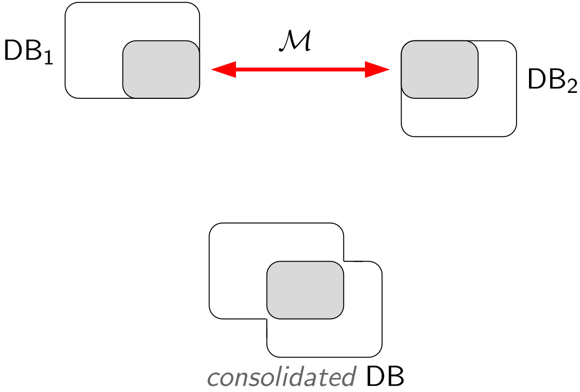


Information and *redundancy*:
fundamental concepts in schema mapping management

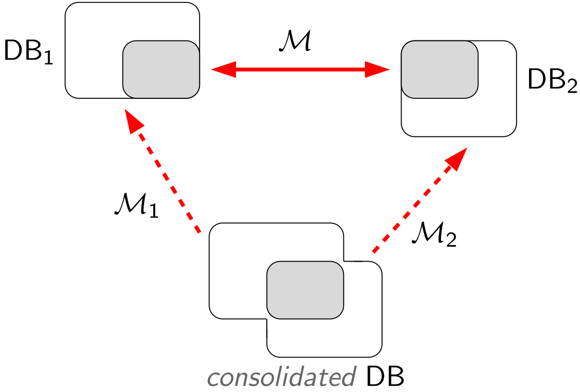




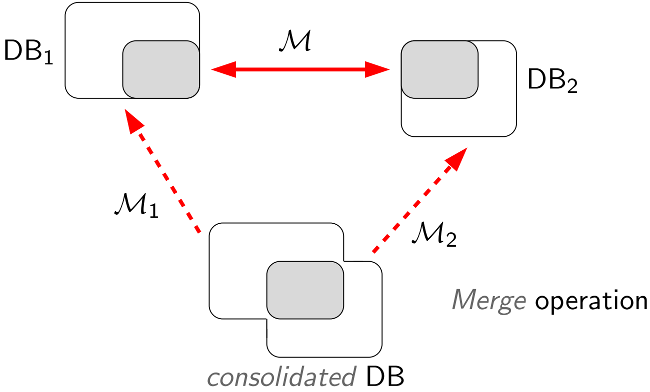
Information and *redundancy*:
fundamental concepts in schema mapping management



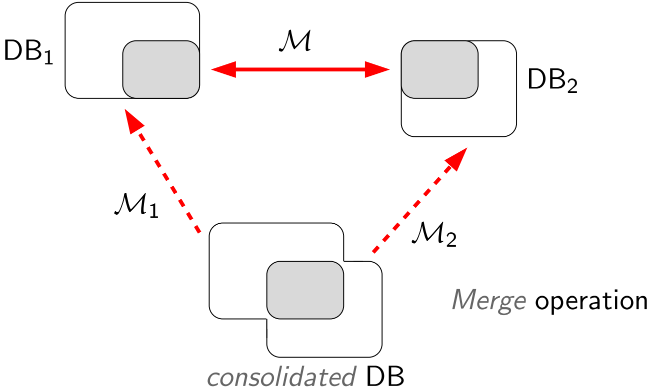
Information and *redundancy*: fundamental concepts in schema mapping management



Information and *redundancy*:
fundamental concepts in schema mapping management

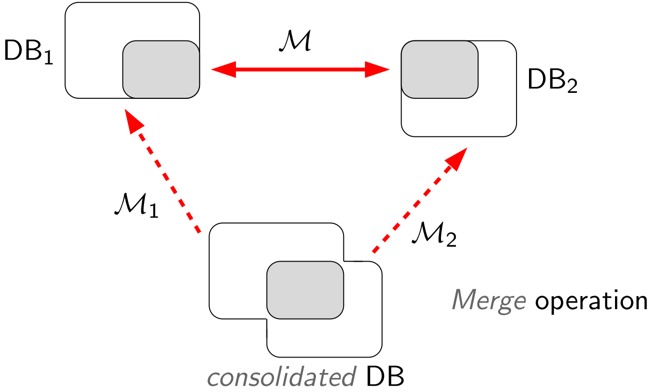


Information and *redundancy*: fundamental concepts in schema mapping management



The new DB should only store the non redundant information w.r.t. $\mathcal{M}$.

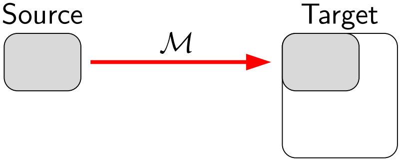
Information and *redundancy*: fundamental concepts in schema mapping management



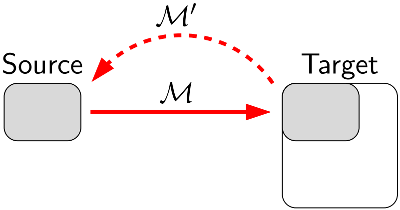
The new DB should only store the non redundant information w.r.t. $\mathcal{M}$.

Information and *redundancy*: fundamental concepts in schema mapping management

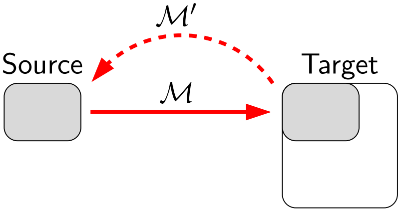
Information and *redundancy*:
fundamental concepts in schema mapping management



Information and *redundancy*: fundamental concepts in schema mapping management

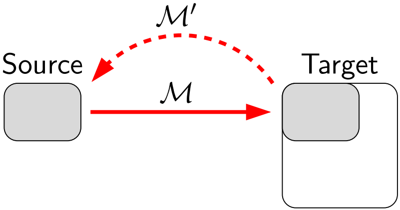


Information and *redundancy*: fundamental concepts in schema mapping management



Inverse operation

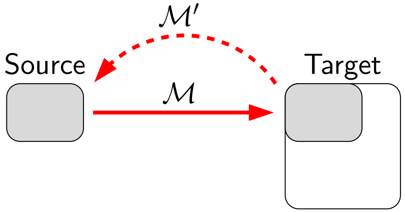
Information and *redundancy*:
fundamental concepts in schema mapping management



Inverse operation

Invertibility for $\mathcal{M}$ should coincide with no loss of information.

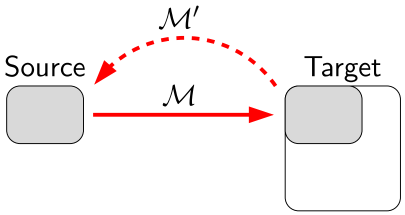
Information and *redundancy*: fundamental concepts in schema mapping management



Inverse operation

Invertibility for $\mathcal{M}$ should coincide with no loss of information.

Information and *redundancy*: fundamental concepts in schema mapping management



Inverse operation

Although fundamental, the notions of *information* and *redundancy* have received little attention in the schema mapping context.

Foundations of Schema Mapping Management

Marcelo Arenas, Jorge Pérez, Juan L. Reutter, Cristian Riveros

PUC Chile, U. Edinburgh, U. Oxford

We provide foundations for schema mapping management by formalizing the notions of *information* and *redundancy*.

We provide foundations for schema mapping management by formalizing the notions of *information* and *redundancy*.

Main contributions:

1. *Information* and *redundancy* in schema mappings
   - general formalization
   - characterizations and algorithmic issues

We provide foundations for schema mapping management by formalizing the notions of *information* and *redundancy*.

Main contributions:

1. *Information* and *redundancy* in schema mappings
   - general formalization
   - characterizations and algorithmic issues
2. Applications of the notions:
   - schema evolution problem
   - *Extract*, *Merge* and *Inverse* operators

Outline

A bit of notation...

A *mapping* $\mathcal{M}$ is a set of pairs (I, J) with

- I a source instance and J a target instance
  (J is called a *solution* for I under $\mathcal{M}$).

A bit of notation...

A *mapping* $\mathcal{M}$ is a set of pairs (I, J) with

- I a source instance and J a target instance
  (J is called a *solution* for I under $\mathcal{M}$).
- the *composition of mappings*, $\mathcal{M} \circ \mathcal{M}'$, is the usual composition of binary relations.

A bit of notation...

A *mapping* $\mathcal{M}$ is a set of pairs (I, J) with

- I a source instance and J a target instance (J is called a *solution* for I under $\mathcal{M}$).
- the *composition of mappings*, $\mathcal{M} \circ \mathcal{M}'$, is the usual composition of binary relations.

Mappings can be specified by formulas (dependencies):

$$\varphi_{\mathbf{S}}(\bar{x}) \rightarrow \psi_{\mathbf{T}}(\bar{x})$$

A bit of notation...

A *mapping* $\mathcal{M}$ is a set of pairs (I, J) with

- I a source instance and J a target instance (J is called a *solution* for I under $\mathcal{M}$).
- the *composition of mappings*, $\mathcal{M} \circ \mathcal{M}'$, is the usual composition of binary relations.

Mappings can be specified by formulas (dependencies):

$$\varphi_{\mathbf{S}}(\bar{x}) \rightarrow \psi_{\mathbf{T}}(\bar{x})$$

with $\varphi_{\mathbf{S}}(\bar{x})$ formula over the source and $\psi_{\mathbf{T}}(\bar{x})$ over the target.

A bit of notation...

A *mapping* $\mathcal{M}$ is a set of pairs (I, J) with

- I a source instance and J a target instance (J is called a *solution* for I under $\mathcal{M}$).
- the *composition of mappings*, $\mathcal{M} \circ \mathcal{M}'$, is the usual composition of binary relations.

Mappings can be specified by formulas (dependencies):

$$\varphi_{\mathbf{S}}(\bar{x}) \rightarrow \psi_{\mathbf{T}}(\bar{x})$$

with $\varphi_{\mathbf{S}}(\bar{x})$ formula over the source and $\psi_{\mathbf{T}}(\bar{x})$ over the target.

- **$\mathbf{L}_1$-to-$\mathbf{L}_2$** dependency: $\varphi_{\mathbf{S}}(\bar{x}) \in \mathbf{L}_1$ and $\psi_{\mathbf{T}}(\bar{x}) \in \mathbf{L}_2$.

A bit of notation...

A *mapping* $\mathcal{M}$ is a set of pairs (I, J) with

- I a source instance and J a target instance (J is called a *solution* for I under $\mathcal{M}$).
- the *composition of mappings*, $\mathcal{M} \circ \mathcal{M}'$, is the usual composition of binary relations.

Mappings can be specified by formulas (dependencies):

$$\varphi_{\mathbf{S}}(\bar{x}) \rightarrow \psi_{\mathbf{T}}(\bar{x})$$

with $\varphi_{\mathbf{S}}(\bar{x})$ formula over the source and $\psi_{\mathbf{T}}(\bar{x})$ over the target.

- $\mathbf{L}_1$**-to-**$\mathbf{L}_2$ dependency: $\varphi_{\mathbf{S}}(\bar{x}) \in \mathbf{L}_1$ and $\psi_{\mathbf{T}}(\bar{x}) \in \mathbf{L}_2$.
- **CQ-to-CQ** $=$ st-tgds.
- we are also interested in $\mathbf{CQ}^{\neq}$**-to-CQ** and **FO-to-CQ**.

Source information transferred by a mapping: Intuition

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

Source information transferred by a mapping: Intuition

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

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

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

Source information transferred by a mapping: Intuition

Source: $\{\text{Emp}(\textbf{name}, \textbf{lives\_in}, \textbf{works\_in})\ \}$

Target_1: $\{\text{Person}(\textbf{ssn}, \textbf{name})\ \}$

Target_2: $\{\text{ENames}(\textbf{name}), \text{WorksIn}(\textbf{name}, \textbf{place})\ \}$

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

Source information transferred by a mapping: Intuition

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

Target$_1$: {Person(**ssn**, **name**) }

Target$_2$: {ENames(**name**), WorksIn(**name**, **place**) }

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

Source information transferred by a mapping: Intuition

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

Target$_1$: {Person(**ssn**, **name**) }

Target$_2$: {ENames(**name**), WorksIn(**name**, **place**) }

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

Intuitively:

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

Source information transferred by a mapping: Intuition

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

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

Intuitively:

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

Source information transferred by a mapping: Intuition

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

Target$_1$: {Person(**ssn**, **name**) }

Target$_2$: {ENames(**name**), WorksIn(**name**, **place**) }

Target$_3$: {Workplace(**place**) }

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

Intuitively:

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

Source information transferred by a mapping: Intuition

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

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

Intuitively:

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

Source information transferred by a mapping: Intuition

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

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

Intuitively:

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

Source information transfered by a mapping: Formalization

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

Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2,$$

Source information transfered by a mapping: Formalization

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

Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2,$$

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

Source information transfered by a mapping: Formalization

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

Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2,$$

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

$\mathcal{M}_2$ transfers information enough to reconstruct $\mathcal{M}_1$.

Source information transfered by a mapping: Formalization

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

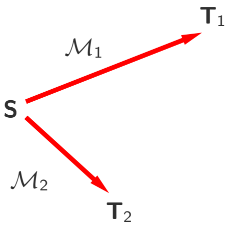
Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2,$$

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

$\mathcal{M}_2$ transfers information enough to reconstruct $\mathcal{M}_1$.



Source information transfered by a mapping: Formalization

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

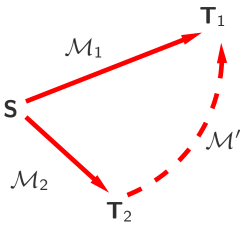
Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2,$$

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

$\mathcal{M}_2$ transfers information enough to reconstruct $\mathcal{M}_1$.



Source information transfered by a mapping: Formalization

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

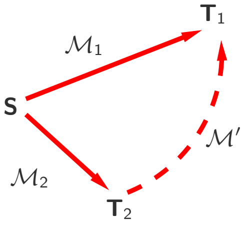
Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2,$$

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

$\mathcal{M}_2$ transfers information enough to reconstruct $\mathcal{M}_1$.



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

Source information transfered by a mapping: Formalization

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

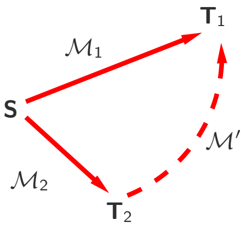
Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2,$$

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

$\mathcal{M}_2$ transfers information enough to reconstruct $\mathcal{M}_1$.



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

Source information transfered by a mapping: Formalization

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

Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2,$$

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

$\mathcal{M}_2$ transfers information enough to reconstruct $\mathcal{M}_1$.

$$\begin{tikzcd}
 & \mathbf{T}_1 \\
\mathbf{S} \arrow[ur, "\mathcal{M}_1"] \arrow[dr, "\mathcal{M}_2"'] & \\
 & \mathbf{T}_2 \arrow[uu, dashed, bend right, "\mathcal{M}'"']
\end{tikzcd}$$

$$\begin{aligned}
\mathcal{M}_1 :&\quad \mathrm{Emp}(x, y, z) \rightarrow \exists u\, \mathrm{Person}(u, x) \\
\mathcal{M}_2 :&\quad \mathrm{Emp}(x, y, z) \rightarrow \mathrm{ENames}(x) \wedge \mathrm{WorksIn}(x, z) \\
\mathcal{M}' :&\quad \mathrm{ENames}(x) \rightarrow \exists u\, \mathrm{Person}(u, x)
\end{aligned}$$

Source information transfered by a mapping: Formalization

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

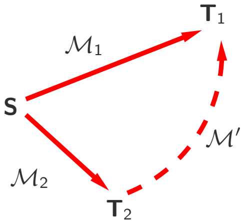
Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2,$$

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

$\mathcal{M}_2$ transfers information enough to reconstruct $\mathcal{M}_1$.



$$\begin{aligned}
\mathcal{M}_1: &\quad \mathrm{Emp}(x,y,z) \rightarrow \exists u\ \mathrm{Person}(u,x) \\
\mathcal{M}_2: &\quad \mathrm{Emp}(x,y,z) \rightarrow \mathrm{ENames}(x) \wedge \mathrm{WorksIn}(x,z) \\
\mathcal{M}': &\quad \mathrm{ENames}(x) \rightarrow \exists u\ \mathrm{Person}(u,x)
\end{aligned}$$

$$\mathcal{M}_2 \circ \mathcal{M}' = \mathcal{M}_1 \implies \mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2$$

Axiomatization of $\preceq_{\mathrm{S}}$

In the paper, we first define 4 *axioms* for an order $\preceq$ on mappings

Axiomatization of $\preceq_{\mathrm{S}}$

In the paper, we first define 4 *axioms* for an order $\preceq$ on mappings

(C1) *reflexivity* : $\mathcal{M} \preceq \mathcal{M}$

(C2) *transitivity* : $\mathcal{M}_1 \preceq \mathcal{M}_2$ and $\mathcal{M}_2 \preceq \mathcal{M}_3$, then $\mathcal{M}_1 \preceq \mathcal{M}_3$

Axiomatization of $\preceq_{\mathrm{S}}$

In the paper, we first define 4 *axioms* for an order $\preceq$ on mappings

(C1) *reflexivity* : $\mathcal{M} \preceq \mathcal{M}$

(C2) *transitivity* : $\mathcal{M}_1 \preceq \mathcal{M}_2$ and $\mathcal{M}_2 \preceq \mathcal{M}_3$, then $\mathcal{M}_1 \preceq \mathcal{M}_3$

(C3) *maximum* : $\mathcal{M} \preceq \mathrm{Id} = \{(I, I) \mid I \text{ is a source instance}\}$

Axiomatization of $\preceq_{\mathrm{S}}$

In the paper, we first define 4 *axioms* for an order $\preceq$ on mappings

(C1) *reflexivity* : $\mathcal{M} \preceq \mathcal{M}$

(C2) *transitivity* : $\mathcal{M}_1 \preceq \mathcal{M}_2$ and $\mathcal{M}_2 \preceq \mathcal{M}_3$, then $\mathcal{M}_1 \preceq \mathcal{M}_3$

(C3) *maximum* : $\mathcal{M} \preceq \mathsf{Id} = \{(I, I) \mid I \text{ is a source instance}\}$

(C4) *preservation* : $\mathcal{M}_1 \preceq \mathcal{M}_2$ then $\mathcal{M} \circ \mathcal{M}_1 \preceq \mathcal{M} \circ \mathcal{M}_2$

Axiomatization of $\preceq_{\mathrm{S}}$

In the paper, we first define 4 *axioms* for an order $\preceq$ on mappings

(C1) *reflexivity* : $\mathcal{M} \preceq \mathcal{M}$

(C2) *transitivity* : $\mathcal{M}_1 \preceq \mathcal{M}_2$ and $\mathcal{M}_2 \preceq \mathcal{M}_3$, then $\mathcal{M}_1 \preceq \mathcal{M}_3$

(C3) *maximum* : $\mathcal{M} \preceq \mathsf{Id} = \{(I, I) \mid I \text{ is a source instance}\}$

(C4) *preservation* : $\mathcal{M}_1 \preceq \mathcal{M}_2$ then $\mathcal{M} \circ \mathcal{M}_1 \preceq \mathcal{M} \circ \mathcal{M}_2$

(C4)
$$\begin{tikzcd}
 & & \mathbf{T}_1 \\
\mathbf{S}' \arrow[r, "\mathcal{M}"] & \mathbf{S} \arrow[ur, "\mathcal{M}_1"] \arrow[dr, "\mathcal{M}_2"'] & \\
 & & \mathbf{T}_2
\end{tikzcd}$$

Axiomatization of $\preceq_{\mathrm{S}}$

In the paper, we first define 4 *axioms* for an order $\preceq$ on mappings

(C1) *reflexivity* : $\mathcal{M} \preceq \mathcal{M}$

(C2) *transitivity* : $\mathcal{M}_1 \preceq \mathcal{M}_2$ and $\mathcal{M}_2 \preceq \mathcal{M}_3$, then $\mathcal{M}_1 \preceq \mathcal{M}_3$

(C3) *maximum* : $\mathcal{M} \preceq \mathsf{Id} = \{(I, I) \mid I \text{ is a source instance}\}$

(C4) *preservation* : $\mathcal{M}_1 \preceq \mathcal{M}_2$ then $\mathcal{M} \circ \mathcal{M}_1 \preceq \mathcal{M} \circ \mathcal{M}_2$

(C4)
$$\begin{tikzcd} & & \mathbf{T}_1 \\ \mathbf{S}' \arrow[r, "\mathcal{M}"] & \mathbf{S} \arrow[ur, "\mathcal{M}_1"] \arrow[dr, "\mathcal{M}_2"'] & \\ & & \mathbf{T}_2 \end{tikzcd}$$

Theorem

The order $\preceq_{\mathrm{S}}$ is the strictest *relation that satisfies (C1-C4).*

Towards deciding $\preceq_{\mathrm{S}}$: target rewritability

Certain answers

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

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

Towards deciding $\preceq_{\mathrm{S}}$: target rewritability

Certain answers

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

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

Definition

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.

Towards deciding $\preceq_S$: target rewritability

Certain answers

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

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

Definition

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 $Q_\mathbf{S}$.

Source information transferred by a mapping can be characterized in terms of queries.

Theorem

Let $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *be specified by* **FO-to-CQ**, *then:*

$$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2 \text{ if and only if}$$

every source query *that is target rewritable under* $\mathcal{M}_1$
is also target rewritable under $\mathcal{M}_2$.

Source information transferred by a mapping can be characterized in terms of queries.

Theorem

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be specified by **FO-to-CQ**, *then:*

$\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2$ *if and only if*

every source query *that is target rewritable under $\mathcal{M}_1$ is also target rewritable under $\mathcal{M}_2$.*

- The characterization is particular for **FO-to-CQ**. For example, it does not work for **CQ-to-UCQ**.

Deciding $\preceq_{\mathrm{S}}$

Theorem

For mappings specified by **FO-to-CQ**:

testing $\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2$ *is* undecidable

Deciding $\preceq_S$

Theorem

For mappings specified by **FO-to-CQ***:*

testing $\mathcal{M}_1 \preceq_S \mathcal{M}_2$ *is* undecidable

Theorem

For mappings specified by $\mathbf{CQ}^{\neq}$**-to-CQ**

testing $\mathcal{M}_1 \preceq_S \mathcal{M}_2$ *is* decidable

Deciding $\preceq_{\mathrm{S}}$

Theorem

For mappings specified by **FO-to-CQ***:*

$$\textit{testing } \mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2 \textit{ is } \text{undecidable}$$

Theorem

For mappings specified by **CQ$^{\neq}$-to-CQ**

$$\textit{testing } \mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2 \textit{ is } \text{decidable}$$

Proof idea

For **CQ$^{\neq}$-to-CQ** mappings, we prove that:

- checking target rewritability for **UCQ$^{\neq}$** is decidable,
- only a finite number of queries in **UCQ$^{\neq}$** need to be checked to determine if $\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}_2$.

Application: Invertibility can be characterized using $\preceq_{\mathrm{S}}$.

Let $\overline{\mathsf{Id}}$ be a mapping specified by a set of *copying*
(rules of the form $R(\bar{x}) \rightarrow \hat{R}(\bar{x})$ with R a source relation).

Application: Invertibility can be characterized using $\preceq_{\mathrm{S}}$.

Let $\overline{\mathsf{Id}}$ be a mapping specified by a set of *copying*
(rules of the form $R(\bar{x}) \rightarrow \hat{R}(\bar{x})$ with R a source relation).

Definition [F06]: $\mathcal{M}'$ is an *inverse* of $\mathcal{M}$ if $\mathcal{M} \circ \mathcal{M}' = \overline{\mathsf{Id}}$.

Application: Invertibility can be characterized using $\preceq_{\mathrm{S}}$.

Let $\overline{\mathsf{Id}}$ be a mapping specified by a set of *copying*
(rules of the form $R(\bar{x}) \rightarrow \hat{R}(\bar{x})$ with R a source relation).

Definition [F06]: $\mathcal{M}'$ is an *inverse* of $\mathcal{M}$ if $\mathcal{M} \circ \mathcal{M}' = \overline{\mathsf{Id}}$.

Theorem

Consider the class of total and closed-down on the left mappings:

- $\mathcal{M}$ *is invertible* $\iff$ $\overline{\mathit{Id}} \preceq_{\mathrm{S}} \mathcal{M}$
- $\mathcal{M}$ *is invertible* $\iff$ $\mathcal{M}$ *is* $\preceq_{\mathrm{S}}$-maximal

Application: Invertibility can be characterized using $\preceq_{\mathrm{S}}$.

Let $\overline{\mathsf{Id}}$ be a mapping specified by a set of *copying*
(rules of the form $R(\bar{x}) \rightarrow \hat{R}(\bar{x})$ with R a source relation).

Definition [F06]: $\mathcal{M}'$ is an *inverse* of $\mathcal{M}$ if $\mathcal{M} \circ \mathcal{M}' = \overline{\mathsf{Id}}$.

Theorem

Consider the class of total and closed-down on the left mappings:

- $\mathcal{M}$ *is invertible* $\iff$ $\overline{Id} \preceq_{\mathrm{S}} \mathcal{M}$
- $\mathcal{M}$ *is invertible* $\iff$ $\mathcal{M}$ *is* $\preceq_{\mathrm{S}}$-maximal

Invertibility do coincide with
transferring the maximum amount of source information!

Application: Invertibility can be characterized using $\preceq_{\mathrm{S}}$.

Let $\overline{\mathrm{Id}}$ be a mapping specified by a set of *copying*
(rules of the form $R(\bar{x}) \rightarrow \hat{R}(\bar{x})$ with R a source relation).

Definition [F06]: $\mathcal{M}'$ is an *inverse* of $\mathcal{M}$ if $\mathcal{M} \circ \mathcal{M}' = \overline{\mathrm{Id}}$.

Theorem

Consider the class of total and closed-down on the left mappings:

- $\mathcal{M}$ *is invertible* $\iff$ $\overline{Id} \preceq_{\mathrm{S}} \mathcal{M}$
- $\mathcal{M}$ *is invertible* $\iff$ $\mathcal{M}$ *is* $\preceq_{\mathrm{S}}$-maximal

Invertibility do coincide with
transferring the maximum amount of source information!

Corollary [FN09]

Testing invertibility for **CQ$^{\neq}$-to-CQ** mappings is decidable.

Covering target information: the *dual* definition

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

Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{T}} \mathcal{M}_2,$$

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

Covering target information: the *dual* definition

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

Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{T}} \mathcal{M}_2,$$

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

- *Universal solutions* [FKMP05]: A solution $J^\star$ for an instance I that represents the *entire space of solutions* of I under $\mathcal{M}$.

Covering target information: the *dual* definition

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

Definition

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

$$\mathcal{M}_1 \preceq_{\mathrm{T}} \mathcal{M}_2,$$

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

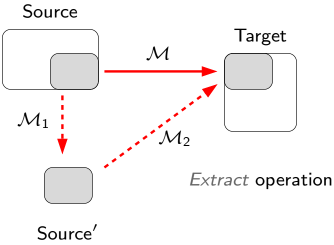
- *Universal solutions* [FKMP05]: A solution $J^\star$ for an instance I that represents the *entire space of solutions* of I under $\mathcal{M}$.

Theorem

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be specified by **FO-to-CQ**, *then:*

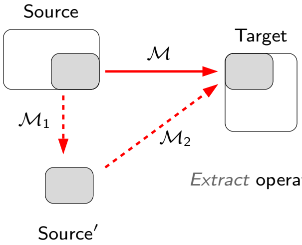
$\mathcal{M}_1 \preceq_{\mathrm{T}} \mathcal{M}_2$ *if and only if*
every target instance that is universal solution under $\mathcal{M}_1$
is also universal solution under $\mathcal{M}_2$.

Application 2: formalization of *Extract* (first attempt)



- We model the *extract* of $\mathcal{M}$ as a pair $(\mathcal{M}_1, \mathcal{M}_2)$ s.t.

Application 2: formalization of *Extract* (first attempt)

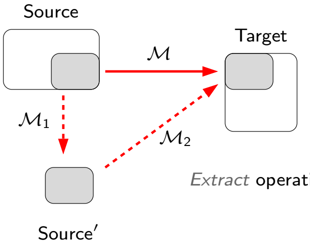


Extract operation

- We model the *extract* of $\mathcal{M}$ as a pair $(\mathcal{M}_1, \mathcal{M}_2)$ s.t.

  (E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$

Application 2: formalization of *Extract* (first attempt)

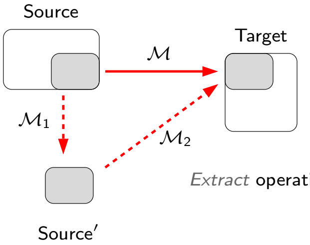


Extract operation

- We model the *extract* of $\mathcal{M}$ as a pair $(\mathcal{M}_1, \mathcal{M}_2)$ s.t.

  (E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$ (i.e. $\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}$ and $\mathcal{M} \preceq_{\mathrm{S}} \mathcal{M}_1$).

Application 2: formalization of *Extract* (first attempt)



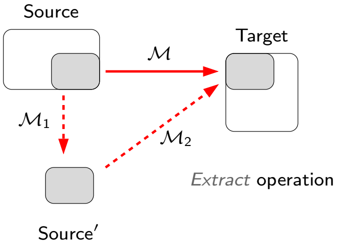
Extract operation

- We model the *extract* of $\mathcal{M}$ as a pair $(\mathcal{M}_1, \mathcal{M}_2)$ s.t.

  (E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$ (i.e. $\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}$ and $\mathcal{M} \preceq_{\mathrm{S}} \mathcal{M}_1$).

  (E2) $\mathcal{M}_2 \equiv_{\mathrm{T}} \mathcal{M}$

Application 2: formalization of *Extract* (first attempt)

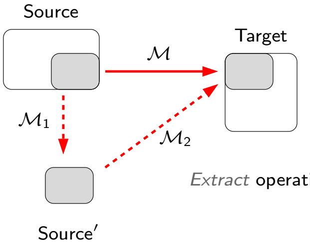


- We model the *extract* of $\mathcal{M}$ as a pair $(\mathcal{M}_1, \mathcal{M}_2)$ s.t.

  (E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$ (i.e. $\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}$ and $\mathcal{M} \preceq_{\mathrm{S}} \mathcal{M}_1$).

  (E2) $\mathcal{M}_2 \equiv_{\mathrm{T}} \mathcal{M}$ (i.e. $\mathcal{M}_2 \preceq_{\mathrm{T}} \mathcal{M}$ and $\mathcal{M} \preceq_{\mathrm{T}} \mathcal{M}_2$).

Application 2: formalization of *Extract* (first attempt)



Extract operation

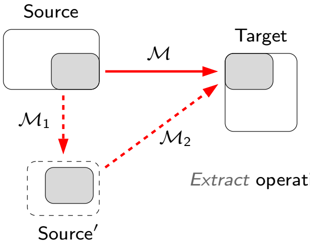
- We model the *extract* of $\mathcal{M}$ as a pair $(\mathcal{M}_1, \mathcal{M}_2)$ s.t.

(E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$ (i.e. $\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}$ and $\mathcal{M} \preceq_{\mathrm{S}} \mathcal{M}_1$).

(E2) $\mathcal{M}_2 \equiv_{\mathrm{T}} \mathcal{M}$ (i.e. $\mathcal{M}_2 \preceq_{\mathrm{T}} \mathcal{M}$ and $\mathcal{M} \preceq_{\mathrm{T}} \mathcal{M}_2$).

(E3) $\mathcal{M} = \mathcal{M}_1 \circ \mathcal{M}_2$

Application 2: formalization of *Extract* (first attempt)



Extract operation

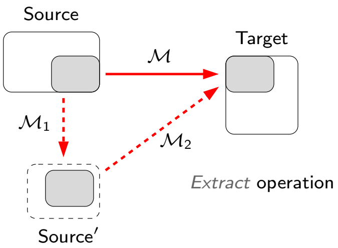
- We model the *extract* of $\mathcal{M}$ as a pair $(\mathcal{M}_1, \mathcal{M}_2)$ s.t.

(E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$ (i.e. $\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}$ and $\mathcal{M} \preceq_{\mathrm{S}} \mathcal{M}_1$).

(E2) $\mathcal{M}_2 \equiv_{\mathrm{T}} \mathcal{M}$ (i.e. $\mathcal{M}_2 \preceq_{\mathrm{T}} \mathcal{M}$ and $\mathcal{M} \preceq_{\mathrm{T}} \mathcal{M}_2$).

(E3) $\mathcal{M} = \mathcal{M}_1 \circ \mathcal{M}_2$

Application 2: formalization of *Extract* (first attempt)



- We model the *extract* of $\mathcal{M}$ as a pair $(\mathcal{M}_1, \mathcal{M}_2)$ s.t.

(E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$ (i.e. $\mathcal{M}_1 \preceq_{\mathrm{S}} \mathcal{M}$ and $\mathcal{M} \preceq_{\mathrm{S}} \mathcal{M}_1$).

(E2) $\mathcal{M}_2 \equiv_{\mathrm{T}} \mathcal{M}$ (i.e. $\mathcal{M}_2 \preceq_{\mathrm{T}} \mathcal{M}$ and $\mathcal{M} \preceq_{\mathrm{T}} \mathcal{M}_2$).

(E3) $\mathcal{M} = \mathcal{M}_1 \circ \mathcal{M}_2$

¿How do we ensure the *optimality* of the new source schema?

Outline

Target redundancy in mappings: Intuition

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

Target redundancy in mappings: Intuition

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

Target$_1$: {ENames(**name**), WorksIn(**name**, **place**) }

Target redundancy in mappings: Intuition

Source: $\{\text{Emp}(\textbf{name}, \textbf{lives\_in}, \textbf{works\_in})\ \}$
Target$_1$: $\{\text{ENames}(\textbf{name}), \text{WorksIn}(\textbf{name}, \textbf{place})\ \}$

$$\mathcal{M}_1: \qquad \text{Emp}(x, y, z) \quad \rightarrow \quad \text{ENames}(x) \wedge \text{WorksIn}(x, z)$$

Target redundancy in mappings: Intuition

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {ENames(**name**), WorksIn(**name**, **place**) }
Target$_2$: {Worker(**name**, **working_place**) }

$$\mathcal{M}_1: \quad \mathrm{Emp}(x, y, z) \quad \rightarrow \quad \mathrm{ENames}(x) \wedge \mathrm{WorksIn}(x, z)$$

Target redundancy in mappings: Intuition

Source: $\{\text{Emp}(\textbf{name}, \textbf{lives\_in}, \textbf{works\_in})\ \}$

Target$_1$: $\{\text{ENames}(\textbf{name}), \text{WorksIn}(\textbf{name}, \textbf{place})\ \}$

Target$_2$: $\{\text{Worker}(\textbf{name}, \textbf{working\_place})\ \}$

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

Target redundancy in mappings: Intuition

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {ENames(**name**), WorksIn(**name**, **place**) }
Target$_2$: {Worker(**name**, **working_place**) }

$$\mathcal{M}_1: \quad \text{Emp}(x, y, z) \rightarrow \text{ENames}(x) \land \text{WorksIn}(x, z)$$
$$\mathcal{M}_2: \quad \text{Emp}(x, y, z) \rightarrow \text{Worker}(x, z)$$

Intuitively:

- $\mathcal{M}_1$ is *target redundant*:
  employee names are stored twice in the target schema.

Target redundancy in mappings: Intuition

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {ENames(**name**), WorksIn(**name**, **place**) }
Target$_2$: {Worker(**name**, **working_place**) }

$$\begin{array}{llcl} \mathcal{M}_1: & \mathrm{Emp}(x,y,z) & \rightarrow & \mathrm{ENames}(x) \wedge \mathrm{WorksIn}(x,z) \\ \mathcal{M}_2: & \mathrm{Emp}(x,y,z) & \rightarrow & \mathrm{Worker}(x,z) \end{array}$$

Intuitively:

- $\mathcal{M}_1$ is *target redundant*:
  employee names are stored twice in the target schema.
- $\mathcal{M}_2$ is *not target redundant*:
  all information in the target is *essential* for $\mathcal{M}_2$.

Target redundancy in mappings: Intuition

Source: {Emp(**name**, **lives_in**, **works_in**) }
Target$_1$: {ENames(**name**), WorksIn(**name**, **place**) }
Target$_2$: {Worker(**name**, **working_place**) }

$$\begin{array}{lccc} \mathcal{M}_1: & \mathrm{Emp}(x,y,z) & \rightarrow & \mathrm{ENames}(x) \wedge \mathrm{WorksIn}(x,z) \\ \mathcal{M}_2: & \mathrm{Emp}(x,y,z) & \rightarrow & \mathrm{Worker}(x,z) \end{array}$$

Intuitively:

- $\mathcal{M}_1$ is *target redundant*:
  employee names are stored twice in the target schema.
- $\mathcal{M}_2$ is *not target redundant*:
  all information in the target is *essential* for $\mathcal{M}_2$.

Notice that $\mathcal{M}_1$ and $\mathcal{M}_2$ are *equally source-informative*, $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}_2$

Target redundancy in mappings: Formalization

Definition

$\mathcal{M}$ is *target redundant* if there is an instance $J^\star \in \text{range}(\mathcal{M})$ such that the mapping

$$\mathcal{M}' = \{(I, J) \in \mathcal{M} \mid J \neq J^\star\}$$

and $\mathcal{M}$ are equally source-informative ($\mathcal{M} \equiv_{\mathrm{S}} \mathcal{M}'$).

Target redundancy in mappings: Formalization

Definition

$\mathcal{M}$ is *target redundant* if there is an instance $J^\star \in \text{range}(\mathcal{M})$ such that the mapping

$$\mathcal{M}' = \{(I, J) \in \mathcal{M} \mid J \neq J^\star\}$$

and $\mathcal{M}$ are equally source-informative ($\mathcal{M} \equiv_{\mathrm{S}} \mathcal{M}'$).

We can *lose a target instance*, and still be able
to transfer the same amount of source information.

Target redundancy in mappings: Formalization

Definition

$\mathcal{M}$ is *target redundant* if there is an instance $J^\star \in \text{range}(\mathcal{M})$ such that the mapping

$$\mathcal{M}' = \{(I, J) \in \mathcal{M} \mid J \neq J^\star\}$$

and $\mathcal{M}$ are equally source-informative ($\mathcal{M} \equiv_{\mathrm{S}} \mathcal{M}'$).

We can *lose a target instance*, and still be able to transfer the same amount of source information.

$\mathcal{M}_1$: $\quad \text{Emp}(x, y, z) \quad \rightarrow \quad \text{ENames}(x) \wedge \text{WorksIn}(x, z)$

$J^\star$:

ENames:

name
Juan
Cristian

WorksIn:

name	place
Juan	Santiago

Source redundancy in mappings: the *dual* definition

Definition

$\mathcal{M}$ is *source redundant* if there is an instance $I^\star \in \text{dom}(\mathcal{M})$ such that the mapping

$$\mathcal{M}' = \{(I, J) \in \mathcal{M} \mid I \neq I^\star\}$$

and $\mathcal{M}$ are equally target-informative ($\mathcal{M} \equiv_{\mathrm{T}} \mathcal{M}'$).

Source redundancy in mappings: the *dual* definition

Definition

$\mathcal{M}$ is *source redundant* if there is an instance $I^\star \in \mathrm{dom}(\mathcal{M})$ such that the mapping

$$\mathcal{M}' = \{(I, J) \in \mathcal{M} \mid I \neq I^\star\}$$

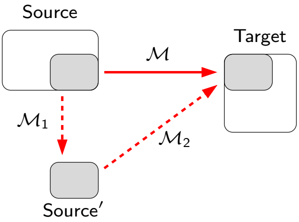
and $\mathcal{M}$ are equally target-informative ($\mathcal{M} \equiv_{\mathrm{T}} \mathcal{M}'$).

Theorem

Let $\mathcal{M}$ be specified by **FO-to-CQ***, then:*

- *$\mathcal{M}$ is target redundant iff there is a target instance that is not a universal solution under $\mathcal{M}$* (onto mapping [FN09]).
- *$\mathcal{M}$ is source redundant iff there are two source instances with the same space of solutions under $\mathcal{M}$* (unique solutions *property* [F06]).

Application 2: formalization of *Extract*



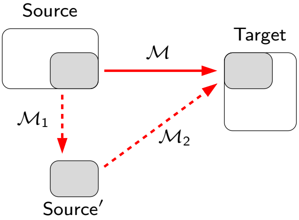
$(\mathcal{M}_1, \mathcal{M}_2)$ is an *extract* of $\mathcal{M}$ iff:

(E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$

(E2) $\mathcal{M}_2 \equiv_{\mathrm{T}} \mathcal{M}$

(E3) $\mathcal{M} = \mathcal{M}_1 \circ \mathcal{M}_2$

Application 2: formalization of *Extract*



$(\mathcal{M}_1, \mathcal{M}_2)$ is an *extract* of $\mathcal{M}$ iff:

(E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$

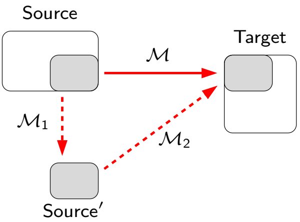
(E2) $\mathcal{M}_2 \equiv_{\mathrm{T}} \mathcal{M}$

(E3) $\mathcal{M} = \mathcal{M}_1 \circ \mathcal{M}_2$

(E4) $\mathcal{M}_1$ is *not target redundant*

(E5) $\mathcal{M}_2$ is *not source redundant*

Application 2: formalization of *Extract*



$(\mathcal{M}_1, \mathcal{M}_2)$ is an *extract* of $\mathcal{M}$ iff:

(E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$

(E2) $\mathcal{M}_2 \equiv_{\mathrm{T}} \mathcal{M}$

(E3) $\mathcal{M} = \mathcal{M}_1 \circ \mathcal{M}_2$

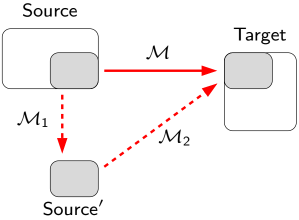
(E4) $\mathcal{M}_1$ is *not target redundant*

(E5) $\mathcal{M}_2$ is *not source redundant*

Theorem

For mappings specified by **FO-to-CQ** *an extract* always exists.

Application 2: formalization of *Extract*



$(\mathcal{M}_1, \mathcal{M}_2)$ is an *extract* of $\mathcal{M}$ iff:

(E1) $\mathcal{M}_1 \equiv_{\mathrm{S}} \mathcal{M}$

(E2) $\mathcal{M}_2 \equiv_{\mathrm{T}} \mathcal{M}$

(E3) $\mathcal{M} = \mathcal{M}_1 \circ \mathcal{M}_2$

(E4) $\mathcal{M}_1$ is *not target redundant*

(E5) $\mathcal{M}_2$ is *not source redundant*

Theorem

For mappings specified by **FO-to-CQ** *an extract always exists.*

In the paper: an algorithm to compute an extract.

Information and redundancy are fundamental notions for schema mappings

In our work:

- we provide a formalization for both notions
- we study algorithmic issues, and natural characterizations
- we use these notions to re-study some schema mapping operators (schema evolution, extract, merge, inverse).

Foundations of Schema Mapping Management

Marcelo Arenas, Jorge Pérez, Juan L. Reutter, Cristian Riveros

PUC Chile, U. Edinburgh, U. Oxford

Outline