Schema Mappings
and
Data Examples

Phokion G. Kolaitis

University of California Santa Cruz
&
IBM Research – Almaden
Schema Mappings and Data Examples

- Characterizing Schema Mappings via Data Examples
  Bogdan Alexe, Phokion Kolaitis, Wang-Chiew Tan

- Database Constraints and Homomorphism Dualities
  Balder ten Cate, Phokion Kolaitis, Wang-Chiew Tan
  Principles and Practice of Constraint Programming (CP) 2010.
Schema Mappings

- **Schema Mapping** $\mathbf{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$
  - **Source** schema $\mathbf{S}$, **Target** schema $\mathbf{T}$
  - High-level, declarative assertions $\Sigma$ that specify the relationship between $\mathbf{S}$ and $\mathbf{T}$.
  - Typically, $\Sigma$ is a finite set of formulas in some suitable logical formalism.
- Schema mappings are the essential **building blocks** in formalizing **data integration** and **data exchange**.
Data Exchange

- Data exchange: transforming data structured under a source schema into data structured under a different target schema.
- [Fagin-K …-Miller-Popa 2003]
Semantics of Schema Mappings

\[ M = (S, T, \Sigma) \] a schema mapping

- **Data Example:** A pair \((I, J)\) where \(I\) is a source instance and \(J\) is a target instance.
- **Positive Data Example for** \(M\): \((I, J) \models \Sigma\)
  - In this case, we say that \(J\) is a **solution** for \(I\) w.r.t. \(M\)
  - From a semantic point of view, \(M\) can be identified with \(\text{Sem}(M) = \{ (I, J): (I, J) \text{ is a positive data example for } M \} \)
Semantics of Schema Mappings

Note: If $M = (S, T, \Sigma)$ is a schema mapping, then $M$ is a finite syntactic representation of the infinite collection $\text{Sem}(M) = \{ (I,J): (I,J) \text{ is a positive data example for } M \}$

Problem:
- Is there a finite semantic representation of $\text{Sem}(M)$?
- Can $M$ be “captured” by finitely many data examples?
In real-life applications, schema mappings can be quite complex, even when derived manually.

There is a clear need to illustrate, understand, and refine schema mappings using “good” data examples.

- This is analogous to the venerable tradition of using test cases in understanding and debugging programs.
- Earlier work by the database community includes:
  - Yan, Miller, Haas, Fagin – 2001
    “Understanding and Refinement of Schema Mappings”
  - Olston, Chopra, Srivastava – 2009
    “Generating Example Data for Dataflow Programs”.

Motivation
Goals

- Develop a foundation for the systematic investigation of “good” data examples for schema mappings.

- Obtain technical results that shed light on both the capabilities and limitations of data examples in capturing schema mappings.
GLAV Schema Mappings

Here, we focus on GLAV schema mappings, that is, schema mappings $\mathbf{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$, where $\Sigma$ is a finite set of Global-And-Local-As-View (GLAV) constraints, also known as source-to-target tuple-generating dependencies (s-t tgds).

Note:
GLAV schema mappings are the most extensively studied and widely used class of schema mappings to date.
GLAV Schema Mappings

- The relationship between source and target is given by **Global-And-Local-As-View (GLAV)** constraints, also known as **source-to-target tuple generating dependencies (s-t tgds):**
  \[ \forall x \ (\varphi(x) \rightarrow \exists y \ \psi(x, y)) \],

- \( \varphi(x) \) is a conjunction of atoms over the source;
- \( \psi(x, y) \) is a conjunction of atoms over the target.

**Examples:**
1. \( \forall s \ \forall c \ (\text{Student}(s) \land \text{Enrolls}(s, c) \rightarrow \exists g \ \text{Grade}(s, c, g)) \)
2. \( \forall s \ \forall c \ (\text{Student}(s) \land \text{Enrolls}(s, c) \rightarrow \exists t \ \exists g \ (\text{Teaches}(t, c) \land \text{Grade}(s, c, g))) \)
LAV and GAV Schema Mappings

Fact: GLAV constraints:

(1) Generalize **LAV (local-as-view)** constraints:
\[ \forall x \ ( P(x) \rightarrow \exists y \ \psi(x, y) ) \], where \( P \) is a source relation.

(2) Generalize **GAV (global-as-view)** constraints:
\[ \forall x \ ( \varphi(x) \rightarrow R(x) ) \], where \( R \) is a target relation.
LAV and GAV Constraints

Examples of LAV (local-as-view) constraints:
- Copy: $\forall x \forall y (P(x,y) \rightarrow R(x,y))$
- Decomposition: $\forall x \forall y \forall z (Q(x,y,z) \rightarrow R(x,y) \land T(y,z))$
- $\forall x \forall y (E(x,y) \rightarrow \exists z (H(x,z) \land H(z,y)))$

Examples of GAV (global-as-view) constraints:
- Copy: $\forall x \forall y (P(x,y) \rightarrow R(x,y))$
- Projection: $\forall x \forall y \forall z (Q(x,y,z) \rightarrow T(y,z))$
- Join: $\forall x \forall y \forall z (E(x,y) \land E(y,z) \rightarrow H(x,z))$

Note:
$\forall s \forall c (\text{Student}(s) \land \text{Enrolls}(s,c) \rightarrow \exists g \text{ Grade}(s,c,g))$

is a GLAV constraint that is neither a LAV nor a GAV constraint
GLAV Mappings and Universal Solutions

**Note:** A key property of GLAV schema mappings is the **existence of universal solutions**; intuitively, they are the most general solutions.

**Theorem** (FKMP 2003) \(M = (S, T, \Sigma)\) a GLAV schema mapping.

- Every source instance \(I\) has a **universal solution** \(J\) w.r.t. \(M\), i.e., a solution \(J\) for \(I\) such that if \(J'\) is another solution for \(I\), then there is a homomorphism \(h: J \rightarrow J'\) that is constant on \(\text{adom}(I)\) (\(h(c) = c\), for \(c \in \text{adom}(I)\)).

- Moreover, the **chase procedure** can be used to construct, given a source instance \(I\), a canonical universal solution \(\text{chase}_M(I)\) for \(I\) in polynomial time.
Universal Solutions in Data Exchange

**Defn:** A homomorphism $h : J \rightarrow J'$ is a function sending every constant (non-null) value to itself, and preserving facts $(P(a_1...a_n) \in J \Rightarrow P(h(a_1)...h(a_n)) \in J')$
Example

Consider the schema mapping \( M = (\{E\}, \{F\}, \Sigma) \), where
\[
\Sigma = \{ \ E(x,y) \rightarrow \exists z \ (F(x,z) \land F(z,y)) \ \}
\]

Source instance \( I = \{ E(1,2) \} \)

**Solutions** for \( I \):

- \( J_1 = \{ F(1,X), F(X,2) \} \) (universal)
- \( J_2 = \{ F(1,2), F(2,2) \} \) (not universal)
- \( J_3 = \{ F(1,X), F(X,2), F(Y,Y) \} \) (not universal) (where \( X \) and \( Y \) are labeled null values)
- ...
Unique Characterizations via Universal Examples

**Definition:** Let $M = (S, T, \Sigma)$ be a GLAV schema mapping.

- A **universal example** for $M$ is a data example $(I, J)$ such that $J$ is a universal solution for $I$ w.r.t. $M$.

- Let $U$ be a finite set of universal examples for $M$, and let $C$ be a class of GLAV constraints. We say that $U$ **uniquely characterizes $M$ w.r.t. $C$** if for every finite set $\Sigma' \subseteq C$ such that $U$ is a set of universal examples for the schema mapping $M' = (S, T, \Sigma')$, we have that $\Sigma \equiv \Sigma'$. 
Unique Characterizations via Universal Examples

**Question:**
Which GLAV schema mappings can be uniquely characterized by a finite set of universal examples?
Theorem: Let $M$ be the schema mapping specified by the binary \textit{copy} constraint $\forall x \, \forall y \, (E(x,y) \rightarrow F(x,y))$.

- There is a finite set $U$ of universal examples that uniquely characterizes $M$ w.r.t. the class of all LAV constraints.
- There is a finite set $U'$ of universal examples that uniquely characterizes $M$ w.r.t. the class of all GAV constraints.
- There is \textbf{no} finite set of universal examples that uniquely characterizes $M$ w.r.t. the class of all GLAV constraints.
The set $U' = \{ (I_1, J_1), (I_2, J_2), (I_3, J_3) \}$ uniquely characterizes the copy schema mapping w.r.t. to the class of all GAV constraints.
Summary of Main Results

PODS 2010 paper (Alexe, K ..., Tan):

- Connection between unique characterizations and Armstrong bases.
- Every LAV schema mapping is uniquely characterizable by a finite set of universal examples w.r.t. the class of all LAV constraints.
- There are GAV schema mappings that are not uniquely characterizable by any finite set of universal examples w.r.t. the class of all GAV constraints.

CP 2010 Paper (ten Cate, K ..., Tan):

- Necessary and sufficient condition for a GAV schema mapping to be uniquely characterizable by a finite set of universal examples w.r.t. to the class of all GAV constraints.
- Algorithmic criterion for such a unique characterizability of GAV schema mappings.
Unique Characterizations of LAV Mappings

**Theorem:** If $M = (S, T, \Sigma)$ is a LAV schema mapping, then there is a finite set $U$ of universal examples that uniquely characterizes $M$ w.r.t. the class of all LAV constraints.

**Hint of Proof:**
- Let $d_1, d_2, ..., d_k$ be $k$ distinct elements, where $k = \text{maximum arity of the relations in } S$.
- $U$ consists of all universal examples $(I, J)$ with $I = \{ R(c_1,\ldots,c_m) \}$ and $J = \text{chase}_M(\{ R(c_1,\ldots,c_m) \})$, where each $c_i$ is one of the $d_j$’s.
Further Unique Characterizations

**Definition:** (ten Cate, K ... - 2009) Let \( n \) be a positive integer. A schema mapping \( M = (S, T, \Sigma) \) is \( n \)-modular if for every data example \((I,J)\) that does not satisfy \( \Sigma \), there is a sub-instance \( I' \) of \( I \) with \( |\text{adom}(I')| \leq n \) such that \((I',J)\) does not satisfy \( \Sigma \).

**Theorem:** If \( M = (S, T, \Sigma) \) is a \( n \)-modular GLAV schema mapping, then there is a finite set \( U \) of universal examples that uniquely characterizes \( M \) w.r.t. the class of all \( n \)-modular constraints.

**Corollary:** Every self-join-free on the source GLAV schema mapping is uniquely characterizable via universal examples.
Unique Characterizations of GAV Mappings

**Note:** Recall that for the schema mapping specified by the binary copy constraint $\forall x \forall y (E(x,y) \rightarrow F(x,y))$, there is a finite set of universal examples that uniquely characterizes it w.r.t. the class of all GAV constraints.

In contrast,

**Theorem:** Let $\mathcal{M}$ be the GAV schema mapping specified by $\forall x \forall y \forall u \forall v \forall w (E(x,y) \land E(u,v) \land E(v,w) \land E(w,u) \rightarrow F(x,y))$. There is no finite set of universal examples that uniquely characterizes $\mathcal{M}$ w.r.t. the class of all GAV constraints.
Unique Characterizations of GAV Mappings

**Theorem:** Let $\mathbf{M}$ be the GAV schema mapping specified by
\[
\forall x \forall y \forall u \forall v \forall w (E(x,y) \land E(u,v) \land E(v,w) \land E(w,u) \rightarrow F(x,y)).
\]
There is no finite set of universal examples that uniquely characterizes $\mathbf{M}$ w.r.t. the class of all GAV constraints.

**Note:**
- This extends to every GAV schema mapping specified by
  \[
  \forall x \forall y (E(x,y) \land Q_G \rightarrow F(x,y)),
  \]
  where $Q_G$ is the canonical conjunctive query of a graph $G$ containing a cycle.
- The proof uses a generalization, due to Nešetřil and Rödl, of Erdös’ result about the existence of graphs of arbitrarily large girth and chromatic number.
(Non-)Characterizable GAV Mappings

- $E(x,y) \rightarrow F(x,y)$ is uniquely characterizable by these 3 universal examples:

  - $E(x,y) \rightarrow F(x,y)$
  - $E(u,v) \rightarrow F(u,v)$
  - $E(v,w) \rightarrow F(v,w)$
  - $E(w,u) \rightarrow F(w,u)$

In contrast,

- $E(x,y) \land E(u,v) \land E(v,w) \land E(w,u) \rightarrow E(u,v)$ is **not** uniquely characterizable.
Characterizing GAV Schema Mappings

- **Question:**
  - What is the reason that some GAV schema mappings are uniquely characterizable w.r.t. the class of all GAV constraints while some others are not?
  - Is there an algorithm for deciding whether or not a given GAV schema mapping is uniquely characterizable w.r.t. the class of all GAV constraints?

- **Answer:**
  - The answers to these questions are closely connected to database constraints and homomorphism dualities.
Homomorphisms

**Notation:**  $A$, $B$ relational structures (e.g., graphs)

- $A \rightarrow B$ means there is a **homomorphism** $h$ from $A$ to $B$, i.e., a function $h$ from the universe of $A$ to the universe of $B$ such that if $P(a_1,\ldots,a_m)$ is a fact of $A$, then $P(h(a_1),\ldots,h(a_m))$ is a fact of $B$.

  - **Example:** $G \rightarrow K_2$ if and only if $G$ is 2-colorable

- $A \rightarrow A = \{B : B \rightarrow A \}$
  - **Example:** $\rightarrow K_2 = \text{Class of 2-colorable graphs}$

- $A \rightarrow = \{B : A \rightarrow B\}$
  - **Example:** $K_2 \rightarrow = \text{Class of graphs with at least one edge.}$
Homomorphism Dualities

**Definition:** Let $D$ and $F$ be two relational structures
- $(F,D)$ is a **duality pair** if for every structure $A$
  - $A \rightarrow D$ if and only if $(F \leftrightarrow A)$.

  In symbols, $\rightarrow D = F\leftrightarrow$
- In this case, we say that $F$ is an **obstruction** for $D$.

**Examples:**
- For graphs, $(K_2, K_1)$ is a duality pair, since
  - $G \rightarrow K_1$ if and only if $K_2 \leftrightarrow G$.
- **Gallai-Hasse-Roy-Vitaver Theorem (~1965)** for directed graphs
  Let $T_k$ be the linear order with $k$ elements, $P_{k+1}$ be the path with $k+1$ elements. Then $(P_{k+1}, T_k)$ is a duality pair, since for every $H$
  - $H \rightarrow T_k$ if and only if $P_{k+1} \leftrightarrow H$. 

Homomorphism Dualities

- **Theorem** (*König 1936*): A graph is 2-colorable if and only if it contains no cycle of odd length. In symbols, $\rightarrow K_2 = \bigcap_{i\geq 0} (C_{2i+1})$.

- **Definition**: Let $F$ and $D$ be two sets of structures. We say that $(F, D)$ is a **duality pair** if for every structure $A$, TFAE:
  - There is a structure $D$ in $D$ such that $A \rightarrow D$.
  - For every structure $F$ in $F$, we have $F \not\leftrightarrow A$.

In symbols, $\bigcup_{D \in D} (\rightarrow D) = \bigcap_{F \in F} (F \not\leftrightarrow)$. In this case, we say that $F$ is an **obstruction set** for $D$. 
Homomorphism Dualities

Duality Pair \((F,D)\), where

\[ F = \{F_1, F_2, \ldots\} \]
\[ D = \{D_1, D_2, \ldots\} \]

The Yin
“Dreams”:
\[ \cup_i (\rightarrow D_i) \]

The Yang
“Fears”:
\[ \cup_i (F_i \rightarrow) \]
Homomorphism Dualities and Constraint Satisfaction

- **Theorem** (Atserias 2005, Rossman 2005)
  For every structure $D$, TFAE
  - $D$ is first-order definable.
  - $\{D\}$ has a **finite** obstruction set.

- **Theorem** (Feder - Vardi 1993, K ... - Vardi – 1998)
  For every structure $D$, TFAE
  - $D$ is definable in co-Datalog (hence, it is in PTIME).
  - $\{D\}$ has an obstruction set of **bounded treewidth**.
  - $D$ is definable in finite-variable infinitary logic.
  **Illustration:** 2-Colorability
  - $\{C_{2i+1} : i \geq 1\}$ is an obstruction set for $K_2$. 
Unique Characterizations and Homomorphism Dualities

**Theorem:** Let $M = (S, T, \Sigma)$ be a GAV schema mapping. Then the following statements are equivalent:

- $M$ is uniquely characterizable via universal examples w.r.t. the class of all GAV constraints.

- For every target relation symbol $R$, the set $F(M,R)$ of the **canonical structures** of the GAV constraints in $\Sigma$ with $R$ as their head is the obstruction set of some finite set of structures.
Definition:

The **canonical structure** of a GAV constraint

$$\forall x \left( \varphi_1(x) \land \ldots \land \varphi_\kappa(x) \rightarrow R(x_{i_1},\ldots,x_{i_m}) \right)$$

is the structure consisting of the atomic facts $\varphi_1(x), \ldots, \varphi_\kappa(x)$ and having constant symbols $c_1,\ldots,c_m$ interpreted by the variables $x_{i_1},\ldots,x_{i_m}$ in the atom $R(x_{i_1},\ldots,x_{i_m})$.

Let $\mathbf{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ be a GAV schema mapping. For every relation symbol $R$ in $\mathbf{T}$, let $\mathbf{F(M,R)}$ be the set of all canonical structures of GAV constraints in $\Sigma$ with the target relation symbol $R$ in their head.
Canonical Structures

Examples:

- GAV constraint $\sigma$
  
  \[(E(x,y) \land E(y,z) \rightarrow F(x,z))\]
  
  - Canonical structure: $A_\sigma = (\{x,y,z\}, \{(E(x,y),E(y,z)\},x,z)\]
  
  - Constants $c_1$ and $c_2$ interpreted by the distinguished elements $x$ and $z$.

- GAV constraint $\tau$
  
  \[(E(x,y) \land E(z,z) \rightarrow F(x,y))\]
  
  - Canonical structure: $A_\tau = (\{x,y,z\}, \{E(x,y),E(z,z)\},x,y)\]
  
  - Constants $c_1$ and $c_2$ interpreted by the distinguished elements $x$ and $y$.

- GAV constraint $\theta$
  
  \[(E(x,y) \land E(z,z) \rightarrow F(x,x))\]
  
  - Canonical structure: $A_\theta = (\{x,y,z\}, \{E(x,y),E(z,z)\},x,x)\]
  
  - Constants $c_1$ and $c_2$ both interpreted by the distinguished element $x$. 
Theorem: Let $M = (S, T, \Sigma)$ be a GAV schema mapping. Then the following statements are equivalent:

- $M$ is uniquely characterizable via universal examples w.r.t. the class of all GAV constraints.
- For every target relation symbol $R$, the set $F(M, R)$ of the canonical structures of the GAV constraints in $\Sigma$ with $R$ as their head is the obstruction set of some finite set of structures.

Note: For structures $A$ and $B$ with distinguished elements, a homomorphism $h: A \rightarrow B$ maps each distinguished element of $A$ to the corresponding distinguished element of $B$. 
Unique Characterizations and Homomorphism Dualities

**Question:**

- Is there an algorithm to tell when a GAV schema mapping is uniquely characterizable via a finite set of universal examples w.r.t. to the class of all GAV constraints?

Equivalently,

- Is there an algorithm to tell when a finite set of structures with constants is the obstruction set of some finite set of structures with constants?
When do Homomorphism Dualities Exist?

**Theorem** (Foniok, Nešetřil, Tardif – 2008):
Let $F$ be a finite set of relational structures (without constants) consisting of homomorphically incomparable core structures.

- Then the following statements are equivalent:
  - $F$ is an obstruction set of some finite set $D$ of structures.
  - Each structure $F$ in $F$ is “acyclic”.

- Moreover, there is an algorithm that, given such a set $F$ consisting of acyclic structures, computes a finite set $D$ of structures such that $(F, D)$ is a duality pair.
Acyclicity

**Definition:** Let $A = (A, R_1, \ldots, R_m)$ be a relational structure (no constants)
- The *incidence graph* $\text{inc}(A)$ of $A$ is the bipartite graph with
  - nodes the elements of $A$ and the facts of $A$
  - edges between elements and facts in which they occur
- The structure $A$ is **acyclic** if
  - $\text{Inc}(A)$ is an acyclic graph, and
  - No element occurs in the same fact twice.

**Example:**
- $A = (\{1,2,3\}, \{R(1,2,3), P(1)\})$ is acyclic.
- $A = (\{1,2,3\}, \{R((1,2,3), Q(1,2))\})$ is **not** acyclic
  because $1$, $R(1,2,3)$, $2$, $Q(1,2)$, and $1$ form a cycle.
c-Acyclicity

**Definition:** Let $A = (A, R_1, \ldots, r_m, c_1, \ldots c_k)$ be a relational structure with constants $c_1, \ldots, c_k$.

- The **incidence graph** $\text{inc}(A)$ of $A$ is the bipartite graph with
  - nodes the elements of $A$ and the facts of $A$
  - edges between elements and facts in which they occur
- The structure $A$ is **c-acyclic** if
  - Every cycle of $\text{Inc}(A)$ contains at least one constant $c_i$, and
  - Only constants may occur more than once in the same fact.

**Example:**
- $A = (\{1,2,3\}, \{R((1,2,3), Q(1,2), 1)\})$ is c-acyclic
  - the cycle $1, R(1,2,3), 2, Q(1,2), 1$ contains the constant $1$, and it is the only cycle of $\text{inc}(A)$.
- $A = (\{1,2,3\}, \{R((1,2,3), Q(1,2), 3)\})$ is not c-acyclic
  - the cycle $1, R(1,2,3), 2, Q(1,2), 1$ contains no constant.
When do Homomorphism Dualities Exist?

**Theorem:**
Let $F$ be a finite set of relational structures with constants consisting of homomorphically incomparable core structures.
- Then the following statements are equivalent:
  - $F$ is an **obstruction set** of some finite set $D$ of structures.
  - Each structure $F$ in $F$ is **c-acyclic**.
- Moreover, there is an algorithm that, given such a set $F$ consisting of c-acyclic structures, computes a finite set $D$ of structures such that $(F, D)$ is a duality pair.

**Proof:**
A (lengthy) reduction to the Foniok- Nešetřil, Tardif Theorem.
Unique Characterizations and Homomorphism Dualities

**Theorem:** Let \( M = (S, T, \Sigma) \) be a GAV schema mapping such that for every target relation symbol \( R \), the set \( F(M, R) \) of the canonical structures of the GAV constraints in \( \Sigma \) with \( R \) as their head consists of homomorphically incomparable cores. Then the following statements are equivalent:

- \( M \) is **uniquely characterizable via universal examples** w.r.t. the class of all GAV constraints.
- For every target relation symbol \( R \), the set \( F(M, R) \) is the **obstruction set** of some finite set of structures.
- For every target relation symbol \( R \), the set \( F(M, R) \) consists entirely of **c-acyclic** structures.
Applications

- The GAV schema mapping $\mathcal{M}$ specified by
  \[ \forall x \forall y \ (E(x,y) \rightarrow F(x,y)) \]
is uniquely characterizable (the canonical structure is c-acyclic).
- More generally, if $\mathcal{M}$ is a GAV schema mapping specified by a tgd in which all variables in the LHS are exported to the RHS, then $\mathcal{M}$ is uniquely characterizable.
- The GAV schema mapping $\mathcal{M}$ specified by
  \[ \forall x \forall y \forall u \forall v \forall w \ (E(x,y) \land E(u,v) \land E(v,w) \land E(w,u) \rightarrow F(x,y)) \]
is not uniquely characterizable: the canonical structure contains a cycle with no constant on it, namely,
  \[ u, E(u,v), v, E(v,w), w, E(w,u), u \]
- The GAV schema mapping $\mathcal{M}$ specified by
  \[ \forall x \forall y \forall u \ (E(x,y) \land E(u,u) \rightarrow F(x,y)) \]
is not uniquely characterizable.
Let $M$ be the GAV schema mappings specified by the constraints

- $\sigma : \forall x \forall y \forall z (E(x,y) \land E(y,z) \land E(z,x) \rightarrow F(x,z))$
- $\tau : \forall x \forall y (E(x,y) \land E(y,x) \rightarrow F(x,x))$

- The canonical structures of these constraints are
  - $A_\sigma = \{x,y,x\} \{E(x,y), E(y,z), E(z,x)\}, x, z)$
  - $A_\tau = \{x,y\} \{E(x,y), E(y,x)\}, x, x)$

- Both are c-acyclic; hence $\{A_\sigma, A_\tau\}$ is an obstruction set of a finite set of structures.

- Therefore, $M$ is uniquely characterizable via universal examples.
Algorithmic Consequences

**Note:** Every GAV schema mapping $M$ is logically equivalent to one in **normal form**, i.e., to a GAV schema mapping $M^*$ such that for every target relation symbol $R$, the set $F(M^*,R)$ of The canonical structures of the GAV constraints in $\Sigma$ with as their head consists of homomorphically incomparable cores.

**Theorem:** The following problem is NP-complete: Given a GAV schema mapping $M$, is it uniquely characterizable w.r.t. the class of all GAV constraints? The same problem is in LOGSPACE for GAV schema mappings $M$ in normal form.
Synopsis

- Introduced and studied the notion of unique characterization of a schema mapping by a finite set of universal examples.
- Every LAV (n-modal) schema mapping is uniquely characterizable via universal examples w.r.t. the class of all LAV (n-modal) constraints.
- There are GAV schema mappings that are not uniquely characterizable by any set of universal examples w.r.t. the class of all GAV constraints.
- Necessary and sufficient condition, and an algorithmic criterion for a GAV schema mapping to be uniquely characterizable via universal examples w.r.t. the class of all GAV constraints.
- **Open Problem:**
  Unique characterizations of GLAV schema mappings?