uOttawa
L'Université canadienne
Canada's university

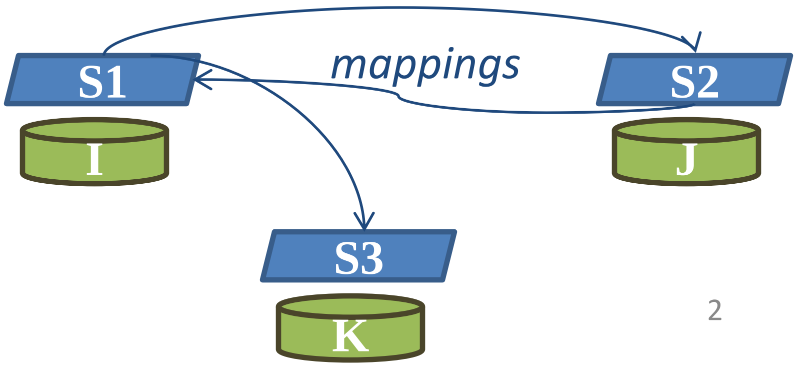
Sharing and Exchanging Data

Rana Awada and Iluju Kiringa
University Of Ottawa
Pablo Barcelo
University of Chile

Introduction

- Data Exchange: Transform an instance ***I*** of a source schema ***S***, according to s-t rules, and generate a target instance ***J*** to conform to a target schema ***T*** and materialize it in the target.
  - d : $\forall \mathbf{x} \, (\varphi_S(\mathbf{x}) \rightarrow \exists \mathbf{y} \Psi_T(\mathbf{x},\mathbf{y}))$

- Data Coordination: Integrate information by propagating updates and by allowing access to information that possibly belong to different sets of vocabularies, using mapping rules, between different sources.
  - d : $\forall \mathbf{x} \, (\varphi_S(\mathbf{x}) \rightarrow \exists \mathbf{y} \Psi_T(\mathbf{x},\mathbf{y}))$
  - *Mapping Tables*
  - Etc..
- ….



Data Exchange

University of Carleton

Student

Sname	Sage
Alex	18

Enroll

Sname	Cid	Cgrade
Alex	ECOR1606	80

??

University of Ottawa

St

Sname	Sage	Address

Take

Sname	Cid	Cgrade

Course

Cid	Cname	Pname
ECOR1606	Introduction to Computers	ENG

Cr

Cid	Cname	Pname

Vocabulary UOC

COMP 4001, COMP1005, 80, 90, …

Vocabulary UOO

CSI1390, CSI4109, A, A+, B, B+ …

Data Coordination

GLAV

University of Carleton

Global Schema

GAV

Student

Sname	Sage
Alex	18

Enroll

Sname	Cid	Cgrade
Alex	ECOR1606	80

Course

Cid	Cname	Pname
ECOR1606	…	ENG

M

S	T
CS	CS
ENG	ENG
ECOR1606	CSI1390

University of Ottawa

Global Schema

GAV

Cr

Cid	Cname	Pname
CSI1390	…	CS

…

M

1 Q_A

2 Q'_A = SELECT Sname, Cid, Cgrade
FROM Enroll
WHERE Pname ='CS'

3 Result Q'_A

Sname	Cid	Cgrade

Data Sharing and Exchange

- Data Sharing and Exchange: Transform an instance ***I*** of a source schema ***S***, according to s-t mappings, and generate a target instance ***J*** that conforms to a target schema ***T*** and to the vocabulary of the target, then materialize it in the target.

- Formally: DSE is a tuple $\vartheta = (\mathbf{S}, \mathbf{T}, M, \sum_{st})$
  - **S** : Source Schema
  - **T**: Target Schema
  - M : Binary relation symbol (not in **S** nor in **T**) with domain in ($Const^S$ x $Const^T$)
  - $\sum_{st}$: $\forall \mathbf{x}\ \forall \mathbf{x'} (\ \varphi_S(\mathbf{x}) \wedge \mu\ (\mathbf{x},\mathbf{x'}) \rightarrow \exists \mathbf{y}\ \Psi_T(\mathbf{x'},\mathbf{y}))$

- DSE Solutions: ?

Motivating Example

Student(x,y) ∧ Enroll(x,z,w) ∧ Course(z,u,`CS`) ∧ M(x,x') ∧ M(y,y') ⟶ ∃z' ST(x',y',z')

Student(x,y) ∧ Enroll(x,z,w) ∧ Course(z,u,`CS`) ∧ M(x,x') ∧ M(z,z') ∧ M(w,w') ⟶ Take(x',z',w')

UOC

Student

Sname	Sage
Alex	18

Enroll

Sname	Cid	Cgrade
Alex	ECOR1606	80

Course

Cid	Cname	Pname
ECOR1606	….	ENG
COMP1005	….	CS

Vocabulary A

COMP 4001, COMP1005, 80, 90

M

S	T
ECOR1606	CSI1390
COMP1005	CSI1390
COMP1005	CSI1790
95	A+
30	D
80	B
18	18
Alex	Alex

UOO

ST

Sname	Sage	Address

Take

Sname	Cid	Cgrade

Cr

Cid	Cname	Pname
CSI1390	….	CS

Vocabulary B

CSI1390, A, A+, B, B+ …

Motivating Example (Cont.)

Student(x,y) ∧ Enroll(x,z,w) ∧ Course(z,u,`CS`) ∧ M(x,x') ∧ M(y,y') ⟶ ∃ z' ST(x',y',z')

Student(x,y) ∧ Enroll(x,z,w) ∧ Course(z,u,`CS`) ∧ M(x,x') ∧ M(z,z') ∧ M(w,w') ⟶ Take(x',z',w')

UOC

Student

Sname	Sage
Alex	18

Enroll

Sname	Cid	Cgrade
Alex	ECOR1606	80
Alex	COMP1005	80

Course

Cid	Cname	Pname
ECOR1606	….	ENG
COMP1005	….	CS

Vocabulary A

COMP 4001, COMP1005, 80, 90

M

S	T
ECOR1606	CSI1390
COMP1005	CSI1390
COMP1005	CSI1790
95	A+
30	D
80	B
18	18
Alex	Alex

UOO

ST

Sname	Sage	Address
Alex	18	Null

Take

Sname	Cid	Cgrade
Alex	CSI1390	B

Cr

Cid	Cname	Pname
CSI1390	….	CS

Vocabulary B

CSI1390, A, A+, B, B+ …

DSE: Knowledge Exchange

- DSE a Knowledge Exchange setting: $\vartheta = (\mathbf{S}, \mathbf{T}, M, \sum_{st})$:
  - Source KB: $(I \cup \{M\}, \sum_s)$
  - Target KB: $(J \cup \{M\}, \sum_t)$

- DSE Solution:
  - For each $K \in Mod(\sum_t (J \cup \{M\}))$, there exists $K' \in Mod(I \cup \{M\})$ such that $K'_M \subseteq K_M$ and K_J is a DSE solution for K'_I and K'_M.

- Universal DSE Solution:
  - For each $K' \in Mod(I \cup \{M\})$ there exists a $K \in Mod(\sum_{\mathbf{t}} (J \cup \{M\}))$, such that $K_M \subseteq K'_M$ and K_J is a DSE solution for K'_I and K'_M.

Motivating Example (Cont.)

Student(x,y) ∧ Enroll(x,z,w) ∧ Course(z,u,`CS`) ∧ M(x,x’) ∧ M(y,y’) ⟶ ∃ z’ ST(x’,y’,z’)

Student(x,y) ∧ Enroll(x,z,w) ∧ Course(z,u,`CS`) ∧ M(x,x’) ∧ M(z,z’) ∧ M(w,w’) ⟶ Take(x’,z’,w’)

UOC

Student

Sname	Sage
Alex	18

Enroll

Sname	Cid	Cgrade
Alex	ECOR1606	80

Course

Cid	Cname	Pname
ECOR1606	….	ENG
COMP1005	….	CS

Vocabulary A

COMP 4001, COMP1005, 80, 90

M

S	T
ECOR1606	CSI1390
ECOR1606	CSI1790
COMP1005	CSI1390
COMP1005	CSI1790
95	A+
30	D
80	B
18	18
Alex	Alex

UOO

ST

Sname	Sage	Address
Alex	18	Null

Take

Sname	Cid	Cgrade
Alex	CSI1390	B
Alex	CSI1790	B

Cr

Cid	Cname	Pname

Vocabulary B

CSI1390, A, A+, B, B+ …

DSE: Knowledge Exchange (Cont.)

- **Source Completion Rules $\sum_s$:**
  - For each $R \in S \cup \{M\}$ of arity n and $1 \leq i \leq n$:
    - $R(x_1, x_i, \ldots, x_n) \rightarrow EQUAL(x_i, x_i)$
  - $EQUAL(x,y) \rightarrow EQUAL(y,x)$
  - $EQUAL(x, z) \wedge EQUAL(z, y) \rightarrow EQUAL(x,y)$
  - $M(x,z) \wedge M(y,z) \rightarrow EQUAL(x,y)$ – *($\sum_t$:$M(z,x) \wedge M(z,y) \rightarrow EQUAL(x,y)$)*
  - **$M(x,y) \wedge EQUAL(x,z) \wedge EQUAL(y,w) \rightarrow M(z,w)$**
  - **For each $R \in S$ of arity n and $1 \leq i \leq n$:**
    - **$R(x_1, x_i, \ldots, x_n) \wedge^{n}_{i=1} EQUAL(x_i, y_i) \rightarrow R(y_1, y_i, \ldots, y_n)$**

Universal DSE Solution:

ST

Sname	Sage	Address
Alex	18	Null

Take

Sname	Cid	Cgrade
Alex	CSI1390	B
Alex	CSI1790	B

Minimal DSE Solution

- Best: Minimal universal DSE solution J:
  1. No proper subset J' of J is a universal DSE solution.
  2. No universal DSE solution J' with a domain (dom(J') $\cap Const^T$) that is properly contained in (dom(J) $\cap Const^T$).

Universal DSE Solution:

ST

Sname	Sage	Address
Alex	18	Null

Take

Sname	Cid	Cgrade
Alex	CSI1390	B
Alex	CSI1790	B

Minimal DSE Solution:

ST

Sname	Sage	Address
Alex	18	Null

Take

Sname	Cid	Cgrade
Alex	CSI1390	B

Minimal DSE Solution (Cont.)

- Exchange Algorithm:
- Input: Source KB ($I \cup \{M\}, \sum_s$)
- Output: Target KB ($J^* \cup \{M\}, \sum_t$) with minimal $dom(J) \cap Const^S$

1. Apply the completion process $\sum_s$ to I and M to generate I_1 and M_1
2. Compute equivalence classes $\{C_1, \ldots, C_n\}$ on $dom(M_1) \cap Const^T$ such that $c_1 \sim c_2$ if $M_1(a, c_1)$ and $M_1(a, c_2)$ hold.
3. Choose a set of witnesses $w_i \in C_i$, $1 \leq i \leq n$, and replace each $c \in C_i \cap dom(M_1)$ with w_i and generate M_2.
4. Apply a procedure based on the *Chase* to $I_1 \cup M_2$ and generate a universal pre-solution J for $I_1 \cup M_2$.
5. Apply a procedure based on the *Core* to J and generate J*.

Motivating Example (Cont.)

UOC

Student

Sname	Sage
Alex	18

Enroll

Sname	Cid	Cgrade
Alex	ECOR1606	80

Course

Cid	Cname	Pname
ECOR1606	….	ENG
COMP1005	….	CS

Vocabulary A

COMP 4001, COMP1005, 80, 90

EQUAL

X	Y
CSI1390	CSI1790
CSI1790	CSI1390
A+	A+
…	…

→

M

S	T
ECOR1606	CSI1390
ECOR1606	CSI1790
COMP1005	CSI1390
COMP1005	CSI1790
95	A+
30	D
80	B
18	18
Alex	Alex

UOO

ST

Sname	Sage	Address
Alex	18	Null

Take

Sname	Cid	Cgrade
Alex	CSI1390	B

Cr

Cid	Cname	Pname
CSI1390	….	CS

Vocabulary B

CSI1390, A, A+, B, B+ …

Query Answering

- DE Setting:
  - Certain(I,Q) = ∩Q(J) for each solution J of I.
- KB setting:
  - Certain (Q, K) = ∩Q(I) for each model I of K.
- DSE Setting:
  - Certain (I ∪{M}, Q) = ∩Q(K) for each model *K* of $\sum_t$(J) , where J is a universal DSE solution of the source KB (I ∪{M}).

Query Answering (Cont.)

- Query:
  - Q(x,u) = ∃vTAKE(x, u,v).

Univesrsal DSE Solution J:

ST

Sname	Sage	Address
Alex	18	Null

Take

Sname	Cid	Cgrade
Alex	CSI1390	B

+

EQUAL

X	Y
CSI1390	CSI1790
CSI1790	CSI1390
A+	A+
B	B
…	…

→

Univesrsal DSE Solution J':

ST

Sname	Sage	Address
Alex	18	Null

Take

Sname	Cid	Cgrade
Alex	CSI1390	B
Alex	CSI1790	B

Q(J')

Sname	Cid
Alex	CSI1390
Alex	CSI1790

Query Answering (Cont.)

- Computing Certain(I ∪{M}, Q) using a Minimal DSE solutions:
  - Let $Q(\mathbf{x}) = \forall\mathbf{x}\exists\mathbf{y}(\ \varphi(\mathbf{x}, \mathbf{y}) \wedge \Psi(\mathbf{y}))$ is such that:
    - $\varphi(\mathbf{x}, \mathbf{y})$ is a conjunction of atomic relations over **T**.
    - $\Psi(\mathbf{y})$ is a conjunction of equality formulas of the form: $y_1 = y_2$.
    - **x** is a set of variables $x_1, .., x_n$
    - **y** is a set of variables $y_1, .., y_n$
  - Rewrite Q to Q' :
    - $Q'(y_1, \ldots y_n) = Q(x_1, \ldots x_n) \bigwedge_{i=1-n} EQUAL(x_i, y_i)$.

Query Answering (Cont.)

- Query re-writing:
  - $Q(x,u) = \exists v TAKE(x, u,v)$.
  - $Q'(x_1, u_1) = \exists v\ (TAKE(x, u,v) \wedge EQUAL(x, x_1) \wedge EQUAL(u, u_1))$.

Minimal DSE Solution J:

ST

Sname	Sage	Address
Alex	18	Null

Take

Sname	Cid	Cgrade
Alex	CSI1390	B

EQUAL

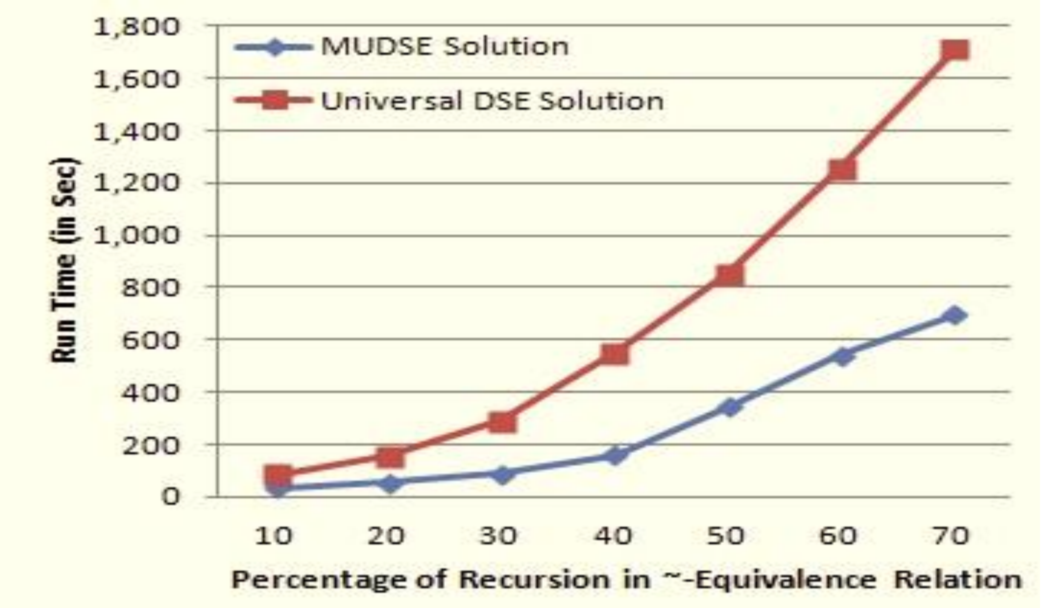
X	Y
ECOR1606	COMP1005
COMP1005	ECOR1606
CSI1390	CSI1790
CSI1790	CSI1390
…	…

Q'(J)

Sname	Cid
Alex	CSI1390
Alex	CSI1790

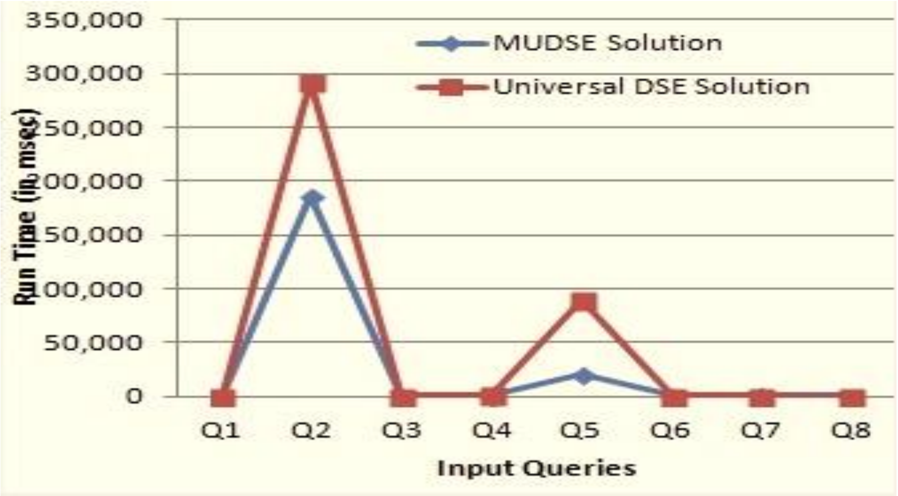
Experimental Results

- Runtime of generating Minimal DSE versus DSE solutions:
  - Source instance I had 4,500 records and a course in the source is mapped to a maximum of two courses in the target.



Experimental Results (Cont.)

- Runtime of Query Answering:



Q1	*Fetch all the students names and the name of courses they have taken*
Q2	*Fetch the list of pairs of students ids and names that took the same course*
Q3	*Fetch all the students names and the grades they have received*
Q4	*Fetch the list of pairs of courses names that belong to the same program*
Q5	*Fetch for each student id the pair of courses that he has finished with the same grade*
Q6	*Fetch all the courses ids and their names*
Q7	*Fetch all the students ids and their names*
Q8	*Fetch the list of pairs of students ids that possess the same address*

Main Results

- Let ϑ = (**S**, **T**, M, $\sum_{st}$) be a fixed DSE setting:
  - Computing a universal DSE solution J for a source instance I and an st-mapping table M is in Logspace.
  - Computing a Minimal DSE solution J for a source instance I and an st-mapping table M is in Logspace.
  - Any two Minimal DSE solutions J_1 and J_2, it is the case that J_1 and J_2 are isomorphic.
  - Given a fixed conjunctive query Q over **T**, the computing certain$(I \cup \{M\},Q)_{\vartheta}$ is in Logspace.

Thank You.