CS4221/CS5421

Tutorial 5: Dependencies, entity-relationship modelling and the Chase

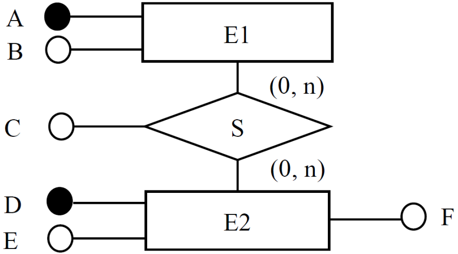
Mark Meng Huasong

School of Computing
National University of Singapore

Week 7, 2022 Spring

Question 1 Entity-relationship Design

Consider the entity-relationship diagram of Figure 1.



(a) Without other knowledge than that captured by the entity-relationship diagram, what are the **functional** and **multi-valued** dependencies?

Question 1 (Cont.)

Solution: The entity-relationship diagram tells us the following functional dependencies.

$\Sigma = \{\{A\} \rightarrow \{B\}, \{D\} \rightarrow \{E, F\}, \{A, D\} \rightarrow \{C\}\}$

Multi-valued dependencies depend on the translation of this design into tables. The entity- relationship diagram and its translation will only result in "not interesting" multi-valued dependencies, those are trivial or correspond to the table or the functional dependencies.

If we canonically translate this design into three tables, we have $\{A\} \twoheadrightarrow \{B\}$ and $\{D\} \twoheadrightarrow \{E, F\}$ and many more (none of them being useful for normalisation), for instance.

Question 1 (Cont.)

Note that additional knowledge of the application could tell us additional functional and multivalued dependencies not captured by the design.

For instance, we could know that $\{E\} \rightarrow \{F\}$ (which would suggest that the entity-relationship design is probably missing entities and relationships that have been merged too early), which we could use to produce a design in the Boyce-Codd normal by splitting the table $R_3(D, E, F)$ into two tables $R_{3.1} = (\underline{D}, E)$ and $R_{3.2} = (\underline{E}, F)$.

For instance, we could know that $\{E\} \twoheadrightarrow \{F\}$, which we could use to produce a design in the fifth normal by splitting the table $R_3(D, E, F)$ into two tables $R_{3.1} = (\underline{D}, E)$ and $R_{3.2} = (\underline{E, F})$.

Question 2 MVD

Consider the relational schema $R = \{A, B, C, D, E\}$ with the following set of functional and multi-valued dependencies.

$\Sigma = \{\{C\} \rightarrow \{A\}, \{D\} \rightarrow \{D, B\}, \{B\} \rightarrow \{E\}, \{E\} \twoheadrightarrow \{A, D\}, \{A, B, D\} \rightarrow \{A, B, C, D\}, \{B\} \rightarrow \{D\}\}$

(a) Prove that $\{E\} \rightarrow \{D\}$ using the Armstrong and multi-valued dependencies axioms.

Question 2 MVD (Cont.)

Quick Recap:

Complementation $(X \twoheadrightarrow Y) \Longrightarrow (X \twoheadrightarrow R - X - Y)$

Augmentation $((X \twoheadrightarrow Y) \wedge (V \subset W)) \Longrightarrow (X \cup W \twoheadrightarrow Y \cup V)$

Transitivity $((X \twoheadrightarrow Y) \wedge (Y \twoheadrightarrow Z)) \Longrightarrow (X \twoheadrightarrow Z - Y)$

Replication/Promotion $(X \rightarrow Y) \Longrightarrow (X \twoheadrightarrow Y)$

Coalescence
$((X \twoheadrightarrow Y) \wedge (W \rightarrow Z) \wedge (Z \subset Y) \wedge (W \cap Y = \emptyset)) \Longrightarrow (X \rightarrow Z)$

Union $((X \twoheadrightarrow Y) \wedge (X \twoheadrightarrow Z)) \Longrightarrow (X \twoheadrightarrow Y \cup Z)$

Intersection $((X \twoheadrightarrow Y) \wedge (X \twoheadrightarrow Z)) \Longrightarrow (X \twoheadrightarrow Y \cap Z)$

Difference $((X \twoheadrightarrow Y) \wedge (X \twoheadrightarrow Z)) \Longrightarrow (X \twoheadrightarrow Y - Z)$

Question 2 MVD (Cont.)

Solution:

1. We know that $\{E\} \twoheadrightarrow \{A, D\}$.
2. We know that $\{B\} \rightarrow \{D\}$.
3. We see that $\{D\} \subset \{A, D\}$.
4. We see that $\{B\} \cap \{A, D\} = \emptyset$
5. Therefore $\{E\} \rightarrow \{D\}$ by Coalescence of (1), (2), (3) and (4).

Q.E.D.

Try the same question with the Chase (answer not provided).

Question 3 Chase

Consider the relation $R(A, B, C, D, E, G)$ with the following set, F, of functional and multi-valued dependencies.

$F = \{\{A, B\} \rightarrow \{C\}, \{A, B\} \twoheadrightarrow \{E\}, \{C, D\} \twoheadrightarrow \{A, B\}\}$

Prove that the decomposition of R into $R_1(A, B, C, D, G)$ and $R_2(C, D, E)$ is lossless using the Chase algorithm (as shown in the lecture).

How to solve this question?

Recall the lecture note "*Testing if a decomposition is lossless*". We first analyze the decomposition to find out what is the X of "$R = X \cup Y \cup Z$". We have $X = \{C, D\}$.

Then we need to use the Chase method to chase $\{C, D\} \twoheadrightarrow \{E\}$ (or $\{C, D\} \twoheadrightarrow \{A, B, G\}$ equivalently, according to Complementation rule of Armstrong Axiom).

Now the question becomes:
$\{\{A, B\} \rightarrow \{C\}, \{A, B\} \twoheadrightarrow \{E\}, \{C, D\} \twoheadrightarrow \{A, B\}\} \models \{C, D\} \twoheadrightarrow \{E\}$?.

Question 3 (Cont.)

Solution:

1. Initial table.

A	B	C	D	E	G
a1	b1	c1	d1	e1	g1
a2	b2	c2	d2	e2	g2

2. We want to chase $\{C, D\} \twoheadrightarrow \{E\}$, make C and D values the same.

A	B	C	D	E	G
a1	b1	**c1**	**d1**	e1	g1
a2	b2	**c1**	**d1**	e2	g2

Question 3 (Cont.)

3. Apply $\{C, D\} \twoheadrightarrow \{A, B\}$ by copying two tuples that have the same C and D values but swapping their A and B values.

Below are cited from the "Chasing the Chase" part of the lecture note:

For each multi-valued dependency $Z \twoheadrightarrow V \in \Sigma$, if there are 2 tuples in the table with the ***SAME** Z**-value***, *then add two new tuples with* ***all the same values and except*** *for their V-values that are* ***SWAPPED***.

A	B	C	D	E	G
a1	b1	c1	d1	e1	g1
a2	b2	c1	d1	e2	g2
a2	**b2**	c1	d1	e1	g1
a1	**b1**	c1	d1	e2	g2

Question 3 (Cont.)

4. Apply $\{A, B\} \twoheadrightarrow \{E\}$ by copying two tuples that have the same A and B values but swapping their A and E values.

A	B	C	D	E	G
a1	b1	c1	d1	e1	g1
a2	b2	c1	d1	e2	g2
a2	b2	c1	d1	e1	g1
a1	b1	c1	d1	e2	g2
a1	b1	c1	d1	**e2**	g1
a2	b2	c1	d1	**e1**	g2
a2	b2	c1	d1	**e2**	g1
a1	b1	c1	d1	**e1**	g2

Question 3 (Cont.)

5. Applying $\{A, B\} \rightarrow \{C\}$ do not change the table.

Because we cannot find for any two tuples $\{a1, b1, c1, ...\}$ *and* $\{a2, b2, c2, ...\}$ *s.t.* $\{a1, b1\} = \{a2, b2\}$ *but* $c1 \neq c2$.

6. Sort the table and proved that $\{C, D\} \twoheadrightarrow \{E\}$

A	B	C	D	E	G
a1	b1	c1	d1	e1	g1
a1	b1	c1	d1	e2	g1
a1	b1	c1	d1	e1	g2
a1	b1	c1	d1	g2	g2
a2	b2	c1	d1	e1	g1
a2	b2	c1	d1	e2	g1
a2	b2	c1	d1	e1	g2
a2	b2	c1	d1	g2	g2

Refer to the lecture slides "Dependencies" page 77 for the definition of a MVD.

Q.E.D

Question 4 Chase

Consider the relation $R(A, B, C, D, E, F, G)$ with the following set, Σ, of functional dependencies.

$$\Sigma = \{\{A, B\} \rightarrow \{C\}, \{C\} \rightarrow \{D, E\}, \{E\} \rightarrow \{D\}, \{F\} \rightarrow \{G\}\}$$

Prove that the decomposition of R into $R_1(A, B, C, D, E)$ and $R_2(A, B, F, G)$ is lossless using the Chase algorithm (as shown in the lecture).

Solution:

1. Initial table.

A	B	C	D	E	F	G
a1	b1	c1	d1	e1	f1	g1
a2	b2	c2	d2	e2	f2	g2

2. We want to chase $\{A, B\} \twoheadrightarrow \{C, D, E\}$, make A and B values the same.

A	B	C	D	E	F	G
a1	**b1**	c1	d1	e1	f1	g1
a1	**b1**	c2	d2	e2	f2	g2

Question 4 (Cont.)

3. Apply $\{A, B\} \to \{C\}$, make C with the same A and B values the same.

A	B	C	D	E	F	G
a1	b1	**c1**	d1	e1	f1	g1
a1	b1	**c1**	d2	e2	f2	g2

4. Apply $\{C\} \to \{D, E\}$, make D and E with the same C values the same.

A	B	C	D	E	F	G
a1	b1	c1	**d1**	**e1**	f1	g1
a1	b1	c1	**d1**	**e1**	f2	g2

5. Applying $\{E\} \to \{D\}$ does not change the table.
6. Applying $\{F\} \to \{G\}$ does not change the table.
7. Proved that $\{A, B\} \twoheadrightarrow \{C, D, E\}$.

Q.E.D

Question 5 Chase with the Distinguished Attributes

Consider the relation $R(A, B, C, D, E)$ with the following set, F, of functional dependencies.

$$\Sigma = \{\{A\} \rightarrow \{B, C\}, \{B\} \rightarrow \{A\}, \{C\} \rightarrow \{D\}\}$$

Check whether the decomposition of R into $R_1(A, E)$, $R_2(C, D)$ and $R_3(A, B, C)$ is lossless using the Chase algorithm with the distinguished attributes.

Solution:

1. Initial table.

	A	**B**	**C**	**D**	**E**
R_1	a				a
R_2			a	a	
R_3	a	a	a		

2. Apply $\{A\} \rightarrow \{B, C\}$.

	A	**B**	**C**	**D**	**E**
R_1	a	**a**	**a**		a
R_2			a	a	
R_3	a	a	a		

Question 5 (Cont.)

3. Applying $\{B\} \rightarrow \{A\}$ does not change the table.
4. Apply $\{C\} \rightarrow \{D\}$.

	A	**B**	**C**	**D**	**E**
R_1	a	a	a	**a**	a
R_2			a	a	
R_3	a	a	a	**a**	

5. R_1 has distinguished variables in all of the columns, therefore the decomposition is lossless.

Q.E.D

Extra Practice (A): MVD & Chase

Q2 Revisit: consider the relational schema $R = \{A, B, C, D, E\}$ with the following set of functional and multi-valued dependencies.

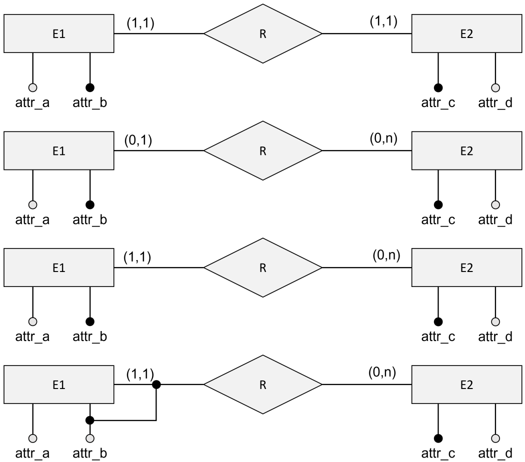
$$\Sigma = \{\{C\} \rightarrow \{A\}, \{D\} \rightarrow \{D, B\}, \{B\} \rightarrow \{E\}, \{E\} \twoheadrightarrow \{A, D\}, \{A, B, D\} \rightarrow \{A, B, C, D\}, \{B\} \rightarrow \{D\}\}$$

Prove that $\{E\} \rightarrow \{D\}$ using the **Chase algorithm**.

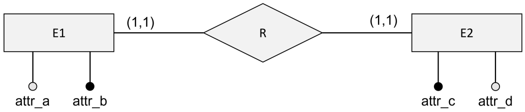
Q3-4 Revisit: prove them by using Armstrong axiom.

Extra Practice (B): ER Participation vs. Logical Design

Can you spot the difference among the 4 cases below? Can you explain what they look like in the logical design? Can you give an example from the real world for each case?



Extra Practice (Case 1)



E1 as **political_party**,
E2 as **secretary_general**,
R as **led_by**

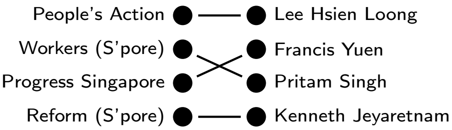
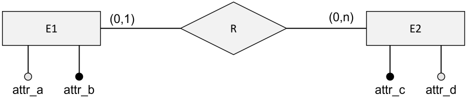


Table E1:
🔑 name,
secretary_general

or alternatively, we define

Table E2:
🔑 name
leading_party,

Extra Practice (Case 2)



E1 as **parliament_member**,
E2 as **political_party**,
R as **belongs_to**

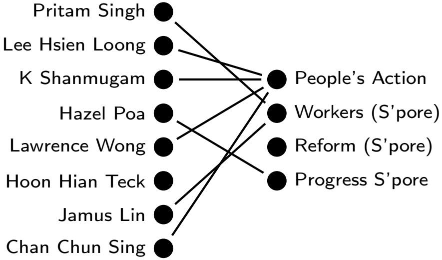
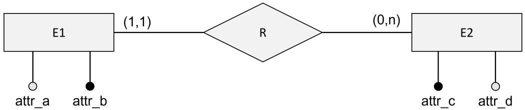


Table E1:
name

Table E2:
name

Table R:
→ member_name (E1.name),
→ affiliation_party (E2.name)

Extra Practice (Case 3)



E1 as **government**,
E2 as **parliament_member**,
R as **head**

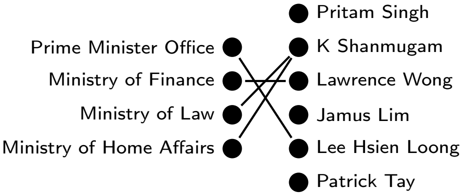


Table E1:

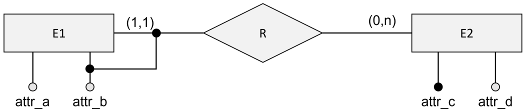
government_name

→ name (E2.name)

Table E2:

name

Extra Practice (Case 4)



E1 as **party**,
E2 as **country**,
R as **belongs_to**

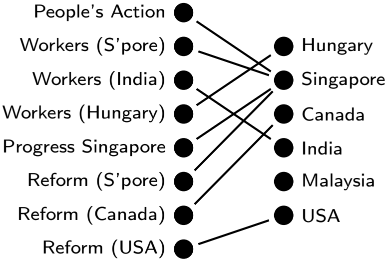


Table E1:
party_name
→ country (E2.name)

Table E2:
name

For any further question, please feel free to email me:

huasong.meng@u.nus.edu

Cases in the extra practice are contributed by our students.