LOGICAL DATABASE DESIGN Part #1/2

Functional Dependencies

• Informally, a FD appears when the values of a set of attributes uniquely determines the values of another set of attributes.
  
  Example: schedule (PILOT, FLIGHT#, DATE, TIME)
  1. TIME functionally depends upon FLIGHT#
  2. PILOT functionally depends upon {FLIGHT#, DATE}
  3. FLIGHT# functionally depends upon {PILOT, DATE, TIME}

  OR
  1. FLIGHT# uniquely determines TIME
  2. {FLIGHT#, DATE} uniquely determines PILOT
  3. {PILOT, DATE, TIME} uniquely determines FLIGHT#

  OR
  1. FLIGHT# → TIME
  2. {FLIGHT#, DATE} → PILOT
  3. {PILOT, DATE, TIME} → FLIGHT#

  Note, the "→" sign is read "functionally determines"

• Formal Definition:

  Let r(R), X ⊆ R, and Y ⊆ R.

  Let s(X) = Π_X(r) and q(Y) = Π_Y(r).
There is a FD $X \rightarrow Y$ for every state of $r$ if the mapping $f : s(X) \rightarrow q(Y)$ is a function.

- In algebraic sense, $r(R)$ satisfies the FD $X \rightarrow Y$, if $\Pi_Y(\sigma_{X=X'}(r))$ always produces at most one tuple. (Note, $X'$ is values of the set of attributes $X$.)

- **Properties of FDs**
  1. If $X \rightarrow Y$, then $\forall t_1, t_2 \in r(R), t_1[Y]=t_2[Y]$ if $t_1[X]=t_2[X]$.
  2. FD is a time invariant property of a relation $r$, i.e. it holds for all states of $r$.

- **Discussion**
  - If $X$ is a key of $r(R)$, then for arbitrary $Y \subseteq R$, $r$ satisfies the FD $X \rightarrow Y$. 

Dr. Byunggu Yu
- If the relationship cardinality between $\Pi_X(R)$ and $\Pi_Y(r)$ is many-to-one, then $r$ satisfies $X \rightarrow Y$ (if the relationship is 1-to-1, then $Y \rightarrow X$ holds too.)

- Looking at the current contents of a relation, one cannot prove the existence of a FD; but it is sometimes possible to say which FDs are not satisfied.

- FDs are statements about the real world made by the DBA in accordance with the semantics of the attributes and the reality which is modeled.

- FDs are axioms about the world which cannot be proven.

- Most DBMSs preserve the FDs of the form Key $\rightarrow$ Attributes, but some systems are capable of enforcing all FDs specified by the DBA.

**Trivial FD and Nontrivial FD**

supplies (S_NAME, S_ADDR, PART#, PRICE)

- Nontrivial FDs:
  1. S_NAME $\rightarrow$ ADDR (if only one address is important)
  2. \{S_NAME, PART#\} $\rightarrow$ PRICE
- Trivial FDs:
  \( S\_\text{NAME} \rightarrow S\_\text{NAME} \)
  \( \{ S\_\text{NAME}, \text{ADDR} \} \rightarrow S\_\text{NAME} \)
  \( \{ S\_\text{NAME}, \text{ADDR} \} \rightarrow \text{ADDR} \)
  \( \{ S\_\text{NAME}, \text{PART}\# \} \rightarrow S\_\text{NAME} \)
  \( \{ S\_\text{NAME}, \text{PART}\# \} \rightarrow \text{PART}\# \)

  Note, \( A \rightarrow B \) is trivial if and only if \( B \subseteq A \)

- From Nontrivial FDs 1 and 2, one can derive
  \( \{ S\_\text{NAME}, \text{PART}\# \} \rightarrow \{ \text{ADDR}, \text{PRICE} \} \)

  Note, \( \{ S\_\text{NAME}, \text{PART}\# \} \) is a key of the relation "supplies" and, therefore, it functionally determines all other attributes in the relation.

**Logical Implications**

- Suppose that \( A \rightarrow B \) and \( B \rightarrow C \) are FDs on a schema \( R=ABC \). Then, \( A \rightarrow C \) is also an FD on \( R \).

  **Proof**

  Let \( r(R) \) be a relation that satisfies FDs \( A \rightarrow B \) and \( B \rightarrow C \). Suppose that \( \exists \ t_1, t_2 \in r(R), \) such that \( t_1[A] = t_2[A] \) and \( t_1[C] \neq t_2[C] \).
If \( t_1[B] \neq t_2[B] \) then \( A \rightarrow B \) does not hold. Otherwise, \( B \rightarrow C \) does not hold.

By contradiction, we conclude that, for no two \( t_1, t_2 \in r(R) \), both \( t_1[A] = t_2[A] \) and \( t_1[C] \neq t_2[C] \). This is necessary and sufficient condition to uphold \( A \rightarrow C \).

Let \( F \) be a set of FDs on \( R \) and let \( X \rightarrow Y \) be an FD on \( R \) which may or may not be in \( F \).

We say that \( F \) logically implies \( X \rightarrow Y \), denoted by \( F \models X \rightarrow Y \), if every relation \( r(R) \) that satisfies FDs in \( F \) also satisfies \( X \rightarrow Y \) (i.e., for no two \( t_1, t_2 \in r(R) \), both \( t_1[X] = t_2[X] \) and \( t_1[Y] \neq t_2[Y] \)).

Example:
\[ F = \{ A \rightarrow B, B \rightarrow C \} \models A \rightarrow C \]

**Closure**

- Let \( F \) be a set of FDs on a schema \( R \). The closure \( F^+ \) of the set \( F \) is the set of all FDs logically implied by \( F \), i.e.
\[ F^+ = \{ X \rightarrow Y \mid F \models X \rightarrow Y \} \]

Example:
\[ F = \{ A \rightarrow B, B \rightarrow C \} \]
\[ F^+ = \{ A \rightarrow B, B \rightarrow C, A \rightarrow C, \]
AB→C, ABC→C, AB→BC, BC→B,
A→∅, B→∅, C→∅, ...

• **Candidate Key**
  - Let r(R=A1A2...An) be a relation for which the set of FDs F is satisfied. Let X⊆R.
  
  We say that X is a candidate key for r(R) if
  1. X→A1A2...An ∈ F+
  2. For no Y⊂X, Y→A1A2...An ∈ F+

  - X is a superkey if only 1 is satisfied.
  - But how can we check the condition 1? ➔ Use X+ computation or Armstrong Axioms

• **Closure of a set of attributes**
  - Let X be a set of attributes s.t. X ⊆ R and let F be a set of FDs on R. The closure of X under F, denoted by X+, is the set of all the attributes that are determined by X under F, i.e.
  \[ X^+ = \{ A \mid X \rightarrow A \} \]

  **X+ Computation**

• **Complexity of the algorithm is proportional to the cardinality of the set of FDs, F.**

• **Algorithm:**
  
  Input: A relation schema R, set of FDs F, and an attribute set X⊆R.
Output: $X^+$, i.e. the closure of $X$ under $F$.

Method:

begin

1. $X^{(0)}=X$
2. $X^{(i+1)}=X^{(i)} \cup \{ A \mid \exists Y,Z (Y \rightarrow Z \in F \land A \in Z \land Y \subseteq X^{(i)}) \}$
3. repeat step 2 until no more changes

end

- **Example 1 of 2:**

- $F = \{ AB \rightarrow C, C \rightarrow A, BC \rightarrow D, ACD \rightarrow B,$
  $D \rightarrow EG, BE \rightarrow C, CG \rightarrow BD, CE \rightarrow AG \}$
- $X = BD$
- begin. $X^{(0)}=BD$

  $X^{(1)}=BD \cup \{ EG \mid D \rightarrow EG \} = BDEG$

  $X^{(2)}=BDEG \cup \{ C \mid BE \rightarrow C \} = BCDEG$

  $X^{(3)}=BCDEG \cup \{ ABDG \mid CE \rightarrow AG, CG \rightarrow BD, BC \rightarrow D, C \rightarrow A \}$

  $X^{(4)}=ABCDEG \cup \{ ABCDEG \mid \ldots \} = ABCDEG$

  $X^{(4)} = X^{(3)}$ end!

- $X^+ = BD^+ = ABCDEG$

- **Example 2 of 2**

  - $F = \{ A \rightarrow C, B \rightarrow DC, AC \rightarrow B \}$. Is $A \rightarrow BD$?
  - $X=A$
  - begin. $X^{(0)}=A$

    $X^{(1)}=AC$

    $X^{(2)}=ABC$

    $X^{(3)}=ABCD$
\[ X^{(4)} = ABCD \]
\[ X^{(4)} = X^{(3)} \text{ end!} \]

- \[ X^+ = A^+ = ABCD \implies BD \subseteq A^+ \implies A \rightarrow BD. \]

**Armstrong Axioms**

- **Sound and complete set of rules for the derivation of FDs in** \( F^+ \), **given a set of FDs** \( F \).

- **Primary axioms**
  - Let \( R \) be a set of attributes, and let \( F \) be the set of FDs on \( R \).
A1. Reflexivity: If $Y \subseteq X \subseteq R$, then $F^{A1} = X \rightarrow Y$ (trivial FDs).

A2. Additivity: $X \rightarrow Y \overset{A2}{=} XZ \rightarrow YZ$ for every $Z \subseteq R$.

A3. Transitivity: $\{X \rightarrow Y, Y \rightarrow Z\} \overset{A3}{=} X \rightarrow Z$ for $X,Y,Z \subseteq R$.

Example: address (STREET, CITY, ZIP)

$F = \{ZIP \rightarrow CITY, \{STREET,CITY\} \rightarrow ZIP\}$

Prove that $\{ZIP,STREET\}$ is a superkey of the relation "address"!

1. $ZIP \rightarrow CITY$ (given)
2. $\{ZIP,STREET\} \rightarrow \{STREET,CITY\}$ (A2. additivity)
3. $\{STREET,CITY\} \rightarrow ZIP$ (given)
4. $\{STREET,CITY\} \rightarrow \{STREET,CITY,ZIP\}$ (A2. additivity)
5. $\{ZIP,STREET\} \rightarrow \{STREET,CITY,ZIP\}$ (A3. transitivity from 2 and 4)

$\Rightarrow \{ZIP,STREET\}$ is a superkey.

- **Secondary axioms**
  - Can be derived from the primary axioms

A4. Union: $\{X \rightarrow Y, X \rightarrow Z\} \overset{A4}{=} X \rightarrow YZ$.

A5. Pseudotransitivity: $\{X \rightarrow Y, WY \rightarrow Z\} \overset{A5}{=} WX \rightarrow Z$.

A6. Decomposition: If $X \rightarrow Y$ and $Z \subseteq Y$, then $X \rightarrow Z$. 
- The derivation of the secondary rules:

A4. \( F = \{ X \rightarrow Y, X \rightarrow Z \} \) (given)

1. \( X \rightarrow Y \) \[ A2 \]
2. \( X \rightarrow Z \) \[ A2 \]
3. \( \{ X \rightarrow XY, XY \rightarrow YZ \} \) \[ A3 \]

A5. \( F = \{ X \rightarrow Y, WY \rightarrow Z \} \) (given)

1. \( X \rightarrow Y \) \[ A2 \]
3. \( \{ WX \rightarrow WY, WY \rightarrow Z \} \) \[ A3 \]

A6. \( F = \{ X \rightarrow Y \}, Z \subseteq Y \) (given)

1. \( Z \subseteq Y \) \[ A1 \]
3. \( \{ X \rightarrow Y, Y \rightarrow Z \} \) \[ A3 \]

- Important consequence of A4 and A6:

\( X \rightarrow A_1A_2...A_n \) iff \( X \rightarrow A_i, i=1,2,...,n \).

- To prove that \( X \subseteq R \) is a superkey of a relation \( r(R) \), it is sufficient to show that \( \forall A_i \in R, X \rightarrow A_i \).

**Properties**

- **Soundness:** The axioms do not generate any incorrect FD, i.e. if \( X \rightarrow Y \) can be logically derived from \( F \) by Armstrong rules, then \( X \rightarrow Y \) holds.

- **Completeness:** The axioms allow us to find all FDs in \( F^+ \).
• Examples: $r(ABC), F=\{AB \rightarrow C, C \rightarrow B\}$
  
  - Prove $AC \rightarrow A$
    
    $AC \subseteq AC \quad \vdash AC \rightarrow A$
    
  - Prove $AC \rightarrow B$
    
    1. $C \rightarrow B$  (given)
    2. $AC \rightarrow AB$  (A2)
    3. $AC \rightarrow B$  (A6)

  - Prove $AB \rightarrow ABC$
    
    1. $AB \rightarrow C$  (given)
    2. $AB \rightarrow ABC$  (A2)

  - Prove $AB \rightarrow BC$
    
    1. $AB \rightarrow C$  (given)
    2. $AB \rightarrow BC$  (A2)