Tutorial 7:

Schema Normalization,
Functional Dependencies
Outline

- Schema design
  - Introduction
  - Notation
- Functional dependencies
- Armstrong’s axioms
- Minimal cover
- Keys
- Questions
Which design is better?
• Disadvantages of one table:
  – Redundant storage
    • Harder to update
    • Consistency issues
  – How should one represent a customer that has not ordered any book?
• Good design:
  – No duplications of data
  – Enables simple updates
  – Simple
  – Not too many tables

• We saw a way to design DB’s
  – ERD – has its limitations…

• This lesson we will focus on an alternative way
  – Functional dependencies (FD’s)
Outline

• Schema design
  – Introduction
  – Notation
• Functional dependencies
• Armstrong’s axioms
• Minimal cover
• Keys
• Questions
Notations

• Attributes - A, B, C, ...
• Sets of attributes – X, Y, ...
• We will often replace
  – \{A\} with A
  – \{A, B\} with AB
• Relational Schemas – R, S, T, ...
  – R(A, B, C) or R={A, B, C} or R[A, B, C]
• The content of the relations – r, s, t
  – r={(1,2,3), (2,1,4)}
• Set of functional dependencies – F
  – A single functional dependency - f
Functional Dependencies - definitions

• Let
  – $R = \{A_1, \ldots, A_n\}$ be a relational schema, let
  – $r$ be a relation over $R$ and let
  – $X, Y \subseteq R$ be sets of attributes

• $r$ is said to satisfy the functional dependency $X \rightarrow Y$ if
  – every two tuples that have the same value in $X$ have the same values in $Y$
  – We denote this by $r \models X \rightarrow Y$
Functional Dependencies – definitions

• Let
  – R be a relational schema and let
  – r be a relation over R and let
  – F be a set of FD’s over R

• r satisfies F (r ⊧ F) if
  – for every f in F, r satisfies f (r ⊧ f)

• f is entailed from F (F ⊧ f) if
  – for every relation r over R it holds that if r ⊧ F
    then r ⊧ f
Outline

• Schema design
  – Introduction
  – Notation
• Functional dependencies
• Armstrong’s axioms
• Minimal cover
• Keys
• Questions
Armstrong’s Axioms

• Three axioms for proving existence of dependencies. Given $X, Y, Z \subseteq R$:
  – Reflexivity: if $X \subseteq Y$ then $Y \rightarrow X$
  – Inclusion: if $X \rightarrow Y$ then $XZ \rightarrow YZ$
  – Transitivity: if $X \rightarrow Y$ and $Y \rightarrow Z$ then $X \rightarrow Z$

• Other modus ponens:
  – Union: if $X \rightarrow Y$ and $X \rightarrow Z$ then $X \rightarrow YZ$
  – Decomposition: if $X \rightarrow Y$ and $Z \subseteq Y$ then $X \rightarrow Z$
  – Semi-transitivity: if $X \rightarrow Y$ and $WY \rightarrow Z$ then $WX \rightarrow Z$
Armstrong’s Axioms

• Let F be a set of functional dependencies and let f be a FD.
  – f is provable from F ($F \vdash f$) if f can be deduced from F by Armstrong’s axioms.
  – That is, we can formally prove f from F
Armstrong’s Axioms - example

• $F = \{\text{Cust}_\text{Id} \to \text{Track}, \text{Track} \to \text{Faculty}\}$
• We show that
  – $F \vdash \text{Cust}_\text{Id} \to \{\text{Track, Faculty}\}$

1. Track $\to$ Faculty $\in F$
2. Track $\to \{\text{Track, Faculty}\}$ Inclusion, 1
3. Cust_Id $\to$ Track $\in F$
4. Cust_Id $\to \{\text{Track, Faculty}\}$ 2,3,Trans.
Outline

• Schema design
  – Introduction
  – Notation
• Functional dependencies
• Armstrong’s axioms
• Minimal cover
• Keys
• Questions
Closure of FD’s

• Let F be a set of FD’s
  – The closure of F (F+) is the set
    \[ \{ X \rightarrow Y \mid F \not\models X \rightarrow Y \} \]

• Example: F=\{A \rightarrow B, B \rightarrow C\}, the following FD’s are in F+:
  – A \rightarrow C, AB \rightarrow C, AC \rightarrow C, B \rightarrow B, A \rightarrow B, \emptyset \rightarrow \emptyset, C \rightarrow \emptyset
  – Note that F+ contains also other FD’s

• Note that the set F+ is exponential and therefore we will try to avoid computing it.
Closure of a property

• Let
  – X be a set of properties and let
  – F be a set of FD’s

• The closure of X with respect to F ($X^+_F$) is the set \{A | F ⊬ X → A\}
  – Note that A is a single attribute
A set of FD’s might have “redundant” information, for instance the sets

- \( F = \{A \rightarrow B, B \rightarrow C, A \rightarrow C\} \)
- \( G = \{A \rightarrow B, B \rightarrow C\} \)

are equivalent in the sense that \( F^+ = G^+ \)

- The dependency \( A \rightarrow C \) is redundant

Our goal: A unified form for FD’s
Minimal set of FD’s

• Let F be a set of FD’s, F is minimal if for every FD $X \rightarrow Y \in F$ the following hold:
  – $|Y| = 1$
  – $F^+ \neq (F \setminus \{X \rightarrow Y\})^+$
  – For all $Z \subset X$ the following hold:
    • $F^+ \neq [(F \setminus \{X \rightarrow Y\}) \cup \{Z \rightarrow Y\}]^+$
    • i.e., F does not contain an FD $X \rightarrow A$ such that X includes redundant attributes
Minimal Cover

• Let F and G be sets of FD’s
  – G is a cover for F (and vice versa) if $F^+ = G^+$
  – In this case we can use F instead of G

• A set of FD’s $F_C$ is a minimal cover of F if
  – It is a cover for F
  – It is minimal

• Note that there might be more than one minimal cover
Algorithm for finding a minimal cover

Let F be a set of FD’s we define
  \[ G \leftarrow \{(X \rightarrow A) \mid \exists Y ((X \rightarrow Y) \in F \land A \in Y)\}; \]

Repeat:
1. For each \( f = X \rightarrow A \in G \) do:
   - if \( A \in X^{+} _{G \setminus \{f\}} \) then \( G \leftarrow G \setminus \{f\}; \)
2. For each \( f = X \rightarrow A \in G \) and \( B \in X \) do
   - if \( A \in (X \setminus \{B\})^{+} _{G} \) then
     \[ G \leftarrow (G \setminus \{X \rightarrow A\}) \cup \{ X \setminus B \rightarrow A\}; \]

Until no more changes to \( G \)
Finding a minimal cover - example

- Let
  - \( R = \{A, B, C, D\} \)
  - \( F = \{A \rightarrow B, BC \rightarrow A, ABC \rightarrow D, D \rightarrow A\} \)
- Find a minimal cover of \( F \)

1\(^{st}\) stage \( G \leftarrow F \)

- Step 1: no change
- Step 2: We omit the \( A \) from the FD \( ABC \rightarrow D \) (since \( D \in (\{A, B, C\}\setminus\{A\})^+_G \) and obtain
  - \( G = \{A \rightarrow B, BC \rightarrow A, BC \rightarrow D, D \rightarrow A\} \)

- Step 1: We omit the FD \( BC \rightarrow A \) (since \( A \in BC^+_G \setminus \{BC \rightarrow A\} \)) and obtain
  - \( G = \{A \rightarrow B, BC \rightarrow D, D \rightarrow A\} \)
- There are no more changes and therefore \( G \) is the minimal cover of \( F \).
Outline

• Schema design
  – Introduction
  – Notation
• Functional dependencies
• Armstrong’s axioms
• Minimal cover
• Keys
• Questions
• Let
  – $R$ be a relational schema and let
  – $X \subseteq R$ be a subset of attributes and let
  – $F$ be a set of FD’s

• $X$ is a superkey of $R$ if and only if $F \models X \rightarrow R$, or equivalently:
  – $X$ is a superkey of $R$ iff $X \rightarrow R \in F^+$
  – $X$ is a superkey of $R$ iff $X^+_F = R$
Keys – cont’d

• Let $R=\{A, B, C, D\}$ and $F=\{A \rightarrow C, B \rightarrow D\}$
  – $ABC$ is a superkey of $R$
    • However, it is not unique
    • And not minimal

• $X$ is a key of $R$ if
  – It is a superkey of $R$
  – There does not exist $Y \subset X$ such that $Y$ is also a superkey
Outline

• Schema design
  – Introduction
  – Notation
• Functional dependencies
• Armstrong’s axioms
• Minimal cover
• Keys
• Questions
Let U be a schema and let F be a non empty set of non trivial FDs over U.

Assume that the FDs in F are of the form $X \rightarrow A$ where A is a single attribute.

True/False

- For every non-trivial FD $X \rightarrow A$ in F there exists a key K such that $A \in K$

- No!

- $U = \{A, B\}$

- $F = \{A \rightarrow B\}$

- The only key is A and B does not belong to the set \{A\}
• For every key K there exists a non trivial FD $X \rightarrow A$ in F such that $A \notin K$?

– This statement is true.
  • K does not equal U
  • Therefore, there exists $A \in U$ such that $A \notin K$
  • Since K is a key, it holds that $A \in K^+_F$
  • That is, there exists an attribute $A \in K^+_F$ such that
    – $A \notin K$
    – $X \rightarrow A \in F$
  • Therefore there exists $X \rightarrow A$ in F where $A \notin K$
The number of keys is at most the number of attributes in \( U \)?

- False.
- Let \( U=\{A,B,C,D\} \) and let
- \( F=\{AB \rightarrow CD, AC \rightarrow BD, AD \rightarrow BC, BC \rightarrow AD, BD \rightarrow AC, CD \rightarrow AB\} \)
- Then \( AB, AC, AD, BC, BD, CD \) are keys
- But there are only 4 attributes
Let
- \( F C \) be a minimal cover of \( F \) and let
- \( L \) be the set of attributes that appear in the l.h.s of FD’s in \( F \),
- \( R \) be the set of attributes that appear in the r.h.s of FD’s in \( F \)

If \( R \cap L = \emptyset \) then there is a unique key?
- True.
- \( U \setminus R \) is a superkey \( X \)
- If \( X \) does not contain \( B \in U \setminus R \) then the closure does not contain \( B \)
- Thus, \( U \setminus R \) is a unique key