Design: Normalization

Textbook (new edition), Chapter 14 & 15

Functional Dependencies

- Going back to FDs….
  - Notion of minimal sets
  - Notion of closure

- Goal: Given a set of dependencies Y, identify a set of dependencies X such that X is as small as possible and every dependency in Y is implied by X.
  - The dependencies inferred by a set of dependencies are, as a unit, called the *closure* of the original set; often written as X^+ given an initial set of dependencies X

Inferring Functional Dependencies

- Given a set of dependencies X, there are mechanisms to infer dependencies Y which hold from X

- Already aware of *transitivity:*

$$\text{If } A \rightarrow B \text{ and } B \rightarrow C\text{, then } A \rightarrow C$$

Inferring Functional Dependencies

- Armstrong's Axioms:
  - Reflexivity: If B is a subset of A, then A → B
  - Augmentation: If A → B, then A,C → B,C
  - Transitivity: If A → B and B → C, then A → C

  (A, B, C are subsets of attributes)

These rules are sound and complete – they generate all *(complete)* and only *(sound)* the functional dependencies inferable from a given set of dependencies X

Inferring Functional Dependencies

Additional rules stemming from Armstrong:

- *Self-determination*: A → A

- *Decomposition*: If A → BC, then A → B and A → C
  - We have written these like this already (singleton RHS)
- *Union*: If A → B and A → C, then A→ BC
- *Composition*: If A → B and C → D, then AC → BD

- *General Unification Theorem:* If A → B and C → D, then A union (C-B)→ BD

Proof of GUT

General Unification Theorem:

If A → B and C → D, then A union (C-B)→ BD

1. A→B (given)
2. C → D (given)
3. A → B INTERSECT C (a subset of B) (reflexivity)
4. C-B → C-B (self-determination)
5. A UNION (C-B) → (B INTERSECT C) UNION (C-B) (composition of 3 and 4)
6. A UNION (C-B) → C (simplification of 5)
7. A UNION (C-B) → D (transitivity, 2 and 6)
8. A UNION (C-B) → BD (composition, 1 and 7)

Inferring Functional Dependencies

Given a relation R over attributes A-F, and the following dependencies:

$A \rightarrow BC$

$B \rightarrow E$

$CD \rightarrow EF$

show that $AD \rightarrow F$ also holds

Inferring Functional Dependencies

Given, $A \rightarrow BC$ show: $AD \rightarrow F$

$B \rightarrow E$

$CD \rightarrow EF$

Given $A \rightarrow BC$, this leads to $A \rightarrow C$ (decomposition) which leads to $AD \rightarrow CD$ (augmentation). Together with the given $CD \rightarrow EF$, this leads to $AD \rightarrow EF$ (transitivity), which can be decomposed into $AD \rightarrow F$.

Closure of Attributes

- Given a set of attributes A and a set of FD S, then A^+ is the set of attributes functionally dependent on A under S
  - Note the re-use of the + symbol to have a slightly different meaning here
- This notion is important again for candidate keys – the closure of the candidate key attributes should be the set of all attributes
- Also important: A query of whether FD $X \rightarrow Y$ holds can be answered by seeing if attributes Y are in X^+

Algorithm for Finding Closure of Attribute Sets

- Given set of attributes A, initialize closure to be A.
- Repeat:

  For each FD, add the RHS attributes to the closure set if the LHS attributes are all in the closure as computed so far.

  until no changes have been made.

  (Changes may propagate over iterations)

Attribute Closure Examples

- Here is a set of FDs for a relation R over attributes {A,B,C,D,E,F,G}
  - $A \rightarrow B$
  - $BC \rightarrow DE$
  - $AEF \rightarrow G$

- Compute $\{A,C\}^+$

Attribute Closure Examples

- Given R over {A,B,C,D,E,F,G} and these FDS:
  - A $\rightarrow$ B
  - BC $\rightarrow$ DE
  - AEF $\rightarrow$ G
- Compute $\{A,C\}^+$
  - Start with closure = {A,C}
  - Iteration 1:
    - A $\rightarrow$ B: A is fully contained in {A,C}, so add B to closure; now is {A,B,C}
    - BC$\rightarrow$DE: BC is fully contained in closure, so add DE to closure; now is {A,B,C,D,E}
    - AEF$\rightarrow$G: AEF is not fully contained in closure
  - Revisit each rule: No changes are made: {A,B,C,D,E}

Attribute Closure Examples

- Given R over {A,B,C,D,E,F,G} and these FDS:
  - $A \rightarrow B$
  - $BC \rightarrow DE$
  - $AEF \rightarrow G$

  is $ACF \rightarrow DG$ implied by this set, using attribute closure?

Attribute Closure Examples

- Given R over {A,B,C,D,E,F,G} and these FDS:
  - A → B
  - BC → DE
  - AEF → G

  is ACF → DG implied by this set, using attribute closure?
- Compute $\{A,C,F\}^+$
  - Start with closure = {A,C,F}
  - Iteration 1:
    - A → B: A is fully contained in {A,C,F}, so add B to closure; now is {A,B,C,F}
    - BC→DE: BC is fully contained in closure, so add DE to closure; now is {A,B,C,D,E,F}
    - AEF→G: AEF is fully contained in closure, so add G to closure; now is {A,B,C,D,E,F,G}
  - Revisit each rule: No changes are made
  - DG are in {A,B,C,D,E,F,G}, so yes, ACF→DG is implied

Attribute Closure Examples

- Given R over {A,B,C,D,E,F,G} and these FDS:
  - $A \rightarrow B$
  - $BC \rightarrow DE$
  - $AEF \rightarrow G$

  is $ACF \rightarrow DG$ implied by this set, using previous axioms and rules?

Attribute Closure Examples

- Given R over {A,B,C,D,E,F,G} and these FDS:
  - A → B
  - BC → DE
  - AEF → G

  is ACF → DG implied by this set, using previous axioms and rules?

A→ B, so ACF → BCF by augmentation

ACF → BC by decomposition

ACF → DE by transitivity

ACF → DEG by GUT of (ACF→DE, AEF→G)

ACF→DG by decomposition

Minimal Sets of FDs

- We would prefer to write small sets of FDs
  - A set of FDs Y is considered to be *covered* by a set of FDs X if every FD in Y is also in X^+ (can be inferred from X).
  - A set of FDs X is considered to be minimal if:
    - Every dependency in X has a single attribute on the RHS
    - We cannot replace any dependency $A \rightarrow B$ with dependency $C \rightarrow B$, where C is a subset of A, and still get a set of dependencies equivalent to X
    - We cannot remove any dependencies from X and still have a set of dependencies equivalent to X
- Caveat: The minimal cover of a set of FDs Y (the minimal set equivalent to Y) is not guaranteed to be unique

Minimal Sets of FDs

Here is a set of FDs for relation R{A,B,C,D,E,F}:

$AB \rightarrow C$

$C \rightarrow A$

$BC \rightarrow D$

$ACD \rightarrow B$

$BE \rightarrow C$

$CE \rightarrow FA$

$CF \rightarrow BD$

$D \rightarrow EF$

Find a minimal cover for this set of FDs.

Minimal Sets of FDs

Here is a set of FDs for relation R{A,B,C,D,E,F}:

$AB \rightarrow C$ $C \rightarrow A$ $BC \rightarrow D$

$ACD \rightarrow B$ $BE \rightarrow C$ $CE \rightarrow FA$

$CF \rightarrow BD$ $D \rightarrow EF$

First, break down any RHS into singletons:

$AB \rightarrow C$ $C \rightarrow A$ $BC \rightarrow D$

$ACD \rightarrow B$ $BE \rightarrow C$ $CE \rightarrow F$

$CE \rightarrow A$ $CF \rightarrow B$ $CF \rightarrow D$

$D \rightarrow E$ $D \rightarrow F$

Minimal Sets of FDs

$AB \rightarrow C$	$C \rightarrow A$	$BC \rightarrow D$
$ACD \rightarrow B$	$BE \rightarrow C$	$CE \rightarrow F$
$CE \rightarrow A$	$CF \rightarrow B$	$CF \rightarrow D$
$D \rightarrow E$	$D \rightarrow F$	

A principled approach: Reduce LHS if possible

For a given LHS with multiple attributes, remove an attribute and see attribute closure of modified LHS. If it contains removed attribute, reduce.

Minimal Sets of FDs

AB→C
ACD→B
CE → A
D→E
C→A
BE→C
CF→B
D→F
BC→D
CE→F
CF→D

Using attribute closures:

{AB} {A}+ = {A} {B}+ = {B} // no reducing

{BC} {B}+ = {B}, {C}+ = {CA} // no reducing

{ACD} {AC}+ = {AC}, {AD}+ = {ADEF}, {CD}+ = {ACDBEF} suggests A is not needed (leaving CD→B)

{BE} {B}+ = {B}, {E}+ = {E} // no reducing

{CE} {C}+ = {AC}, {E}+ = {E} // no reducing

{CF} {C}+ = {AC}, {F}+ = {F} // no reducing

Minimal Sets of FDs

$AB \rightarrow C$	$C \rightarrow A$	$BC \rightarrow D$
$CD \rightarrow B$	$BE \rightarrow C$	$CE \rightarrow F$
$CE \rightarrow A$	$CF \rightarrow B$	$CF \rightarrow D$
$D \rightarrow E$	$D \rightarrow F$	

A principled approach: Remove rules if possible

For each rule, test if RHS can be achieved from all other rules using closure

Functional Dependencies

AB→C C→A BC→D

CD→B BE→C CE→F

CE → A CF→B CF→D

D→E D→F

{AB} closure, without using AB→C

{AB}+ = {AB} → does not contain C, so can't be discarded

{C} closure, without using C→ A {C}+ = {C} → can't discard

{BC}+ = {BCA} → can't discard

{CD}+ = {CDAEFB} → DISCARD CD→B

{BE}+ = {BE} → can't discard

{CE}+ not using CE→F = {CEA} → can't discard

{CE}+ not using CE→A = {CEFBDA} → DISCARD CE→A

{CF}+ not using CF→B = {CFDEA} → can't discard

{CF}+ not using CF→D = {CFABDE} →DISCARD CF→D

{D}+ not using D→E = {DF} → can't discard

{D}+ not using D→F = {DE} → can't discard

Functional Dependencies

$AB \rightarrow C$ $C \rightarrow A$ $BC \rightarrow D$

$BE \rightarrow C$ $CE \rightarrow F$ $CF \rightarrow B$

$D \rightarrow E$ $D \rightarrow F$

$\{CD\}^+ = \{CDAEFB\}$ → DISCARD $CD \rightarrow B$

$\{CE\}^+$ not using $CE \rightarrow A$ = $\{CEFBDA\}$ → DISCARD $CE \rightarrow A$

$\{CF\}^+$ not using $CF \rightarrow D$ = $\{CFABDE\}$ → DISCARD $CF \rightarrow D$