Schema Refinement Tutorial (part I)

Announcement

- Course assessment on AEFIS (10 minutes)

https://assess.stevens.edu/

Final Exam Logistics

- The final exam will be held in class next Tuesday (75 minutes)
- The exams will be held online in Canvas.
- The students still need to come to the class to take exams.
  - The seating chart will be signed during the exam

Cheat Sheet

- A one-sided, A4-size cheat sheet is allowed.
- The cheat sheet can be either handwritten or typed.
- The cheat sheet is NOT required to be submitted after the exam.

Type of Questions in Final Exam

- Mainly cover Functional Dependencies (FD)
  - True/false statement (cover the concepts in FD chapter)
  - Validation of correctness of SQL queries
  - Keys and normal forms (use FDs to find candidate keys, and check if the scheme satisfies 3NF and BCNF)
  - Lossless decomposition
  - 3NF decomposition

Make-up and Accommodations

- Make-up exams
  - The students who cannot attend the exams should inform the instructor at least one day BEFORE the exam.
  - The make-up exam request will be approved upon the validity of reasons.
  - The request will NOT be approved if the instructor is informed on the day of or after the exam.
- Accommodations
  - Student with accommodations send an email to the instructor for the arrangement.

Functional Dependencies (FDs)

- A <u>functional dependency</u> $X \rightarrow Y$ means

  Given any two tuples in *r*, if the X values are the same, then the Y values must also be the same. (but not vice versa)

- Can read "$\rightarrow$" as "determines"

FD Inference Rules

- FD Inference: new FDs can be implied by old FDS
- Armstrong's Axioms (X, Y, Z are <u>sets</u> of attributes):
  - *<u>Reflexivity</u>*: If $X \supseteq Y$, then $X \rightarrow Y$
  - *<u>Augmentation</u>*: If $X \rightarrow Y$, then $XZ \rightarrow YZ$ for any Z
  - *<u>Transitivity</u>*: If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$
- Some additional rules (that follow from AA):
  - *Union*: If $X \rightarrow Y$ and $X \rightarrow Z$, then $X \rightarrow YZ$
  - *Decomposition*: If $X \rightarrow YZ$, then $X \rightarrow Y$ and $X \rightarrow Z$

FD Closure

- F^+ = *closure of F* is the set of all FDs that are implied by *F*.
- Computing the closure of a set of FDs can be expensive. (Size of closure is exponential in # attrs!)
- Instead, we compute *attribute closure*
  - given FD F: $X \rightarrow Y$, X^+ = Set of all attributes A such that $X \rightarrow A$ is in F^+

Attribute Closure

- Compute ***attribute closure*** of X (denoted X^+) wrt *F.*
  - $X^+ := X$
  - Repeat until no change: if there is an FD $U \rightarrow V$ in *F* such that U is in X^+, then add V to X^+
- Approach can also be used to find the keys of a relation.
  - If X^+ ={all attributes of the relation}, then X is a superkey for R.

Exercise 1

- Given FD={AB→CD, B→DE, C→F, E→G, A→B}.
- Prove that {A→FG} is logically implied by FD, by using
  - (1) Armstrong's axiom
  - (2) Attribute closure

Solution 1: Armstrong's axiom

1	$A \rightarrow B$	Given
2	$A \rightarrow A\,B$	1, Augmentation
3	$A\,B \rightarrow C\,D$	Given
4	$A \rightarrow C\,D$	2, 3, Transitivity
5	$B \rightarrow D\,E$	Given
6	$A \rightarrow D\,E$	1, 5, Transitivity
7	$A \rightarrow A\,C\,D$	4, Augmentation
8	$A\,C\,D \rightarrow C\,D\,E$	6, Augmentation twice
9	$A \rightarrow C\,D\,E$	7, 8, Transitivity
10	$A \rightarrow C\,E$	9, Trivial dependency
11	$C \rightarrow F$	Given
12	$C\,E \rightarrow F\,E$	11, Augmentation
13	$E \rightarrow G$	Given
14	$F\,E \rightarrow F\,G$	13, Augmentation
15	$C\,E \rightarrow F\,G$	12, 14, Transitivity
16	$A \rightarrow F\,G$	10, 15, Transitivity

Solution 2: Attribute Closure

- Given FD={AB→CD, B→DE, C→F, E→G, A→B}.
- Prove that {A→FG} is logically implied by FD (by using attribute closure)
- Way of thinking
  - Prove that A^+ contains FG!
    - Initialize $A^+ = A$
    - Keep adding attributes into A^+ by following the algorithm of computing attribute closure.
    - $A^+ = ABCDEFG$

Using FDs to Find Keys

- A superkey K of a schema R is a set of attributes in R s.t. K-> R holds
- A (candidate) key is a *minimal* superkey.

Exercise 2

- Consider the relation schema $R(A, B, C, D)$ with FDs: $A \rightarrow C$ and $B \rightarrow D$. Is $\{A,B\}$ a candidate key for R?

Solution

- Step 1: check whether {A,B} is a superkey:
  – Check whether $(AB)^+$ ={ABCD};
- Step 2: check whether {A, B} is minimal (i.e., Is any subset of {A, B} is a superkey?):
  – Check whether $(A)^+$ ={ABCD};
  – Check whether $(A)^+$ ={ABCD};

Lossless and Dependency-preserving Decomposition

- Given a table R and a set of FDs F of R, the decomposition of R into X and Y is lossless with respect to F *if and only if*

  $X \cap Y \rightarrow X$ **or** $X \cap Y \rightarrow Y$

  – i.e., the common attributes of X and Y is a superkey of either X or Y.
- Decomposition of R into X and Y is *dependency preserving* if $(F_X \cup F_Y)^+ = F^+$
  – i.e., consider only dependencies in the closure F^+ that can be checked in X without considering Y, and in Y without considering X, these imply all dependencies in F^+.

Exercise 3

Consider a relation R(A,B,C). It has FDs: A -> B, B->C, and C -> A. Assume R is decomposed into X={A,B} and Y={B,C}. Answer the following questions:

(1) Is this decomposition lossless?

(2) Is this decomposition dependency preserving?

Exercise 3 – Lossless Solution

- Step 1: find common attributes of X and Y.
  - X={A,B} and Y={B,C}.
  - Common attribute: B
- Step 2: Check if B is the superkey of X or Y
  - FDs hold on X: A->B
  - FDs hold on Y: B->C
  - For X: B+= AB: B is a superkey of X
  - For Y: B+= BC: B is a superkey of Y

  So the decomposition (X, Y) is lossless.

Exercise 3 – Dependency Preserving Solution

- F^+= {A->B, B->C, C->A, A->C, B->A, C->B}
- F_x={A->B, B->A}
- F_y={B->C, C->B}
- $(Fx \cup Fy)^+$={A->B, B→C, B->A, C->B, A->C, C->A}
- $(Fx \cup Fy)^+$= F^+, so it is dependency preserving.