COSC265 — Relational Database Systems

Neville Churcher

Department of Computer Science & Software Engineering
University of Canterbury

2021

Data Dependencies

- ☆ Real-world facts (semantics) represented as constraints on database
- ☆ Some (integrity) constraints depend on attribute *values*:

  ```
  Person.Height < 10 meters
  IF YearsOfService = 30 THEN Age >= 15
  ```

  Some support (e.g. SQL domain constraints) but of little help in overall database design.
- ☆ Other constraints are *value-independent*.
- ☆ Functional dependencies (FDs) are an example
- ☆ Such data dependencies represent (time-independent) assertions about the real world.
- ☆ Data dependencies cannot be 'proved' but may be *enforced* by the DBMS

Introducing Functional Dependencies

Definition (Function)

A function is a binary relation between two sets that associates each element of the first set with exactly one element of the second set.

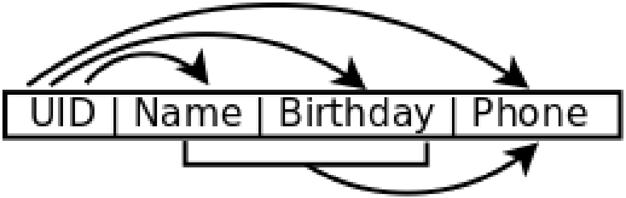
- ☆ FDs are *equality-generating* constraints between (sets of) attributes
- ☆ Whenever attribute X has the value x, attribute Y has the value y
- ☆ $X \rightarrow Y$: X *functionally determines* Y; Y is *functionally dependent on* X LHS, RHS in same relation scheme
- ☆ FDs represent facts about intension, valid extensions $r(R)$ must obey the corresponding constraints
- ☆ $StudentID \rightarrow Birthday$
- ☆ $Birthday \not\rightarrow StudentID$

Example

What can we say about these?

- ☆ $UID \to Name$?
- ☆ $Name \to UID$?
- ☆ $Name, Birthday \to Phone$?
- ☆ $UID \to Name, Birthday, Phone$?

UID	Name	Birthday	Phone
das21	Dave	12 Jan	325-3811
djp54	Dave	21 Feb	351-7683
dsw28	Debby	12 Jan	325-3811
ajd54	Andy	4 Aug	351-7683



Functional Dependencies

More formally . . .

Definition (FD)

Given $X \subseteq R, Y \subseteq R, r(R)$, if the FD $X \to Y$ holds then r cannot have two tuples that agree in their components for the attributes in X but disagree for one or more components of Y.

r satisfies $X \to Y$ if:

- ☆ $\forall x, \pi_Y(\sigma_{X=x}(r))$ has at most one tuple
- ☆ $\nexists u, v \in r$ where $u[X] = v[X]$ and $u[Y] \neq v[Y]$

Can pronounce '$\to$' as 'identifies' or 'functionally determines'

Keys & Superkeys

If $X \to R$ then X is a (super)key

Families, Closures & Inference ...

Given a set of dependencies, we can infer others

Definition ($\mathcal{F}$)

$\mathcal{F}$ is the family of FDs on R that all permissible states of $r(R)$ satisfy.

Definition (Logical Implication)

$\mathcal{F} \models X \to Y$ if every relation satisfying $\mathcal{F}$ also satisfies $X \to Y$

Definition (Closure)

$\mathcal{F}^+ = \{f | \mathcal{F} \models f\}$ is the set of all FDs logically implied by $\mathcal{F}$

- ☆ If $\mathcal{F} = \mathcal{F}^+$ then $\mathcal{F}$ is a *full* family of FDs
- ☆ $\mathcal{F}^+$ can be very large, and hard to calculate

Keys (again)

Definition (Keys)

Given $R = A_1 A_2 \ldots A_n, X \subseteq R$ then X is a key of R if:

1. $X \rightarrow A_1 A_2 \ldots A_n \in \mathcal{F}^+$
2. $\nexists Y \subseteq X, Y \rightarrow A_1 A_2 \ldots A_n \in \mathcal{F}^+$

Homework Exercise

- ☆ Last term you produced a number of conceptual data models using EER diagrams and other table/relation identification techniques.
- ☆ Review the entities (and corresponding relations) from your answers to some of the exercises in tutorials 1 & 2 — exercise 6 from tutorial 2 would be a good one to start with.
- ☆ Identify the FDs that hold between the attributes in the various schemes/tables
- ☆ What patterns do you notice?

Inference

Answering questions about our data model

- ☆ If some members of $\mathcal{F}^+$ are known, then others may be *inferred*
- ☆ *Inference rules* state that, if a relation satisfies certain FDs, then it must also satisfy certain others
- ☆ *Armstrong's axioms* are one such set. They are:

  Complete: Given $\mathcal{F}$, can deduce *all* $f_i \in \mathcal{F}^+$

  Sound: Given $\mathcal{F}$, can not deduce any $f_i \notin \mathcal{F}^+$
- ☆ Other inference rules may be derived from these

Reflexivity: $\forall Y, Y \subseteq X \subseteq \mathcal{U}, \mathcal{F} \models X \rightarrow Y$ (N.B. $\not\propto \mathcal{F}$)

Augmentation: $\forall Z \subseteq \mathcal{U}, X \rightarrow Y \models XZ \rightarrow YZ$

Transitivity: $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$

William Ward Armstrong

Derived Inference Rules

Union Rule: $\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$

Proof.

1. $X \rightarrow Y$ (given)
2. $X \rightarrow XY$ (augment 1 by X)
3. $X \rightarrow Z$ (given)
4. $XY \rightarrow YZ$ (augment 3 by Y)
5. $X \rightarrow YZ$ (transitivity on 2, 4)

$\square$

... continued

Pseudotransitivity Rule: $\{X \rightarrow Y, WY \rightarrow Z\} \models XW \rightarrow Z$

Proof.

1. $X \rightarrow Y$ (given)
2. $WX \rightarrow WY$ (augment 1 by W)
3. $WY \rightarrow Z$ (given)
4. $XW \rightarrow Z$ (transitivity on 2, 3)

□

Decomposition Rule: $\forall Z \subseteq Y, X \rightarrow Y \models X \rightarrow Z$

Equivalently, $X \rightarrow VW \models X \rightarrow V, X \rightarrow W$

Example

Past exam question — just sayin' ...

Example (Using inference rules)

Given $R = ABCD$, and $\mathcal{F} = \{A \to B, AC \to D, BC \to A, BC \to D, CD \to A\}$, show that $AC \to D$ is redundant.

Solution 1: Pseudotransitivity rule

- $A \to B \in \mathcal{F}$
- $CB \to D \in \mathcal{F}$
- $A \to B, CB \to D \models AC \to D$
- We can infer $AC \to D$ from other FDs
- Thus $AC \to D$ is redundant and can be removed from $\mathcal{F}$

$\square$

Example

Past exam question again — solved differently this time

Example (Using inference rules)

Given $R = ABCD$, and $\mathcal{F} = \{A \to B, AC \to D, BC \to A, BC \to D, CD \to A\}$, show that $AC \to D$ is redundant.

Solution 2: Armstrong's axioms

- ☆ $A \to B \in \mathcal{F}$
- ☆ $BC \to D \in \mathcal{F}$
- ☆ Augmentation: $A \to B \models AC \to BC$
- ☆ Transitivity: $AC \to BC, BC \to D \models AC \to D$
- ☆ We can infer $AC \to D$ from other FDs
- ☆ Thus $AC \to D$ is redundant and can be removed from $\mathcal{F}$

□

Computing Closures

- ☆ In general, to tell if $\mathcal{F} \models f$ we must compute $\mathcal{F}^+$ and see if $f \in \mathcal{F}^+$
- ☆ $\mathcal{F}^+$ can be very large — even for "small" $\mathcal{F}$
- ☆ This is bad news — we'll avoid having to compute $\mathcal{F}^+$ whenever we can
- ☆ Consider $\mathcal{F} = \{A \rightarrow B_1, A \rightarrow B_2, \ldots A \rightarrow B_n\}$
- ☆ Additivity (Union Rule) tells us that $\mathcal{F}^+$ includes all $A \rightarrow Y$ where $Y \subseteq B_1 \ldots B_n$
- ☆ There are $\sum_{i=0}^{n} {}^nC_i = 2^n$ of these!
- ☆ Plus all the others ...
- ☆ Is there a simpler way?

Yes — Attribute Closures

Much better news...

Definition (Attribute Closures)

$X^+ = \{A_i | \mathcal{F} \models X \to A_i\}$

Lemma

$X \to Y$ follows from Armstrong's axioms iff $Y \subseteq X^+$

Proof.

Assume $Y \subseteq X^+, Y = A_1 A_2 \ldots A_n$
Then, from definition of $X^+, X \to A_i$ and thus $X \to Y$ by the union rule
Or,
Assume $X \to Y$ follows from the axioms. Then, by the decomposition rule,
$\forall i, i \in \{1 \ldots n\}, X \to A_i$ so $Y \subseteq X^+$ □

Computing X^+

I'll need this for the exam

- ☆ Lemma tells us that testing $X \to Y \in \mathcal{F}^+$ is no harder than computing X^+
- ☆ This is excellent news!
- ☆ Algorithm for computing X^+ is O(N) where N is the number of FDs in $\mathcal{F}$
- ☆ This is also excellent news!
- ☆ Compute the sequence $X^{(0)}, X^{(1)}, \ldots X^{(i)}$ until $X^{(i+1)} = X^{(i)} \equiv X^+$
- ☆ $X^{(0)} = X$
- ☆ $X^{(i+1)} = X^{(i)} \cup A$ where $Y \subseteq X^{(i)}$ and $Y \to Z \in \mathcal{F}$ and $A \subseteq Z$
- ☆ $X^{(j)} \subseteq X^{(j+1)}$

Computing X^+

In plainer language ...

Computing X^+

1. Start with X itself as the initial attribute set $X^{(0)}$
2. Look for an FD in $\mathcal{F}$ whose LHS is X or a subset of X
3. If you find one, then add its RHS to the working set
4. Repeat steps 2 and 3 until no more changes can be made
5. Now you're done!

Previous Example

Past exam question — yet again

Example (Using inference rules)

Given $R = ABCD$, and $\mathcal{F} = \{A \to B, AC \to D, BC \to A, BC \to D, CD \to A\}$, show that $AC \to D$ is redundant.

Solution 3: Attribute Closure

Show $\mathcal{F}' = \mathcal{F} \setminus AC \to D \models AC \to D$

Consider $X = AC$

$X^{(0)} = AC$	look for dependencies in $\mathcal{F}'$ with LHS (Y) of $A, C, or\ AC$ and find $A \to B$
$X^{(1)} = ABC$	find $BC \to D$
$X^{(2)} = X^{(+)} = ABCD$	All attributes, AC is candidate key

☞ Have shown $AC \to D$ since $AC \subseteq X^{+}$ so it is redundant and can be removed from $\mathcal{F}$

□

Bigger Example (Compute BD^+)

$R = ABCDEG$

$$\mathcal{F} = \left\{ \begin{array}{ll} AB \to C & D \to EG \\ C \to A & BE \to C \\ BC \to D & CG \to BD \\ ACD \to B & CE \to AG \end{array} \right\}$$

Consider $X = BD$

$X^{(0)} = BD$	look for dependencies in $\mathcal{F}$ with LHS (Y) of B, D, BD and find $D \to EG$
$X^{(1)} = BDEG$	find $D \to EG$ and $BE \to C$
$X^{(2)} = BCDEG$	add $C \to A$, $BC \to D$, $CG \to BD$, $CE \to AG$
$X^{(3)} = X^+ = ABCDEG$	All attributes

Continued (B^+, D^+)

$$R = ABCDEG \qquad \mathcal{F} = \left\{ \begin{array}{ll} AB \to C & D \to EG \\ C \to A & BE \to C \\ BC \to D & CG \to BD \\ ACD \to B & CE \to AG \end{array} \right\}$$

$B^{(0)} = B^+$	No more attributes can be added
$D^{(0)} = D$	find $D \to EG$ only
$D^{(1)} = D^+ = DEG$	Nothing more can be added

☞ $D^+ = DEG$ tells us that $D \to G \in \mathcal{F}^+, D \to DE \in \mathcal{F}^+$ etc.

Conclusion

$BD^+ = R, B^+ \neq R, D^+ \neq R$ means that:

- ☆ BD is a candidate key for R
- ☆ Neither B nor D is a candidate key

A Big Issue

If $\mathcal{F}$ is so important, how do we know we've got it right?

Real-world example

- ☆ Ron and Hermione each model the same domain
- ☆ They each deliver a set of FDs – $\mathcal{F}_{Ron}$ and $\mathcal{F}_{Hermione}$
- ☆ $\mathcal{F}_{Ron}$ contains 23 FDs; $\mathcal{F}_{Hermione}$ contains 19
- ☆ Some FDs are in both $\mathcal{F}_{Ron}$ and $\mathcal{F}_{Hermione}$; others are in only one

What to do?

Ron: I modelled the domain accurately — you must have done something wrong

Hermione: No, *I* modelled it accurately too — it's *you* that must be wrong

Ron: Maybe we're *both* right?

Hermione: No way! Even if we were, my $\mathcal{F}$ has fewer FDs in it than yours so it's obviously better

Ron: How are we going to settle this — there must be a way

Hermione: I know, how about this? I'll take your $\mathcal{F}$, and if I can use it to infer every FD in mine then I'll believe yours is as good

Ron: OK, I'll do the same.

Hermione: And if you can use my $\mathcal{F}$ to infer every FD in yours then we might as well use mine because it has fewer FDs in it

Ron: OK, let's do that . . .

Covers of Dependency Sets

More formally ...

Let $\mathcal{F}, \mathcal{G}$ be sets of dependencies

- ☆ $\mathcal{F}$ *covers* $\mathcal{G}$ if if every FD in $\mathcal{G}$ can be inferred from $\mathcal{F}$ (i.e., if $\mathcal{G}^+ \subseteq \mathcal{F}^+$)
- ☆ $\mathcal{F}$ and $\mathcal{G}$ are *equivalent* if $\mathcal{F}^+ = \mathcal{G}^+$
- ☆ If $\mathcal{F}^+ = \mathcal{G}^+$ then $\mathcal{F}$ *covers* $\mathcal{G}$ and $\mathcal{G}$ covers $\mathcal{F}$
- ☆ Test for equivalence:

for each FD $X \rightarrow Y$ *in* $\mathcal{F}$
    compute X^+ *using* $\mathcal{G}$
    test that $Y \subseteq X^+$ *(i.e.* $X \rightarrow Y \in \mathcal{G}^+$*)*
Then repeat for each FD in $\mathcal{G}$

Example

Example (Are $\mathcal{F}_R$ and $\mathcal{F}_H$ equivalent?)

$R = ABCDEG$; $\mathcal{F}_H = \{A \to CD, E \to AG\}$ $\mathcal{F}_R = \{A \to C, AC \to D, E \to AD, E \to G\}$

Does $\mathcal{F}_R$ cover $\mathcal{F}_H$?

☞ $\forall f_i \in \mathcal{F}_H$ does $\mathcal{F}_R \models f_i$?

	$A^{(0)} = A$	look for FDs in $\mathcal{F}_R$ with LHS A and find $A \to C$
A^+	$A^{(1)} = AC$	find $AC \to D$
	$A^{(2)} = X^{(+)} = ACD$	$\therefore A \to CD$ ✔

	$E^{(0)} = E$	look for FDs in $\mathcal{F}_R$ with LHS E and find $E \to G$,
E^+		$E \to AD$
	$E^{(1)} = E^{(+)} = ADEG$	$\therefore E \to AG$ ✔

$\therefore \mathcal{F}_R$ covers $\mathcal{F}_H$ □

Example (Are $\mathcal{F}_R$ and $\mathcal{F}_H$ equivalent?)

$R = ABCDEG$; $\mathcal{F}_H = \{A \to CD, E \to AG\}$ $\mathcal{F}_R = \{A \to C, AC \to D, E \to AD, E \to G\}$

Does $\mathcal{F}_H$ cover $\mathcal{F}_R$?

☞ $\forall f_i \in \mathcal{F}_R$ does $\mathcal{F}_H \models f_i$?

A^+	$A^{(0)} = A$	look for FDs in $\mathcal{F}_H$ with LHS A and find $A \to CD$
	$A^{(+)} = ACD$	$\therefore A \to C$ ✔
AC^+	$AC^{(0)} = AC$	FDs in $\mathcal{F}_H$ with LHS A, C, AC — find $A \to C$
	$AC^{(+)} = ACD$	$\therefore AC \to D$ ✔
E^+	$E^{(0)} = E$	look for FDs in $\mathcal{F}_H$ with LHS E and find $E \to AG$
	$E^{(1)} = AEG$	find $A \to CD$
	$E^{(+)} = ACDEG$	$\therefore E \to G$, $E \to AD$ ✔

$\therefore \mathcal{F}_H$ covers $\mathcal{F}_R$. Have already shown $\mathcal{F}_R$ covers $\mathcal{F}_H$ so $\mathcal{F}_R, \mathcal{F}_H$ are equivalent □

More on Covers

Lemma

Every $\mathcal{F}$ is covered by some $\mathcal{G}$ in which no RHS has more than one attribute.

Proof.

Let $\mathcal{G} = \{X \rightarrow A_i | A_i \subset Y \wedge X \rightarrow Y \in \mathcal{F}\}$

$X \rightarrow Y \models X \rightarrow A_i$ (decomposition rule)

$\therefore \mathcal{G} \subseteq \mathcal{F}^+$

$\{X \rightarrow A_1, X \rightarrow A_2, \ldots\} \models X \rightarrow Y$ (union rule)

$\therefore \mathcal{F} \subseteq \mathcal{G}^+$

$\mathcal{G} \subseteq \mathcal{F}^+, \mathcal{F} \subseteq \mathcal{G}^+ \Rightarrow \mathcal{F}^+ = \mathcal{G}^+$

$\therefore \mathcal{G}$ covers $\mathcal{F}$ $\square$

Minimal Covers

- Can also prove that every $\mathcal{F}$ is equivalent to an $\mathcal{F}'$ which is *minimal*
- $\mathcal{F}$ is minimal if:
  - RHS of every $f \in \mathcal{F}$ is a single attribute (*no redundant RHS attributes*)
  - $\nexists X \to A$ where $X \to A \in \mathcal{F}$ and $\{\mathcal{F} - X \to A\}^+ = \mathcal{F}^+$ (*no redundant dependencies*)
  - $\nexists X \to A, Z \subset X$ where $(\{\mathcal{F} - X \to A\} \cup \{Z \to A\})^+ = \mathcal{F}^+$ (*no redundant LHS attributes*)
- Minimal cover is *not unique*
  - may have different FDs
  - may nave different numbers of FDs
- Minimal covers are a useful starting point for database design

Minimal Cover Example

Revisit previous example

$$R = ABCDEG \qquad \mathcal{F} = \left\{ \begin{array}{ll} AB \to C & D \to EG \\ C \to A & BE \to C \\ BC \to D & CG \to BD \\ ACD \to B & CE \to AG \end{array} \right\}$$

☞ Split RHS so every FD has single-attribute RHS

$$\mathcal{F} = \left\{ \begin{array}{llll} AB \to C & BE \to C & C \to A & CG \to B \\ BC \to D & CG \to D & ACD \to B & CE \to A \\ D \to E & CE \to G & D \to G & \end{array} \right\}$$

Minimal Cover Example

remove redundant dependencies

$$\mathcal{F} = \left\{ \begin{array}{llll} AB \to C & BE \to C & C \to A & CG \to B \\ BC \to D & CG \to D & ACD \to B & CE \to A \\ D \to E & CE \to G & D \to G & \end{array} \right\}$$

☞ $CE \to A$ is redundant — remove

1. $C \to A$ (given)
2. $CE \to AE$ (augmentation)
3. $CE \to A$ (decomposition)

Minimal Cover Example

remove redundant dependencies

$$\mathcal{F} = \left\{ \begin{array}{llll} AB \to C & BE \to C & C \to A & CG \to B \\ BC \to D & CG \to D & ACD \to B & D \to E \\ CE \to G & D \to G & & \end{array} \right\}$$

☞ $CG \to B$ is redundant — remove

Compute CG^+

$CG^{(0)} = CG$ add $CG \to D, C \to A$

$CG^{(1)} = ACDG$ add $ACD \to B$

$\therefore$ can get $CG \to B$ indirectly

Minimal Cover Example

Remove redundant LHS attributes

$$\mathcal{F} = \left\{ \begin{array}{lllll} AB \rightarrow C & BE \rightarrow C & C \rightarrow A & BC \rightarrow D & CG \rightarrow D \\ ACD \rightarrow B & D \rightarrow E & CE \rightarrow G & D \rightarrow G & \end{array} \right\}$$

☞ A is redundant in $ACD \rightarrow B$ so replace by $CD \rightarrow B$

1. If we have $CD \rightarrow B$, can we show $ACD \rightarrow B$?
2. $ACD \rightarrow AB$ (augmentation)
3. $ACD \rightarrow B$ (decomposition)

Finally...

☞ We started with (after splitting RHS):

$$\mathcal{F} = \left\{ \begin{array}{llll} AB \to C & BE \to C & C \to A & CG \to B \\ BC \to D & CG \to D & ACD \to B & CE \to A \\ D \to E & CE \to G & D \to G & \end{array} \right\}$$

☞ Removed 2 redundant FDs and 1 LHS attribute to get one minimal cover of $\mathcal{F}$

$$\mathcal{F}' = \left\{ \begin{array}{lll} AB \to C & BE \to C & C \to A \\ BC \to D & CG \to D & CD \to B \\ D \to E & CE \to G & D \to G \end{array} \right\}$$

☞ Another minimal cover of $\mathcal{F}$ is (DIY)

$$\mathcal{F}'' = \left\{ \begin{array}{llll} AB \to C & BE \to C & C \to A & CG \to B \\ BC \to D & D \to E & D \to G & CE \to G \end{array} \right\}$$

☞ Note that $\mathcal{F}'$ and $\mathcal{F}''$ contain different numbers of dependencies and have arisen from the elimination of dependencies in different orders.