

Background for Database Normalization

Tom Kelliher, CS 318

Mar. 6, 2002

1 Administrivia

Announcements

Assignment

Read 8.1–4.

From Last Time

Relationship between SQL, relational algebra, and calculi.

Outline

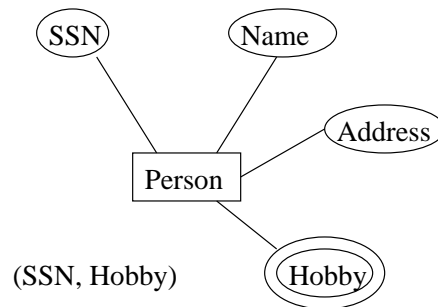
1. The problem.
2. Decomposition.
3. Functional dependencies.
4. Properties of functional dependencies.

Coming Up

Further properties of functional dependencies.

2 The Problem

Consider the following E/R diagram, schema, and relation instance:



```
CREATE TABLE Person (  
  SSN          INTEGER,  
  Name         CHAR(20),  
  Address      CHAR(50),  
  Hobby        CHAR(20),  
  PRIMARY KEY (SSN, Hobby));
```

Note that neither SSN nor Hobby may be NULL, since they're part of a primary key.

SSN	Name	Address	Hobby
111111111	John Doe	123 Main St.	Stamps
111111111	John Doe	123 Main St.	Hiking
111111111	John Doe	123 Main St.	Coins
222222222	Jane Smith	321 Elm St.	Gardening
666666666	Evil Tom	666 Turnmill Ct.	Ice Cream
666666666	Evil Tom	666 Turnmill Ct.	Perpetrating Evil
666666666	Evil Tom	666 Turnmill Ct.	Hiking Devils Postpile

Questions:

1. Suppose we want to insert John with no hobbies?

We have to come up with an effective “No Hobby” hobby.

2. Suppose we later want to add a hobby for John. Do we update the original tuple or add a new tuple?

Special case!

3. Consider how much work must be done if Evil Tom becomes Good Tom.

This is counter-intuitive, making three DB changes to reflect one real change.

4. Suppose Jane no longer gardens.

Another special case!

These *update anomalies* are the result of undesirable redundancy in the DB.

2.1 Decomposition

1. Consider the decomposition of Person into:

Person(SSN, Name, Address)

Hobby(SSN, Hobby)

- (a) How does this solve our problems?
- (b) What redundancy remains? Does it matter?

2. Consider this decomposition of Person:

SSN(SSN)

Name(Name)

Address(Address)

Hobby(Hobby)

- (a) No redundancy, right?

- (b) What's wrong with this decomposition?
- 3. Key idea — E/R design may result in redundancy. We need a set of tools to find and eliminate redundancy.

3 Functional Dependencies

- 1. Functional dependency definition:

A constraint over a relation R of the form $\overline{X} \rightarrow \overline{Y}$ where \overline{X} and \overline{Y} are sets of attributes of \mathbf{R} . If \mathbf{r} is an instance of \mathbf{R} , it is said to satisfy this functional dependency if:

For every pair of tuples s and t in \mathbf{r} , if s and t agree on \overline{X} then s and t agree on \overline{Y} .

Informally, we would say \overline{X} determines \overline{Y} .

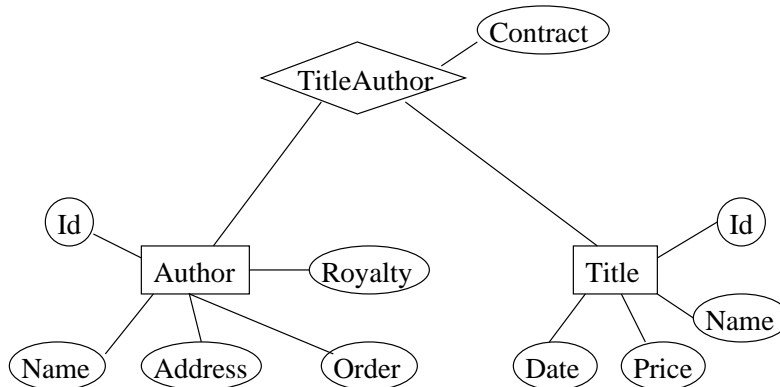
- 2. What type of constraint have we already see which is a functional dependency?
- 3. Example:

$$\begin{aligned} \text{SSN, Hobby} &\rightarrow \text{SSN, Name, Address, Hobby} \\ \text{SSN} &\rightarrow \text{Name, Address} \end{aligned}$$

How are our update anomalies manifest here?

Any other dependencies in the first Person relation?

- 4. What are the functional dependencies in the Person/Hobby relations?
- 5. What are the functional dependencies inherent in this *flawed* E/R diagram?



6. Nomenclature:

(a) Individual attributes: A, B, C .

(b) Sets of attributes: $ABC, \overline{P}, \overline{X}, \overline{XY}$.

4 Properties of Functional Dependencies

Entailment: Consider all relation instances \mathbf{r} over attributes \overline{R} . We say the set of FDs \mathcal{F} entails the FD f if the following property is true:

If \mathbf{r} satisfies every FD in \mathcal{F} , then \mathbf{r} satisfies FD f .

The closure of \mathcal{F} (written \mathcal{F}^+) is the set of all FDs entailed by \mathcal{F} . Note $\mathcal{F} \subseteq \mathcal{F}^+$.

Armstrong's Axioms:

1. Reflexivity of FDs. If $\overline{Y} \subseteq \overline{X}$ then $\overline{X} \rightarrow \overline{Y}$.

Proof. Consider a relation instance \mathbf{r} over attributes \overline{XY} . If $s, t \in \mathbf{r}$ agree on \overline{X} , then they agree on \overline{Y} . Hence, $\overline{X} \rightarrow \overline{Y}$.

Such trivial FDs are entailed by any set of FDs.

2. Augmentation of FDs. If $\overline{X} \rightarrow \overline{Y}$ then $\overline{XZ} \rightarrow \overline{YZ}$ for a relation instance \mathbf{r} containing \overline{XYZ} .

Proof?

3. Transitivity of FDs. If $\overline{X} \rightarrow \overline{Y}$ and $\overline{Y} \rightarrow \overline{Z}$ then $\overline{X} \rightarrow \overline{Z}$.

Proof is a homework problem.