Theory I: Database Foundations

Jan-Georg Smaus (Georg Lausen)

2.8.2011

1. Formal Design
   - Motivation
   - Functional Dependencies
   - Decomposition

Formal Design

- We want to distinguish good from bad database design.
- What kind of additional information do we need?
- Can we transform a bad into a good design?
- (At which cost?)

20.1 Motivation

Relations and anomalies

City

CityNo	CityName	CCode	CSurface
7	Freiburg	D	357
9	Berlin	D	357
40	Moscow	RU	17075
43	St.Petersburg	RU	17075

Continent

ConName	CCode	ConSurface	Percent
Europe	D	3234	100
Europe	RU	3234	20
Asia	RU	44400	80

Having removed anomalies

City'

CityNo	CityName	CCode
7	Freiburg	D
9	Berlin	D
40	Moscow	RU
43	St.Petersburg	RU

Country'

CCode	CSurface
D	357
RU	17075

Location'

CCode	ConName	Percent
D	Europe	100
RU	Europe	20
RU	Asia	80

Continent'

ConName	ConSurface
Europe	3234
Asia	44400

20.2 Functional Dependencies

Definition

- Let a relation schema be given by its format X and let $Y, Z \subseteq X$.
- Let $r \in \text{Rel}(X)$. r fulfills a functional dependency $Y \to Z$, if for all $\mu, \nu \in r$:

$$\mu[Y] = \nu[Y] \Rightarrow \mu[Z] = \nu[Z].$$

- Let $\mathcal{F}$ be a set of functional dependencies over X. The set of all relations $r \in \text{Rel}(X)$, which fulfill all functional dependencies in $\mathcal{F}$, is called $\text{Sat}(X, \mathcal{F})$.

20.2.2 Membership Test

- The functional dependency $Y \to Z$, is implied by $\mathcal{F}$, written $\mathcal{F} \models Y \to Z$, if for each relation r, whenever $r \in \text{Sat}(X, \mathcal{F})$ then r fulfills $Y \to Z$.
- The set $\mathcal{F}^+ = \{Y \to Z \mid \mathcal{F} \models Y \to Z\}$ is called closure of $\mathcal{F}$.
- The question "$Y \to Z \in \mathcal{F}^+$?" is called membership test.

Key

Let $X = \{A_1, \ldots, A_n\}$. $Y \subseteq X$ is called key of X (wrt. $\mathcal{F}$), if

- $Y \rightarrow A_1 \ldots A_n \in \mathcal{F}^+$,
- $Z \subset Y \Rightarrow Z \rightarrow A_1 \ldots A_n \notin \mathcal{F}^+$.

Armstrong axioms

Let $r \in \text{Sat}(X, \mathcal{F})$.

(A1) Reflexivity: If $Z \subseteq Y \subseteq X$, then r fulfills functional dependency $Y \to Z$.

(A2) Augmentation: If $Y \to Z \in \mathcal{F}, V \subseteq X$, then r fulfills functional dependency $YV \to ZV$.

(A3) Transitivity: If $Y \to Z, Z \to V \in \mathcal{F}$, then r fulfills functional dependency $Y \to V$.

(A1): trivial functional dependencies.

Correctness and Completeness

- Every functional dependency derivable by the Armstrong axioms is an element of the closure (correctness).
- Every functional dependency in $\mathcal{F}^+$ is derivable by the Armstrong axioms (completeness)

Membership test

Starting from $\mathcal{F}$ apply (A1)–(A3) until $Y \to Z$ is derived, or $\mathcal{F}^+$ is derived and $Y \to Z \notin \mathcal{F}^+$.

In practice this test is too complex and therefore other tests have been developed.

20.3 Decomposition

Let $\rho = \{Y_1, \dots, Y_k\}$ a decomposition of X, i.e., $Y_1 \cup \dots \cup Y_k = X$. Let $\mathcal{F}$ be a set of functional dependencies.

- Let $r \in \mathrm{Sat}(X, \mathcal{F})$ and let $r_i = \pi[Y_i]r$, $1 \leq i \leq k$.

  ρ is called lossless, if for any $r \in \mathrm{Sat}(X, \mathcal{F})$ it holds that:

$$r = \pi[Y_1]r \bowtie \dots \bowtie \pi[Y_k]r.$$

Example

- $X = \{A, B, C\}$ and $\mathcal{F} = \{A \to B, A \to C\}$.
- $r \in Sat(X, \mathcal{F})$:

$r =$

A	B	C
a_1	b_1	c_1
a_2	b_1	c_2

- $\rho_1 = \{AB, BC\}$ and $\rho_2 = \{AB, AC\}$.
- $r \qquad \pi[AB]r \bowtie \pi[BC]r$,
- $r \qquad \pi[AB]r \bowtie \pi[AC]r$.

Theorem

Let a format X and set $\mathcal{F}$ of functional dependencies. Let $\rho = (Y_1, Y_2)$ be a decomposition of X.

ρ is lossless, iff

$$(Y_1 \cap Y_2) \rightarrow (Y_1 \setminus Y_2) \in \mathcal{F}^+, \text{or } (Y_1 \cap Y_2) \rightarrow (Y_2 \setminus Y_1) \in \mathcal{F}^+.$$

- There is a similar notion called dependency-preserving.
- The aim of good database design is to decompose relations to remove redundancies.