

Databaser 2: Avhengigheter og case

20.1 Avhengigheter (dependencies)

Oppsummering av dagens forelesning:

- Avhengigheter er formler i FOL som uttrykker semantiske føringer på databaseskjemaet
- Vi skal idag se på to vanlige avhengigheter som oppfører seg spesielt pent: funksjonelle avhengigheter og join-avhengigheter
- Chase er en sentral teknikk for å vise implikasjon mellom spørringer mht. avhengigheter
- Chase illustrerer en viktig teknikk i teoretisk databehandling

Definisjon 20.1

La A_1, \dots, A_n være ikke-repeterende liste attributter i U , og I en mengde tupler over U . Projeksjonen π_{A_1, \dots, A_n} er gitt ved mengden av tupler s over A_1, \dots, A_n slik at $\exists t \in I$ der $s(A_i) = t(A_i)$ for hver $i \leq n$.

Definisjon 20.2

La U være en mengde attributter. En *funksjonell avhengighet* (*fd*) over U er et uttrykk på formen $X \rightarrow Y$, der $X, Y \subseteq U$. En instans I over en relasjon $R[U]$ oppfyller $X \rightarrow Y$, skrevet $I \models X \rightarrow Y$, dersom

$$\pi_X(s) = \pi_X(t) \text{ impliserer } \pi_Y(s) = \pi_Y(t)$$

for alle tupler $s, t \in I$.

Definisjon 20.3

La $X_1, \dots, X_n \subseteq U$ og I_j være instans over X_j , for hver $j \leq n$. Den *naturlige join* av hver I_j er:

$$\bowtie_{j=1}^n \{I_j\} = \{s \text{ over } \cup X_j \mid \pi_{X_j}(s) \in I_j \text{ for hver } j \leq n\}.$$

Definisjon 20.4

En *join-avhengighet* (*jd*) over en mengde attributter U er et uttrykk på formen $\bowtie [X_1, \dots, X_n]$, der

- $X_1, \dots, X_n \subseteq U$ og
- $\cup_{j=1}^n X_j = U$.

En relasjon I over U *tilfredsstiller* $\bowtie [X_1, \dots, X_n]$ hvis $I = \bowtie_{j=1}^n \{\pi_{X_j}(I)\}$.

Definisjon 20.5

La Σ være en mengde avhengigheter. Vi sier at q_1 er inneholdt i q_2 relativt til Σ , $q_1 \subseteq_{\Sigma} q_2$, dersom $q_1(\mathbf{I}) \subseteq q_2(\mathbf{I})$ for hver instans \mathbf{I} som oppfyller hver avhengighet i Σ .

Definisjon 20.6

$q_1 \equiv_{\Sigma} q_2$ hviss $q_1 \subseteq_{\Sigma} q_2$ og $q_2 \subseteq_{\Sigma} q_1$.

Merk: Når $q_1 \equiv_{\Sigma} q_2$, vil q_1 og q_2 returnere samme svar fra en database som oppfyller Σ !

20.2 Chase

Gitt en mengde Σ av avhengigheter på R og en ordning \leq på variable. Chase defineres på en tablåpørring (\mathbf{T}, t) over R mhp. Σ .

Definisjon 20.7

fd-regelen: La

- $\sigma \in \Sigma$ være $X \rightarrow A$
- $u, v \in \mathbf{T}$ slik at $\pi_X(u) = \pi_X(v)$ og $u(A) \neq v(A)$
- x være den minste variabelen mht \leq i $\{u(A), v(A)\}$.

σ anvendt på u, v i (\mathbf{T}, t) gir tablået $(\theta(\mathbf{T}), \theta(t))$, der $\theta(y) = x$ og θ ellers er identitetsfunksjonen.

Definisjon 20.8

jd-regelen: La

- $\sigma \in \Sigma$ være $\bowtie [X_1, \dots, X_n]$
- u være fritt tuppel over R som ikke er i \mathbf{T}
- $u_1, \dots, u_n \in \mathbf{T}$ oppfylle $\pi_{X_i}(u_i) = \pi_{X_i}(u)$ for hver $i \leq n$.

σ anvendt på (u_1, \dots, u_n) i (\mathbf{T}, t) gir tablået $(\mathbf{T} \cup \{u\}, t)$.

Teorem 20.9

La Σ være mengde fd'er og jd'er over R , $\sigma \in \Sigma$, og q en tablåpørring over R . Hvis q' er resultatet av å anvende σ på noen tupler i q , er $q \equiv_{\Sigma} q'$. ⊣

Definisjon 20.10

En *chasing-sekvens* er en sekvens av tablå-spøringer, der hvert tablå i sekvensen fremkommer fra det forrige gjennom en anvendelse av en chase-regel. Sekvensen er *terminerende* hvis den er endelig og ingen regel kan anvendes på det siste tablået.

Teorem 20.11

La (\mathbf{T}', t') være resultatet av en terminerende chase-sekvens fra (\mathbf{T}, t) mhp. Σ . Da vil (\mathbf{T}', t') , når det betraktes som en instans, oppfylle Σ . \dashv

Lemma 20.12

La Σ inneholde fd-er og jd-er.

- Enhver chasing-sekvens er endelig og kan utvides til en terminerende sekvens.
- To terminerende chasing-sekvenser mht Σ fra samme tablå-spørring (\mathbf{T}, t) resulterer i identisk tablå. Dette betegnes $\text{chase}(\mathbf{T}, t, \Sigma)$. \dashv

Teorem 20.13

1. $(\mathbf{T}_1, t_1) \subseteq_{\Sigma} (\mathbf{T}_2, t_2)$ hviss $\text{chase}(\mathbf{T}_1, t_1, \Sigma) \subseteq \text{chase}(\mathbf{T}_2, t_2, \Sigma)$.
2. $(\mathbf{T}_1, t_1) \equiv_{\Sigma} (\mathbf{T}_2, t_2)$ hviss $\text{chase}(\mathbf{T}_1, t_1, \Sigma) \equiv \text{chase}(\mathbf{T}_2, t_2, \Sigma)$. \dashv