INF2080 – Logikk og beregninger

Forelesning 20: Databaser 2: Avhengigheter og case

UiO : Institutt for informatikk

Sist oppdatert: 2012-05-02 23:20

20.1 Avhengigheter (dependencies)

Oppsummering av dagens forelesning:

- Avhengigheter er formler i FOL som uttrykker semantiske føringer på databaseskjemaet

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

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

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

Definisjon 20.1

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

Definisjon 20.2

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

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

for alle tupler $s, t \in \mathbf{I}$.

Definisjon 20.3

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

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

Definisjon 20.3

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

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

Definisjon 20.4

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

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

Definisjon 20.3

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

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

Definisjon 20.4

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

- $X_1, \ldots, X_n \subseteq U$ *og*

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

Definisjon 20.3

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

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

Definisjon 20.4

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

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

En relasjon I *over* U *tilfredsstiller* $\bowtie [X_1, \ldots, X_n]$ *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, $\mathsf{q}_1 \subseteq_\Sigma \mathsf{q}_2$, dersom $\mathsf{q}_1(\mathbf{I}) \subseteq \mathsf{q}_2(\mathbf{I})$ for hver instans $\mathbf{I}$ som oppfyller hver avhengighet i Σ.

Merk: Når $\mathsf{q}_1 \equiv_\Sigma \mathsf{q}_2$, vil q_1 og q_2 returnere samme svar fra en database som oppfyller Σ!

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 $\leqslant$ på variable. *Chase* defineres på en tablåspørring $(\mathbf{T}, t)$ over R mhp. Σ.

Definisjon 20.7

fd-regelen: La

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.*

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

Definisjon 20.7

fd-regelen: La

- $\sigma \in \Sigma$ *være* $X \to A$

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.*

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

Definisjon 20.7

fd-regelen: La

- $\sigma \in \Sigma$ *være* $X \to A$
- $u, v \in \mathbf{T}$ *slik at* $\pi_X(u) = \pi_X(v)$ *og* $u(A) \neq v(A)$

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.*

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

Definisjon 20.7

fd-regelen: La

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

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

Definisjon 20.8

jd-regelen: La

- $\sigma \in \Sigma$ *være* $\bowtie [X_1, \ldots, X_n]$

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

Definisjon 20.8

jd-regelen: La

- $\sigma \in \Sigma$ *være* $\bowtie [X_1, \ldots, X_n]$
- u *være fritt tuppel over* R *som ikke er i* $\mathbf{T}$

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

Definisjon 20.8

jd-regelen: La

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

σ *anvendt på* $(u_1, \ldots, u_n)$ *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 tablåspørring over* R. *Hvis* q' *er resultatet av å andvende* σ *på noen tupler i* q, *er* $\mathsf{q} \equiv_{\Sigma} \mathsf{q}'$.

Teorem 20.9

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

Definisjon 20.10

En chasing-sekvens er en sekvens av tablå-spørringer, 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 20.9

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

Definisjon 20.10

En chasing-sekvens er en sekvens av tablå-spørringer, 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 $(\mathbf{T}', \mathrm{t}')$ *være resultatet av en terminerende chase-sekvens fra* $(\mathbf{T}, \mathrm{t})$ *mhp.* Σ. *Da vil* $(\mathbf{T}', \mathrm{t}')$, *når det betraktes som en instans, oppfylle* Σ.

Lemma 20.12

La Σ inneholder fd-er og jd-er.

Lemma 20.12

La Σ inneholder fd-er og jd-er.

- *Enhver chasing-sekvens er endelig og kan utvides til en terminerende sekvens.*

Lemma 20.12

La Σ inneholder 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 $(\mathbf{T}, \mathrm{t})$ resulterer i identisk tablå. Dette betegnes chase$(\mathbf{T}, \mathrm{t}, \Sigma)$.*

Lemma 20.12

La Σ inneholder 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 $(\mathbf{T}, \mathrm{t})$ resulterer i identisk tablå. Dette betegnes chase$(\mathbf{T}, \mathrm{t}, \Sigma)$.*

Teorem 20.13

Lemma 20.12

La Σ inneholder 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 $(\mathbf{T}, \mathrm{t})$ resulterer i identisk tablå. Dette betegnes chase$(\mathbf{T}, \mathrm{t}, \Sigma)$.*

Teorem 20.13

1. $(\mathbf{T}_1, \mathrm{t}_1) \subseteq_\Sigma (\mathbf{T}_2, \mathrm{t}_2)$ *hviss chase*$(\mathbf{T}_1, \mathrm{t}_1, \Sigma) \subseteq$ *chase*$(\mathbf{T}_2, \mathrm{t}_2, \Sigma)$.

Lemma 20.12

La Σ inneholder 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 $(\mathbf{T}, \mathrm{t})$ resulterer i identisk tablå. Dette betegnes* $chase(\mathbf{T}, \mathrm{t}, \Sigma)$.

Teorem 20.13

1. $(\mathbf{T}_1, \mathrm{t}_1) \subseteq_\Sigma (\mathbf{T}_2, \mathrm{t}_2)$ *hviss* $chase(\mathbf{T}_1, \mathrm{t}_1, \Sigma) \subseteq chase(\mathbf{T}_2, \mathrm{t}_2, \Sigma)$.
2. $(\mathbf{T}_1, \mathrm{t}_1) \equiv_\Sigma (\mathbf{T}_2, \mathrm{t}_2)$ *hviss* $chase(\mathbf{T}_1, \mathrm{t}_1, \Sigma) \equiv chase(\mathbf{T}_2, \mathrm{t}_2, \Sigma)$.