Databaser fra et logikkperspektiv del 2

Evgenij Thorstensen

IFI, UiO

Høst 2015

Outline

Oppsummering del 1

Vi har vist ekvivalens av

- $\vec{c} \in Ans(\phi, D)$
- $D^{\mathcal{M}} \models \phi[\vec{c}]$
- $f(D^{\mathcal{M}}) \models \phi[\vec{c}]$ (via $\models f(D^{\mathcal{M}}) \rightarrow \phi[\vec{c}]$)
- $f(D^{\mathcal{M}}) \subseteq \phi[\vec{c}]$

Oppsummering del 1

Vi har vist ekvivalens av

- $\vec{c} \in Ans(\phi, D)$
- $D^{\mathcal{M}} \models \phi[\vec{c}]$
- $f(D^{\mathcal{M}}) \models \phi[\vec{c}]$ (via $\models f(D^{\mathcal{M}}) \to \phi[\vec{c}]$)
- $f(D^{\mathcal{M}}) \subseteq \phi[\vec{c}]$

For $\phi \subseteq \psi$, finnes ingen $\phi^{\mathcal{M}}$ generelt sett. I tillegg har vi

Theorem (Trakhtenbrot 1949)

For førsteordens formler ϕ og ψ er $\phi \models \psi$ ikke avgjørbart, selv over endelige modeller.

SPJ-spørringer

En veldig vanlig type spørring i SQL er select-project-join.

SELECT $v_1, v_2 \ldots$ FROM $T_1, T_2 \ldots$ WHERE $Attr_1 = Attr_2$ AND $Attr_3 = Attr_4 \ldots$

Vi kombinerer flere tabeller med likhet mellom attributter.

Denne typen spørring kan vi enkelt karakterisere ved hjelp av det logiske perspektivet.

Konjunktive spørringer

En *konjunktiv spørring (CQ)* er bygd opp av konjunksjoner av atomer og eksistensielle kvantorer. Formelt:

- Alle atomære formler er med i CQ.
- Hvis ϕ og ψ er i CQ, så er $\phi \wedge \psi$ i CQ.
- Hvis ϕ er i CQ, så er $\exists x \phi$ i CQ.

Vi tillater altså ikke negasjon, disjunksjon, etc.

Siden samme variabel alltid tolkes likt, har vi likhet:
$P(x, y) \wedge P(z, v) \wedge y = z$ er det samme som $P(x, y) \wedge P(y, v)$.

Kan skrive konjunktiv ϕ som $\exists x_1, \ldots, \exists x_n . A_1 \wedge \cdots \wedge A_m$ (prenex NF).

QA for konj. spørringer

Vi vet at QA kan gjøres via modellsjekk eller bevis.

For en boolsk konjunktiv spørring, må man finne vitner slik at

$$D^{\mathcal{M}} \models \left(\bigwedge A_i\right)[\vec{c}/\vec{x}]$$

QA for konj. spørringer

Vi vet at QA kan gjøres via modellsjekk eller bevis.

For en boolsk konjunktiv spørring, må man finne vitner slik at

$$D^{\mathcal{M}} \models (\bigwedge A_i)[\vec{c}/\vec{x}]$$

$(\bigwedge A_i)[\vec{c}/\vec{x}]$ er grunn — så per konstruksjon av $D^{\mathcal{M}}$ må hver $A_i[\vec{c}/\vec{x}]$ finnes i D!

Vi kan formalisere dette via *homomorfier*, vanligvis definert på konjunktive spørringer.

Homomorfier

Homomorfi = mapping fra en relational structure til en annen. Vi definerer den på konjunktive spørringer.

Homomorfier

Homomorfi = mapping fra en relational structure til en annen. Vi definerer den på konjunktive spørringer.

La $Q = \exists x_1, \ldots, x_n \phi$ og $R = \exists y_1, \ldots, y_m \psi$ være to konjunktive spørringer med $FV(Q) = FV(R)$.

En *homomorfi* fra Q til R er en substitusjon θ slik at $\phi\theta$ er en delformel av ψ.

Vi skriver $Q \xrightarrow{h} R$ dersom det finnes en homomorfi fra Q til R.

Homomorfien skal ikke endre de frie variablene i Q og R!

Homomorfier

Homomorfi = mapping fra en relational structure til en annen. Vi definerer den på konjunktive spørringer.

La $Q = \exists x_1, \ldots, x_n \phi$ og $R = \exists y_1, \ldots, y_m \psi$ være to konjunktive spørringer med $FV(Q) = FV(R)$.

En *homomorfi* fra Q til R er en substitusjon θ slik at $\phi\theta$ er en delformel av ψ.

Vi skriver $Q \xrightarrow{h} R$ dersom det finnes en homomorfi fra Q til R.

Homomorfien skal ikke endre de frie variablene i Q og R!

$\exists y \exists z.P(x, y, z)$ har en homomorfi $[a/y, x/z]$ til $\exists v.P(x, a, x) \wedge Q(x, c)$.

QA for konjunktive spørringer

Vi husker at $f(D^{\mathcal{M}})$ er en konjunksjon av positive og negative grunne atomer. La $f^{+}(D^{\mathcal{M}})$ være konjunksjonen av de positive.

Theorem

La D *være en database og* $\phi = \exists y_1, \ldots, y_m \psi$ *en konjunktiv spørring med* $FV(\phi) = \{x_1, \ldots, x_n\}$. *Vi har at* $\langle c_1, \ldots, c_n \rangle \in Ans(\phi, D)$ *hvis og bare hvis* $\phi[c_1/x_1, \ldots, c_n/x_n] \xrightarrow{h} f^{+}(D^{\mathcal{M}})$.

QA for konj, bevis

Theorem

La D *være en database og* $\phi = \exists y_1, \ldots, y_m \psi$ *en konjunktiv spørring med* $FV(\phi) = \{x_1, \ldots, x_n\}$*. Vi har at* $\langle c_1, \ldots, c_n \rangle \in Ans(\phi, D)$ *hvis og bare hvis* $\phi[c_1/x_1, \ldots, c_n/x_n] \xrightarrow{h} f^+(D^{\mathcal{M}})$*.*

Hvis-delen: La θ være homomorfien fra $\phi[c_1/x_1, \ldots, c_n/x_n]$ til $f^+(D^{\mathcal{M}})$.

QA for konj, bevis

Theorem

La D *være en database og* $\phi = \exists y_1, \ldots, y_m \psi$ *en konjunktiv spørring med* $\mathsf{FV}(\phi) = \{x_1, \ldots, x_n\}$*. Vi har at* $\langle c_1, \ldots, c_n \rangle \in \mathsf{Ans}(\phi, D)$ *hvis og bare hvis* $\phi[c_1/x_1, \ldots, c_n/x_n] \xrightarrow{h} f^+(D^{\mathcal{M}})$*.*

Hvis-delen: La θ være homomorfien fra $\phi[c_1/x_1, \ldots, c_n/x_n]$ til $f^+(D^{\mathcal{M}})$.

Da er $\psi[c_1/x_1, \ldots, c_n/x_n]\theta$ en *grunn* delformel av $f^+(D^{\mathcal{M}})$. Det betyr at $D^{\mathcal{M}} \models \psi[c_1/x_1, \ldots, c_n/x_n]\theta$.

Per semantikk for eksistenskvantorer, har vi at $D^{\mathcal{M}} \models \psi[c_1/x_1, \ldots, c_n/x_n]$, og derfor at $\langle c_1, \ldots, c_n \rangle \in \mathsf{Ans}(\phi, D)$.

QA for konj, bevis

Theorem

La D *være en database og* $\phi = \exists y_1, \ldots, y_m \psi$ *en konjunktiv spørring med* $FV(\phi) = \{x_1, \ldots, x_n\}$. *Vi har at* $\langle c_1, \ldots, c_n \rangle \in \mathsf{Ans}(\phi, D)$ *hvis og bare hvis* $\phi[c_1/x_1, \ldots, c_n/x_n] \xrightarrow{h} f^+(D^{\mathcal{M}})$.

Bare hvis-delen: Anta $\langle c_1, \ldots, c_n \rangle \in \mathsf{Ans}(\phi, D)$. Da har vi at $D^{\mathcal{M}} \models \phi[c_1/x_1, \ldots, c_n/x_n]$.

Per semantikk, så finnes det domeneelementer $d_1, \ldots, d_m$ slik at $D^{\mathcal{M}} \models A = \psi[c_1/x_1, \ldots, c_n/x_n][\bar{d}_1/y_1, \ldots, \bar{d}_m/y_m]$.

QA for konj, bevis

Theorem

La D *være en database og* $\phi = \exists y_1, \ldots, y_m \psi$ *en konjunktiv spørring med* $FV(\phi) = \{x_1, \ldots, x_n\}$. *Vi har at* $\langle c_1, \ldots, c_n \rangle \in Ans(\phi, D)$ *hvis og bare hvis* $\phi[c_1/x_1, \ldots, c_n/x_n] \xrightarrow{h} f^+(D^{\mathcal{M}})$.

Bare hvis-delen: Anta $\langle c_1, \ldots, c_n \rangle \in Ans(\phi, D)$. Da har vi at $D^{\mathcal{M}} \models \phi[c_1/x_1, \ldots, c_n/x_n]$.

Per semantikk, så finnes det domeneelementer $d_1, \ldots, d_m$ slik at $D^{\mathcal{M}} \models A = \psi[c_1/x_1, \ldots, c_n/x_n][\bar{d}_1/y_1, \ldots, \bar{d}_m/y_m]$.

Siden A er en *grunn* konjunksjon av atomer, følger det at alle atomene er sanne i $D^{\mathcal{M}}$, og at A er en delformel av $f^+(D^{\mathcal{M}})$.

Det betyr at $[\bar{d}_1/y_1, \ldots, \bar{d}_m/y_m]$ er en homomorfi fra $\phi[c_1/x_1, \ldots, c_n/x_n]$ til $f^+(D^{\mathcal{M}})$.

QA for konj. spørringer, oppsummering

Teoremet karakteriserer QA for konjunktive spørringer.

Sier egentlig at QA er å finne homomorfi fra spørring til database.

Åpenbart avgjørbart: Endelig mange homomorfier å sjekke.

Egenskaper ved konjunktive spørringer

Vi kan skrive konjunktiv ϕ som $\exists x_1, \ldots, \exists x_n A_1 \wedge \cdots \wedge A_m$ (prenex NF).

$\exists y \exists z. P(x, y) \wedge P(y, v) \wedge R(x, z, y)$

Dette kan vi representere som en database — la variablene bli konstanter.

Konj. spørringer som databaser

La alle variable x få ny konstant $\dot{x}$. Ny=ikke brukt noe sted.

Definer $\dot{c} = c$ for eksisterende konstanter.

Gitt konjunktiv spørring $\phi = \exists \vec{x}. \bigwedge A_i(t_1, \ldots, t_{i_n})$, lar vi instansen ϕ_D inneholde $A_i(\dot{t}_1, \ldots, \dot{t}_{i_n})$ for hvert atom i ϕ.

$\exists y \exists z. P(x, y) \wedge P(y, v) \wedge R(a, z, y)$

blir

$\{P(\dot{x}, \dot{y}), P(\dot{y}, \dot{v}), R(a, \dot{z}, \dot{y})\}$

Egenskaper ved ϕ_D

Vi skal vise korrekthet av ϕ_D, nemlig følgende

Theorem (Korrekthet av ϕ_D)

La ϕ *og* ψ *være konjunktive spørringer med* $\mathrm{FV}(\phi) = \mathrm{FV}(\psi) = \{x_1, \ldots, x_n\}$. *Vi har* $\mathrm{Ans}(\psi, \phi_D) \neq \emptyset$ *hvis og bare hvis* $\models \forall x_1, \ldots, x_n(\phi \rightarrow \psi)$.

Egenskaper ved ϕ_D

Vi skal vise korrekthet av ϕ_D, nemlig følgende

Theorem (Korrekthet av ϕ_D)

La ϕ og ψ være konjunktive spørringer med $FV(\phi) = FV(\psi) = \{x_1, \ldots, x_n\}$. Vi har $\mathrm{Ans}(\psi, \phi_D) \neq \emptyset$ hvis og bare hvis $\models \forall x_1, \ldots, x_n(\phi \rightarrow \psi)$.

Vil trenge følgende lemma:

Homomorfier er transitive: $\phi \xrightarrow{h} \psi$ og $\psi \xrightarrow{h} \lambda$ impliserer at $\phi \xrightarrow{h} \lambda$.
Bevis: Ukeoppgave.

Korrekthet av ϕ_D

Theorem (Korrekthet av ϕ_D)

La ϕ og ψ være konjunktive spørringer med $FV(\phi) = FV(\psi) = \{x_1, \ldots, x_n\}$. Vi har $Ans(\psi, \phi_D) \neq \emptyset$ hvis og bare hvis $\models \forall x_1, \ldots, x_n(\phi \rightarrow \psi)$.

Hvis-delen. Anta $\models \forall x_1, \ldots, x_n(\phi \rightarrow \psi)$. Det betyr at $(\phi_D)^{\mathcal{M}} \models \forall x_1, \ldots, x_n(\phi \rightarrow \psi)$.

Holder å vise at $(\phi_D)^{\mathcal{M}} \models \phi[\vec{c}]$ for en $\vec{c}$. Da får vi at $(\phi_D)^{\mathcal{M}} \models \psi[\vec{c}]$, og dermed at $\vec{c} \in Ans(\psi, \phi_D)$.

Fra konstruksjonen av ϕ_D har vi at $\langle \dot{x}_1, \ldots, \dot{x}_n \rangle \in Ans(\phi, \phi_D)$. Ergo har vi $(\phi_D)^{\mathcal{M}} \models \phi[\dot{x}_1/x_1, \ldots, \dot{x}_n/x_n]$ også.

Korrekthet av ϕ_D, forts.

Theorem (Korrekthet av ϕ_D)

La ϕ og ψ være konjunktive spørringer med $FV(\phi) = FV(\psi) = \{x_1, \ldots, x_n\}$. Vi har $Ans(\psi, \phi_D) \neq \emptyset$ hvis og bare hvis $\models \forall x_1, \ldots, x_n(\phi \to \psi)$.

Bare hvis. Anta at det finnes $\langle c_1, \ldots, c_n \rangle \in Ans(\psi, \phi_D)$. Da har vi $\psi[c_1/x_1, \ldots, c_n/x_n] \xrightarrow{h} f^+((\phi_D)^{\mathcal{M}})$ med homomorfi θ.

La oss se på $\psi[c_1/x_1, \ldots, c_n/x_n] = \exists y_1, \ldots, y_m \psi'$. I $\psi'\theta$ blir alle y_i til $\dot{t}$, hvor t er termer fra ϕ.

Korrekthet av ϕ_D, forts.

Theorem (Korrekthet av ϕ_D)

La ϕ og ψ være konjunktive spørringer med $FV(\phi) = FV(\psi) = \{x_1, \ldots, x_n\}$. Vi har $Ans(\psi, \phi_D) \neq \emptyset$ hvis og bare hvis $\models \forall x_1, \ldots, x_n(\phi \to \psi)$.

Bare hvis. Anta at det finnes $\langle c_1, \ldots, c_n \rangle \in Ans(\psi, \phi_D)$. Da har vi $\psi[c_1/x_1, \ldots, c_n/x_n] \xrightarrow{h} f^+((\phi_D)^{\mathcal{M}})$ med homomorfi θ.

La oss se på $\psi[c_1/x_1, \ldots, c_n/x_n] = \exists y_1, \ldots, y_m \psi'$. I $\psi'\theta$ blir alle y_i til $\dot{t}$, hvor t er termer fra ϕ.

Vi definerer en ny substitusjon θ^{-1} ved å la $y\theta^{-1} = t$ hvis og bare hvis $y\theta = \dot{t}$.

Siden θ er en homomorfi fra $\psi[c_1/x_1, \ldots, c_n/x_n]$ til ϕ_D, er θ^{-1} en homomorfi fra $\psi[c_1/x_1, \ldots, c_n/x_n]$ til $\phi[c_1/x_1, \ldots, c_n/x_n]$, og dermed fra ψ til ϕ.

Korrekthet av ϕ_D, forts.

Theorem (Korrekthet av ϕ_D)

La ϕ *og* ψ *være konjunktive spørringer med* $\mathsf{FV}(\phi) = \mathsf{FV}(\psi) = \{x_1, \ldots, x_n\}$. *Vi har* $Ans(\psi, \phi_D) \neq \emptyset$ *hvis og bare hvis* $\models \forall x_1, \ldots, x_n(\phi \rightarrow \psi)$.

Vi har vist at $\psi \xrightarrow{h} \phi$. Det kan vi bruke til å vise $\models \forall x_1, \ldots, x_n(\phi \rightarrow \psi)$.

Vi husker at $\models \forall x_1, \ldots, x_n(\phi \rightarrow \psi)$ er ekvivalent med $\phi \subseteq \psi$. La D og $\langle a_1, \ldots, a_n \rangle \in Ans(\phi, D)$ være vilkårlig. Da gjelder $\phi[a_1/x_1, \ldots, a_n/x_n] \xrightarrow{h} f^+(D^{\mathcal{M}})$, og per lemma om transitivitet at $\psi[a_1/x_1, \ldots, a_n/x_n] \xrightarrow{h} f^+(D^{\mathcal{M}})$.

Ekvivalens av QC og QA

Vi har vist at QA og QC for konjunktive spørringer er samme problem.

Dette problemet er $\models \forall x_1, \ldots, x_k(\phi \rightarrow \psi)$.

Funker også motsatt vei — logisk konsekvens som databaseproblem.

Korrekthet av ϕ_D, ekstra

I beviset for bare hvis-delen av korrekthet av ϕ_D beviste vi følgende underveis:

$$\psi \xrightarrow{h} \phi \text{ impliserer } \phi \subseteq \psi$$

Korrekthet av ϕ_D, ekstra

I beviset for bare hvis-delen av korrekthet av ϕ_D beviste vi følgende underveis:

$$\psi \xrightarrow{h} \phi \text{ impliserer } \phi \subseteq \psi$$

Theorem (Homomorfiteoremet)

La ϕ og ψ være konjunktive spørringer med $\mathsf{FV}(\phi) = \mathsf{FV}(\psi)$. Da har vi at $\phi \subseteq \psi$ hvis og bare hvis $\psi \xrightarrow{h} \phi$.

Homomorfiteoremet

Theorem (Homomorfiteoremet)

La ϕ *og* ψ *være konjunktive spørringer med* $\mathsf{FV}(\phi) = \mathsf{FV}(\psi)$. *Da har vi at* $\phi \subseteq \psi$ *hvis og bare hvis* $\psi \xrightarrow{h} \phi$.

Bare hvis-delen: Anta $\phi \subseteq \psi$. Per korrekthet av ϕ_D følger det at $Ans(\psi, \phi_D) \neq \emptyset$.

Hva mer, fra $Ans(\phi, \phi_D) \neq \emptyset$ kan man bruke beviset for korrekthet av ϕ_D til å vise $\psi \xrightarrow{h} \phi$.

QC er QA for konj. spørringer

Theorem (QC er QA)

La ϕ *og* ψ *være konjunktive spørringer med* $\mathsf{FV}(\phi) = \mathsf{FV}(\psi) = \{x_1, \ldots, x_n\}$. *Vi har at* $\phi \subseteq \psi$ *hvis og bare hvis* $Ans(\psi, \phi_D) \neq \emptyset$.

Per teorem om sammenheng mellom konsekvens og QC, er $\phi \subseteq \psi$ ekvivalent med $\models \forall x_1, \ldots, x_n(\phi \to \psi)$. Dette er ekvivalent med $Ans(\psi, \phi_D) \neq \emptyset$ per korrekthet av ϕ_D.

Oppsummering

For konjunktive spørringer er QA og QC inter-reduserbare.

QA er $\phi[\vec{c}] \xrightarrow{h} f^{+}(D^{\mathcal{M}})$.

QC er $\psi \xrightarrow{h} \phi$.

Oppsummering

For konjunktive spørringer er QA og QC inter-reduserbare.

QA er $\phi[\vec{c}] \xrightarrow{h} f^{+}(D^{\mathcal{M}})$.

QC er $\psi \xrightarrow{h} \phi$.

Homomorfiteoremet har mange flere anvendelser. Mer i Alice-boken.

Eksempel: QC under integritetsregler, $\Sigma \models \forall x_1, \ldots, x_k(\phi \rightarrow \psi)$.