

IN2090 – Databaser og datamodellering

02 – Relasjonsalgebra

Leif Harald Karlsen
leifhka@ifi.uio.no



Universitetet i Oslo

Hva er relasjonsalgebra?

- ◆ Vi har nå sett hvordan vi representerer data
- ◆ Ikke hvordan vi bruker dataene
- ◆ Med bruk mener vi:
 - ◆ Hente ut spesifikke data basert på én eller fler relasjoner
 - ◆ Kombinere og transformere relasjoner til nye relasjoner
- ◆ Relasjonsalgebra er et (presist/matematisk) språk for å bruke dataene beskrevet med relasjoner
- ◆ Danner fundamentet for databasespråket SQL

Relasjonsalgebra

- ◆ En algebra er en mengde med operasjoner/funksjoner
- ◆ Operasjonene/funksjonene har argumenter og resultat fra samme mengde/type
- ◆ Eksempel på en vanlig algebra: $\{+, -, \times, \div\}$ over reelle tall
 - ◆ F.eks. $(5 \times (3 + 4)) \div 4 = 8.75$
- ◆ Annet eksempel på en vanlig algebra: $\{\cap, \cup, \setminus\}$ over mengder
 - ◆ F.eks. $(\{1\} \cup (\{3, 4\} \setminus \{4, 5\})) \cap \{1, 2, 3\} = \{1, 3\}$
- ◆ *Relasjonsalgebra*: mengde med operasjoner over relasjoner
- ◆ En operasjon i relasjonsalgebra tar derfor én eller to relasjoner som argument, og returnerer en ny relasjon
- ◆ En slik algebra lar oss kombinere og hente ut data fra relasjonene

Projeksjon (π)

$\pi_{\text{Brnavn, Etternavn}}(\text{Student}) =$

Brnavn	Etternavn
evgenit	Thorstensen
peternl	Nilsen
leifhka	Karlsen

- ◆ π er unær (tar ett argument)
- ◆ Har en mengde med attributter som subskrift
- ◆ Returnerer ny relasjon med kun attributtene listet opp
- ◆ Den velger altså ut attributter
- ◆ Merk: Operasjonene *endrer* ikke de originale relasjonene i databasen, bare returnerer en ny relasjon
- ◆ Merk: Resultatet er en relasjon uten navn. Kan gi den et navn slik:

$\text{StudentNavn} := \pi_{\text{Brnavn, Etternavn}}(\text{Student})$

Seleksjon (σ)

$$\sigma_{\text{Emnummer} \geq 3000 \wedge \text{AntSP} = 10}(\text{Kurs})$$

=

Fagkode	Emnummer	Tittel	AntSP
AST	3220	Kosmologi 1	10
IN	4070	Logikk	10

- ◆ σ er også unær (tar ett argument)
- ◆ Har et uttrykk med attributt-navnene som variable som subskrift
 - ◆ Bruker symbolene $\wedge, \vee, \neg, \geq, \leq, >, <, =$ og konstanter (3000, 10, *IN*, osv.)
- ◆ Returnerer ny relasjon med kun de tuplene som tilfredstiller uttrykket
- ◆ Men samme attributter
- ◆ Den velger altså ut de radene vi er interessert i

Omdøping (ρ)

$\rho_{\text{Tittel} \rightarrow \text{Navn}, \text{AntSP} \rightarrow \text{Poeng}}(\text{Kurs})$

=

Fagkode	Emnernummer	Navn	Poeng
IN	2090	Databaser og datamodellering	10
IN	2010	Algoritmer og datastrukturer	10
AST	3220	Kosmologi 1	10
MAT	2000	Prosjektarbeid	10
BIOS	9312	Alpine Ecology	5
IN	4070	Logikk	10

- ◆ ρ er også unær (tar ett argument)
- ◆ Har en mengde med attributt-pil-attributt som subskrift
- ◆ $A \rightarrow B$ sier at A skal omdøpes til B
- ◆ Returnerer ny relasjon med nye attributt-navn ihht. omdøpingene
- ◆ Endrer ellers ingenting
- ◆ Enkelte kilder bruker en litt annen syntaks

Eksempel: Unære operasjoner

Kurs			
Fagkode	Emnernr	Tittel	AntSP
IN	2090	Databaser og datamodellering	10
IN	2010	Algoritmer og datastrukturer	10
AST	3220	Kosmologi 1	10
MAT	2000	Prosjektarbeid	10
IN	4070	Logikk	10
BIOS	9312	Alpine Ecology	5
IN	5800	Declarative data engineering	10
MAT	4500	Toplogi	10
IN	4230	Nettverk	10

Finn tittel på alle informatikk-masterkurs:

$$\rho_{\text{Tittel} \rightarrow \text{IfiMasterKurs}}(\pi_{\text{Tittel}}(\sigma_{\text{Fagkode} = \text{'IN'} \wedge \text{Emnenummer} \geq 4000}(\text{Kurs})))$$

=

Fagkode	Emnernr	Tittel	AntSP
IN	2090	Databaser og datamodellering	10
IN	2010	Algoritmer og datastrukturer	10
AST	3220	Kosmologi 1	10
MAT	2000	Prosjektarbeid	10
IN	4070	Logikk	10
BIOS	9312	Alpine Ecology	5
IN	5800	Declarative data engineering	10
MAT	4500	Toplogi	10
IN	4230	Nettverk	10

Kartesisk produkt (\times)

Student \times Karakter

=

Brnavn	Navn	Etternavn	Adresse	Student	Fagkode	Emnenummer	Kara
evgenit	Evgenij	Thorstensen	Gateveien 1a	evgenit	IN	2090	B
evgenit	Evgenij	Thorstensen	Gateveien 1a	peternl	AST	3220	A
evgenit	Evgenij	Thorstensen	Gateveien 1a	evgenit	IN	2010	B
evgenit	Evgenij	Thorstensen	Gateveien 1a	leifhka	IN	2090	B
evgenit	Evgenij	Thorstensen	Gateveien 1a	leifhka	MAT	2000	C
peternl	Petter	Nilsen	Stedplassen 23	evgenit	IN	2090	B
peternl	Petter	Nilsen	Stedplassen 23	peternl	AST	3220	A
peternl	Petter	Nilsen	Stedplassen 23	evgenit	IN	2010	B
peternl	Petter	Nilsen	Stedplassen 23	leifhka	IN	2090	B
peternl	Petter	Nilsen	Stedplassen 23	leifhka	MAT	2000	C
leifhka	Leif H.	Karlsen	Bergfjellet 42	evgenit	IN	2090	B
leifhka	Leif H.	Karlsen	Bergfjellet 42	peternl	AST	3220	A
leifhka	Leif H.	Karlsen	Bergfjellet 42	evgenit	IN	2010	B
leifhka	Leif H.	Karlsen	Bergfjellet 42	leifhka	IN	2090	B
leifhka	Leif H.	Karlsen	Bergfjellet 42	leifhka	MAT	2000	C

- ◆ \times er binær (tar to argumenter)
- ◆ Returnerer ny relasjon med:
 - ◆ Alle attributtene til begge relasjonene
 - ◆ Alle kombinasjoner av tupler fra de to relasjonene

Eksempel: Kartesisk produkt (\times)

Finn info om alle studenter og deres karakterer:

$$\sigma_{\text{Brnavn}=\text{Student}} \left(\begin{array}{|c|c|c|c|} \hline \text{Brnavn} & \text{Navn} & \text{Etternavn} & \text{Adresse} \\ \hline \text{evgenit} & \text{Evgenij} & \text{Thorstensen} & \text{Gateveien 1a} \\ \text{peternl} & \text{Petter} & \text{Nilsen} & \text{Stedplassen 23} \\ \text{leifhka} & \text{Leif H.} & \text{Karlsen} & \text{Bergfjellet 42} \\ \hline \end{array} \times \begin{array}{|c|c|c|c|} \hline \text{Student} & \text{Fagkode} & \text{Emnummer} & \text{Kara} \\ \hline \text{evgenit} & \text{IN} & 2090 & \text{B} \\ \text{peternl} & \text{AST} & 3220 & \text{A} \\ \text{evgenit} & \text{IN} & 2010 & \text{B} \\ \text{leifhka} & \text{IN} & 2090 & \text{B} \\ \text{leifhka} & \text{MAT} & 2000 & \text{C} \\ \hline \end{array} \right)$$

$$= \begin{array}{|c|c|c|c|c|c|c|c|} \hline \text{Brnavn} & \text{Navn} & \text{Etternavn} & \text{Adresse} & \text{Student} & \text{Fagkode} & \text{Emnummer} & \text{Kara} \\ \hline \text{evgenit} & \text{Evgenij} & \text{Thorstensen} & \text{Gateveien 1a} & \text{evgenit} & \text{IN} & 2090 & \text{B} \\ \text{evgenit} & \text{Evgenij} & \text{Thorstensen} & \text{Gateveien 1a} & \text{peternl} & \text{AST} & 3220 & \text{A} \\ \text{evgenit} & \text{Evgenij} & \text{Thorstensen} & \text{Gateveien 1a} & \text{evgenit} & \text{IN} & 2010 & \text{B} \\ \text{evgenit} & \text{Evgenij} & \text{Thorstensen} & \text{Gateveien 1a} & \text{leifhka} & \text{IN} & 2090 & \text{B} \\ \text{evgenit} & \text{Evgenij} & \text{Thorstensen} & \text{Gateveien 1a} & \text{leifhka} & \text{MAT} & 2000 & \text{C} \\ \text{peternl} & \text{Petter} & \text{Nilsen} & \text{Stedplassen 23} & \text{evgenit} & \text{IN} & 2090 & \text{B} \\ \text{peternl} & \text{Petter} & \text{Nilsen} & \text{Stedplassen 23} & \text{peternl} & \text{AST} & 3220 & \text{A} \\ \text{peternl} & \text{Petter} & \text{Nilsen} & \text{Stedplassen 23} & \text{evgenit} & \text{IN} & 2010 & \text{B} \\ \text{peternl} & \text{Petter} & \text{Nilsen} & \text{Stedplassen 23} & \text{leifhka} & \text{IN} & 2090 & \text{B} \\ \text{peternl} & \text{Petter} & \text{Nilsen} & \text{Stedplassen 23} & \text{leifhka} & \text{MAT} & 2000 & \text{C} \\ \text{leifhka} & \text{Leif H.} & \text{Karlsen} & \text{Bergfjellet 42} & \text{evgenit} & \text{IN} & 2090 & \text{B} \\ \text{leifhka} & \text{Leif H.} & \text{Karlsen} & \text{Bergfjellet 42} & \text{peternl} & \text{AST} & 3220 & \text{A} \\ \text{leifhka} & \text{Leif H.} & \text{Karlsen} & \text{Bergfjellet 42} & \text{evgenit} & \text{IN} & 2010 & \text{B} \\ \text{leifhka} & \text{Leif H.} & \text{Karlsen} & \text{Bergfjellet 42} & \text{leifhka} & \text{IN} & 2090 & \text{B} \\ \text{leifhka} & \text{Leif H.} & \text{Karlsen} & \text{Bergfjellet 42} & \text{leifhka} & \text{MAT} & 2000 & \text{C} \\ \hline \end{array}$$

Join (\bowtie)

Student $\bowtie_{\text{Brnavn=Student}}$ Karakter

=

Brnavn	Navn	Etternavn	Adresse	Student	Fagkode	Emnenummer	Kara
evgenit	Evgenij	Thorstensen	Gateveien 1a	evgenit	IN	2090	B
evgenit	Evgenij	Thorstensen	Gateveien 1a	evgenit	IN	2010	B
peternl	Petter	Nilsen	Stedplassen 23	peternl	AST	3220	A
leifhka	Leif H.	Karlsen	Bergfjellet 42	leifhka	IN	2090	B
leifhka	Leif H.	Karlsen	Bergfjellet 42	leifhka	MAT	2000	C

- ◆ \bowtie er binær (tar to argumenter)
- ◆ Har en mengde med attributt= -- attributt som subskrift
- ◆ $A = B$ sier at A skal være lik B i resultatet
- ◆ Kan også ha mer generelle uttrykk (slik som for σ)
- ◆ Returnerer ny relasjon med alle kombinasjoner av tupler fra de to relasjonene som tilfredstiller uttrykket (med alle attributtene fra begge relasjonene)
- ◆ Merk: $R \bowtie_e P$ er ekvivalent med $\sigma_e(R \times P)$.

Eksempel: Join (\bowtie)

Finn karakterer med all info om kurset:

$\pi_{\text{Student, Fagkode, Emnummer, Karakter, Tittel, AntSP}}$ (

Student	Fagkode	Emnummer	Kara
evgenit	IN	2090	B
peternl	AST	3220	A
evgenit	IN	2010	B
leifhka	IN	2090	B
leifhka	MAT	2000	C

$\bowtie_{\text{Fagkode} = \text{Fagkode} \wedge \text{Emnummer} = \text{Emnummer}}$???

$\rho_{\text{Fagkode} \rightarrow \text{Kode}, \text{Emnummer} \rightarrow \text{Nummer}}$

Fagkode	Emnummer	Tittel	AntSP
IN	2090	Databaser og datamodellering	10
IN	2010	Algoritmer og datastrukturer	10
AST	3220	Kosmologi 1	10
MAT	2000	Prosjektarbeid	10
BIOS	9312	Alpine Ecology	5
IN	4070	Logikk	10

)

=

Student	Fagkode	Emnummer	Kara	Kode	Nummer	Tittel	AntSP
evgenit	IN	2090	B	IN	2090	Databaser og datamodellering	10
peternl	AST	3220	A	AST	3220	Kosmologi 1	10
evgenit	IN	2010	B	IN	2010	Algoritmer og datastrukturer	10
leifhka	IN	2090	B	IN	2090	Databaser og datamodellering	10
leifhka	MAT	2000	C	MAT	2000	Prosjektarbeid	10

Eksempel: Join (\bowtie)

Finn karakterer med all info om kurset:

$$\pi_{\text{Student, Fagkode, Emnummer, Karakter, Tittel, AntSP}} \left(\begin{array}{l} \textit{Karakterer} \\ \bowtie_{\text{Fagkode} = \text{Kode} \wedge \text{Emnummer} = \text{Nummer}} \\ \rho_{\text{Fagkode} \rightarrow \text{Kode}, \text{Emnummer} \rightarrow \text{Nummer}} (\textit{Kurs}) \end{array} \right)$$

=

Student	Fagkode	Emnummer	Kara	Tittel	AntSP
evgenit	IN	2090	B	Databaser og datamodellering	10
peternl	AST	3220	A	Kosmologi 1	10
evgenit	IN	2010	B	Algoritmer og datastrukturer	10
leifhka	IN	2090	B	Databaser og datamodellering	10
leifhka	MAT	2000	C	Prosjektarbeid	10

Naturlig join (\star)

Student	Fagkode	Emnenummer	Kara
evgenit	IN	2090	B
peternl	AST	3220	A
evgenit	IN	2010	B
leifhka	IN	2090	B
leifhka	MAT	2000	C

\star

Fagkode	Emnernummer	Tittel	AntSP
IN	2090	Databaser og datamodellering	10
IN	2010	Algoritmer og datastrukturer	10
AST	3220	Kosmologi 1	10
MAT	2000	Prosjektarbeid	10
BIOS	9312	Alpine Ecology	5
IN	4070	Logikk	10

=

Student	Fagkode	Emnenummer	Kara	Tittel	AntSP
evgenit	IN	2090	B	Databaser og datamodellering	10
peternl	AST	3220	A	Kosmologi 1	10
evgenit	IN	2010	B	Algoritmer og datastrukturer	10
leifhka	IN	2090	B	Databaser og datamodellering	10
leifhka	MAT	2000	C	Prosjektarbeid	10

- ◆ \star er binær (tar to argumenter)
- ◆ Utfører en join på alle attributter med likt navn (og velger ut én kopi av hver)
- ◆ Dersom ingen attributter har likt navn får man det kartesiske produktet (\times)
- ◆ Må bruke ρ dersom enten:
 - ◆ Man ønsker å joine på attributter med ulikt navn (bruker ρ for å gjøre dem like)
 - ◆ Man ønsker ikke å joine på to attributter som har likt navn (bruker ρ for å gjøre dem ulike)

Snitt, union og differanse (\cap , \cup , \setminus)

Bachelorkurs

Fagkode	Emnernr	Tittel	AntSP
IN	2090	Databaser og datamodellering	10
IN	2010	Algoritmer og datastrukturer	10
AST	3220	Kosmologi 1	10
MAT	2000	Prosjektarbeid	10

Masterkurs

Fagkode	Emnernr	Tittel	AntSP
IN	4070	Logikk	10
BIOS	9312	Alpine Ecology	5
IN	5800	Declarative data engineering	10
MAT	4500	Toplogi	10

Bachelorkurs \cup Masterkurs =

Fagkode	Emnernr	Tittel	AntSP
IN	2090	Databaser og datamodellering	10
IN	2010	Algoritmer og datastrukturer	10
AST	3220	Kosmologi 1	10
MAT	2000	Prosjektarbeid	10
IN	4070	Logikk	10
BIOS	9312	Alpine Ecology	5
IN	5800	Declarative data engineering	10
MAT	4500	Toplogi	10

- ◆ \cup , \cap og \setminus er alle binære (tar to argumenter)
- ◆ Utfører vanlig union, snitt og differanse på mengden av tupler
- ◆ Krever at begge argument-relasjonene har samme signatur (attributt-navn og typer)

Større eksempel

Student	Fagkode	Emnenummer	Kara
evgenit	IN	3220	A
evgenit	IN	2090	D
peternl	AST	3220	A
evgenit	IN	2010	B
leifhka	IN	2090	C
peternl	MAT	2000	A
leifhka	MAT	2000	E
erihan	IN	2090	A

Brnavn	Navn	Etternavn	Adresse
evgenit	Evgenij	Thorstensen	Gateveien 1a
peternl	Petter	Nilsen	Stedplassen 23
leifhka	Leif H.	Karlsen	Bergfjellet 42
idamo	Ida	Mo	Knutekrysset 78b
erihan	Erik	Hansen	Bryggekaia 4

Finn navn på alle studenter som kun har fått A'er:

$\pi_{\text{Navn, Etternavn}}((\pi_{\text{Student}}(\sigma_{\text{Kara}='A'}(\text{Karakter}))) \setminus \pi_{\text{Student}}(\sigma_{\text{Kara} \neq 'A'}(\text{Karakter}))) \bowtie_{\text{Brnavn} = \text{Student}} \text{Student}$

=

Student	Fagkode	Emnenummer	Kara
evgenit	IN	3220	A
evgenit	IN	2090	B
peternl	AST	3220	A
evgenit	IN	2010	B
leifhka	IN	2090	B
peternl	MAT	2000	A
leifhka	MAT	2000	C
erihan	IN	2090	A

Takk for nå!

Neste uke handler om datamodellering.