# Algebra

**Definition**. *Signatur*:

- et antal *sorter* eller *typer*: (*simple* eller *strukturerede*).
- et antal *konstantsymboler* med type, og
- en antal *operatorsymboler* med *rang*
  $$\tau_1, \ldots, \tau_n \to \tau,$$
  hvor $\tau$'erne er typer.

**Eksempel:** Signatur $\Sigma_1$ svarende til regning med tal:

- typerne $\mathcal{Z}$, $\mathcal{B}$
- konstantsymbolerne $\ldots, -1, 0, 1, \ldots$ af type $\mathcal{Z}$
- konstantsymbolerne "sand", "falsk" af type $\mathcal{B}$
- Operatorsymboler $+ : \mathcal{Z}, \mathcal{Z} \to \mathcal{Z}$, $>: \mathcal{Z}, \mathcal{Z} \to \mathcal{B}$ osv.

**Definition**. *Term* over signatur $\Sigma$:

- Et konstantsymbol $c$ af type $\tau$ er en term af type $\tau$.
- Hvis $t_1, \ldots, t_n$ er termer af type $\tau_1, \ldots, \tau_n$ og *op* er et operatorsymbol af rang $\tau_1, \ldots, \tau_n \to \tau$, da er $op(t_1, \ldots, t_n)$ er term af type $\tau$.
- Intet andet er termer.

I praksis: varieret notation, f.eks. "2+2=4" for "=(+(2,2),4)"; "(1+2)*3" for "*(+(1,2),3)".

**Eksempel:** Under $\Sigma_1$ er "2+2" en term af type $\mathcal{Z}$, og "2+2=4" og "2+2=5" er termer af type $\mathcal{B}$.

**Definition**. En $\Sigma$-*algebra* $\mathcal{A}$ for en signatur $\Sigma$:

- for hver type $\tau$ i $\Sigma$ en mængde $\tau^{\mathcal{A}}$,
- for hvert konstantsymbol $c$ af type $\tau$, et element $c^{\mathcal{A}} \in \tau^{\mathcal{A}}$, og
- for hvert operatorsymbol $f\colon type_1, \ldots, type_n \rightarrow type_{n+1}, \ldots, type_{n+1}$ en funktion $f^{\mathcal{A}}\colon type_1^{\mathcal{A}}, \ldots, type_n^{\mathcal{A}} \rightarrow type_{n+1}^{\mathcal{A}}, \ldots, type_{n+k}^{\mathcal{A}}$

**OBS:** En $\Sigma$-*algebra* $\mathcal{A}$ definerer på naturlig måde til en funktion, som evaluerer vilkårlig $\Sigma$-term $t$ til værdi $t^{A}$.

**OBS:** En signatur kan have mange forskellige algebraer, bl.a.:

- en *kanoniske algebra*,
- *termalgebraen*,
- *abstrakte fortolkninger*

**Eksempel.** Kanonisk $\Sigma_1$-algebra kaldet $N$:

- $\mathcal{Z} \longrightarrow$ mængden af alle hele tal
  $\mathcal{B} \longrightarrow$ mængde af to sandhedsværdier
- $0 \longrightarrow$ det matematiske nul
  $1 \longrightarrow$ det matematiske ét-tal, osv.
- $+ \longrightarrow$ den matematiske funktion, som lægger to tal sammen

F.eks. $(2 + 2)^N \longrightarrow$ det matematiske fire-tal; "$(2 + 2 = 4)^N$" $\longrightarrow$ logisk falskhed

**NB:** $\cdots^N$ ikke er total; "$(1/0)^N$" udefineret.

**Eksempel.** Termalgebra for $\Sigma_1$-algebra $T\Sigma_1$:

- $\mathcal{Z} \longrightarrow$ mængden termer af type $\mathcal{Z}$
  $\mathcal{B} \longrightarrow$ mængden af termer af type $\mathcal{B}$
- $0 \longrightarrow$ termen 0
  $1 \longrightarrow$ termen 1, osv.
- $+ \longrightarrow$ funktion, der tager to termer $t_1$ og $t_2$ og danner termen $t_1 + t_2$, osv.

F.eks. $(2+2)^{T\Sigma_1} \longrightarrow$ termen "$2+2$" som noget andet end termen "4"!!!!
"$(2+2=5)^{T\Sigma_1}$" $\longrightarrow$ termen "$2+2=5$"
**NB:** $\cdots^{T\Sigma_1}$ er total; "$(1/0)^{T\Sigma_1}$" $\longrightarrow$ termen "$1/0$"

**Eksempel** på algebra for abstrakt fortolkning:

- $\mathcal{Z} \longrightarrow \{\text{lige}, \text{ulige}, \text{ved-ikke}\}$
- $\ldots, -4, -2, 0, 2, 4, \ldots \longrightarrow$ "lige"
- $\ldots, -3, -1, 1, 3, 5, \ldots \longrightarrow$ "ulige"

+:

| | | |
|---|---|---|
| lige | lige | lige |
| lige | ulige | ulige |
| ulige | lige | ulige |
| ulige | ulige | lige |
| ved-ikke | lige | ved-ikke |
| ved-ikke | ulige | ved-ikke |
| lige | ved-ikke | ved-ikke |
| ulige | ved-ikke | ved-ikke |

∗:

| | | |
|---|---|---|
| lige | lige | lige |
| lige | ulige | lige |
| ulige | lige | lige |
| ulige | ulige | ulige |
| ved-ikke | lige | lige |
| ved-ikke | ulige | ved-ikke |
| lige | ved-ikke | lige |
| ulige | ved-ikke | ved-ikke |

**Eksempel** på "abstrakt evaluering":

$$1 + 2 * 3 * x + 4$$

$$\leadsto \text{ulige} + \text{lige} * \text{ulige} * \text{ved-ikke} + \text{lige}$$

$$\leadsto \text{ulige} + \text{lige} * \text{ved-ikke} + \text{lige}$$

$$\leadsto \text{ulige} + \text{lige} + \text{lige}$$

$$\leadsto \text{ulige} + \text{lige}$$

$$\leadsto \text{lige}$$

# Mængder som en algebra

Signaturen:

Typerne *Mængde*, *Element* og *Bool*.

Konstantsymboler af type *Mængde*: $\mathcal{N}$, $\mathcal{Z}$, $\mathcal{B}$, $\mathcal{T}$, $\mathcal{U}$ og $\emptyset$.

Konstantsymboler af type *Element*: de sædvanlige talkonstanter, tekststrenge osv.

OBS: *Bool* er en særskilt type, som her er noget andet end konstanten $\mathcal{B}$ af type *Mængde*.

Nogle af operatorsymbolerne i signaturen:

> $\in$: *Element*, *Mængde* $\rightarrow$ *Bool*
>
> $\notin$: *Element*, *Mængde* $\rightarrow$ *Bool*
>
> $\cup$: *Mængde*, *Mængde* $\rightarrow$ *Mængde*
>
> $\subseteq$: *Mængde*, *Mængde* $\rightarrow$ *Bool*
>
> $\{\ldots\}$: *Element*, $\ldots$, *Element* $\rightarrow$ *Mængde*
>
> $\cdots \times \cdots \times \cdots$: *Mængde*, $\ldots$, *Mængde* $\rightarrow$ *Mængde*
>
> $\langle \cdots, \cdots, \cdots \rangle$: *Element*, $\ldots$, *Element* $\rightarrow$ *Element*
>
> – en usynlig operator med aritet *Mængde* $\rightarrow$ *Element* (så vi kan have mængder af mængder)

Kanoniske algebra: afbilder konstantsymbolerne over i deres respektive elementer eller mængder og operatorsymbolerne over i de funktioner, vi forventer.

**Eksempel:** betragt termen $\langle \text{abe}, 7 \rangle \in \{\{1,2\}, \emptyset\} \cup \mathcal{T} \times \mathcal{N}$ af type *Bool*. Delterm $\{\{1,2\}, \emptyset\} \cup \mathcal{T} \times \mathcal{N}$ af type *Bool* evaluerer til mængde som bl.a. indeholder elementer *noteret ved*:

> $\{1,2\}$, $\emptyset$, $\langle \text{abe}, 7 \rangle$, $\langle \text{gris}, 4217 \rangle$.

**At regne med tal, at regne med mængder, at regne med symboler: Det er bare algebra altsammen!**

# Relationel algebra: Mængdelære med typer og navngivning

Men først **redefinere kendte begreber inden algebraen introduceres**.

Mgd. af *attributter* $\mathcal{A}$ — arbitrære navne.

**Ny definition af tupler:** Lad $A$ være endelig delmængde af $\mathcal{A}$; en $A$-*tupel* er en total funktion fra $A$ over i $\mathcal{U}$.

**Notation:** Hvis $A = \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$ er flg. en $A$-tupel:

$$\langle \mathrm{a}\colon 10, \mathrm{b}\colon 20, \mathrm{c}\colon 30\rangle$$

Komponenter vælges ved prik.notation, f.eks. $t.\mathrm{a} = 10$, $t.\mathrm{b} = 20$ og $t.\mathrm{c} = 30$.

**Ny definition af mængdeprodukt:** Lad $A = \{a_1, \ldots, a_n\}$ være endelig delmængde af $\mathcal{A}$; *produktet*

$$a_1\colon M_1, \ldots, a_n\colon M_n$$

består af samtlige $A$-tupler $t$, hvor $t.a_i \in M_i$ for alle $i = 1, \ldots, n$.

**OBS:** $(\mathrm{a}\colon \mathcal{T}, b\colon \mathcal{N}) = (b\colon \mathcal{N}, \mathrm{a}\colon \mathcal{T})$ og $\langle \mathrm{a}\colon \mathrm{ko}, \mathrm{b}\colon 7\rangle = \langle \mathrm{b}\colon 7.\mathrm{a}\colon \mathrm{ko}\rangle$.

**Ny definition af relation med skema**: En *relation* med skema $a_1\colon M_1, \ldots, a_n\colon M_n$ er en **endeling** delmængde af $a_1\colon M_1, \ldots, a_n\colon M_n$.

**NB:** Skema skal forstås som en "syntaktisk" struktur.

**Eksempel** på relation med skema $\mathrm{a}\colon \mathcal{N}, \mathrm{c}\colon \mathcal{N}, \mathrm{b}\colon \mathcal{N}$:

| a | b | c |
|---|---|---|
| 10 | 20 | 30 |
| 17 | 18 | 19 |
| 34 | 43 | 56 |
| 100 | 200 | 5000 |

# Hvad modellerer relational algebra?

Evaluering af forespørgsler til en bestemt database i et tidsrum, hvor databasen ikke ændrer sig (databasetilstand, instans af database).

Database har et *databaseskema* $R_1(S_1), \ldots, R_n(S_n)$ hvor $R_1, \ldots, R_n$ er *relationsnavne* og $R_1, \ldots, R_n$ relationsskemaer.

**Eksempel:**

Kasse(højde: $\mathcal{N}$, bredde: $\mathcal{N}$, dybde: $\mathcal{N}$, farve: {rød,grøn,blå})

Dåse(højde: $\mathcal{N}$, diameter: $\mathcal{N}$, farve: {rød,grøn,blå})

Cylinder(højde: $\mathcal{N}$, diameter: $\mathcal{N}$, farve: {rød,grøn,blå})

På givet tidspunkt har db. en *tilstand* eller *instans*, hvor hver relation(snavn) $R$ har en bestemt *instans af relationen* $IR$.

**Eksempel:**

Kasse:

| højde | bredde | dybde | farve |
|---|---|---|---|
| 110 | 25 | 35 | rød |
| 10 | 10 | 10 | blå |
| 50 | 40 | 17 | blå |

Dåse:

| højde | diameter | farve |
|---|---|---|
| 28 | 7 | grøn |
| 10 | 50 | rød |
| 1 | 30 | rød |
| 5 | 5 | rød |

Cylinder:

| højde | diameter | farve |
|---|---|---|
| 28 | 7 | blå |

**Advarsel:** Terminologi er sløset, "relation" bruges i db-litteratur for

- relationsnavnet "$R$" (blot et symbol),
- den datastruktur "$R$" refererer til, og som er noget man kan ændre i,
- om de endelige relationer (i matematisk forstand) som ligger deri



# Kom så med den algebra!!!! Vi ve' ha' algebra!!!

# Relationel algebra: En algebra *om* relationer

Antag databaseskema $R_1(S_1), \ldots, R_n(S_n)$ hvor $S$'erne er opbygget af $\mathcal{N}$, $\mathcal{Z}$, $\mathcal{B}$, $\mathcal{T}$ eller endelige delmængder $\{a_1, \ldots, a_k\} \subset \mathcal{A}$. [[genbrug af $\mathcal{A}$, f.eks. {rød,grøn,blå}]]

**Signatur:**

- Typer: alle mulige rel'skemaer samt (til betingelser) $\mathcal{N}$, $\mathcal{Z}$, $\mathcal{B}$, $\mathcal{T}$, $\mathcal{A}$.
- Konstanter $R_i$ af type $S_i$.
- For vilkårligt skema $S$ en konstant $\emptyset$ af type $S$.
- Konstanter forventet for $\mathcal{N}$, $\mathcal{Z}$, $\mathcal{B}$, $\mathcal{T}$ og $\mathcal{A}$.
- Sædvanlige op'ner for $\mathcal{N}$, $\mathcal{Z}$, $\mathcal{B}$, $\mathcal{T}$ og $\mathcal{A}$.
- For samtlige skemaer $S$, op'ner $\cup, \cap, \setminus : S, S \to S$, $\subseteq, \subset, = : S, S \to \mathcal{B}$
- ... og en mange flere operationer

**Eksempler** på relationelle udtryk (= termer over $\uparrow$ signatur) med skema:

> Kasse(højde: $\mathcal{N}$, bredde: $\mathcal{N}$, dybde: $\mathcal{N}$, farve: {rød,grøn,blå})
> Dåse(højde: $\mathcal{N}$, diameter: $\mathcal{N}$, farve: {rød,grøn,blå})
> Cylinder(højde: $\mathcal{N}$, diameter: $\mathcal{N}$, farve: {rød,grøn,blå})

- Kasse — med type højde: $\mathcal{N}$, bredde: $\mathcal{N}$, dybde: $\mathcal{N}$,
- Dåse $\cup$ Cylinder — med type højde: $\mathcal{N}$, diameter: $\mathcal{N}$, farve: {rød,grøn,blå}
- Dåse $\cup$ Cylinder $= \emptyset$ — med type $\mathcal{B}$.

**NB:** Kasse $\cap$ Dåse $= \emptyset$ er ikke et udtryk!!!!!

**Den kanoniske algebra** for $\uparrow$ signatur (den relationelle algebra):

- $\mathcal{N}$, $\mathcal{Z}$, $\mathcal{B}$, $\mathcal{T}$, $\mathcal{A}$ med operationer $\longrightarrow$ deres kanoniske betydninger
- Skema $S$ $\longrightarrow$ mgd. af alle relationer med skema $S$
- $\cup, \cap, \setminus, \subseteq, \subset, =$ $\longrightarrow$ restriktioner af de kanoniske betydninger ag tilsv. operationer i klassisk mgd.-lære.

# Type af udtryk afh. af skema, værdi af udtryk af aktuel instans

Følgende skema

Kasse(højde: $\mathcal{N}$, bredde: $\mathcal{N}$, dybde: $\mathcal{N}$, farve: {rød,grøn,blå})
Dåse(højde: $\mathcal{N}$, diameter: $\mathcal{N}$, farve: {rød,grøn,blå})
Cylinder(højde: $\mathcal{N}$, diameter: $\mathcal{N}$, farve: {rød,grøn,blå})

giver følgende typer

Kasse $\longrightarrow$ højde: $\mathcal{N}$, bredde: $\mathcal{N}$, dybde: $\mathcal{N}$,

Dåse $\cup$ Cylinder $\longrightarrow$ højde: $\mathcal{N}$, diameter: $\mathcal{N}$, farve: {rød,grøn,blå}

Dåse $\cup$ Cylinder $= \emptyset \longrightarrow \mathcal{B}$.

Følgende tilstand/instans

Kasse:

| højde | bredde | dybde | farve |
|---|---|---|---|
| 110 | 25 | 35 | rød |
| 10 | 10 | 10 | blå |
| 50 | 40 | 17 | blå |

Dåse:

| højde | diameter | farve |
|---|---|---|
| 28 | 7 | grøn |
| 10 | 50 | rød |
| 1 | 30 | rød |
| 5 | 5 | rød |

Cylinder:

| højde | diameter | farve |
|---|---|---|
| 28 | 7 | blå |

giver følgende værdier:

Kasse $\longrightarrow$

| højde | bredde | dybde | farve |
|---|---|---|---|
| 110 | 25 | 35 | rød |
| 10 | 10 | 10 | blå |
| 50 | 40 | 17 | blå |

Dåse $\cup$ Cylinder $\longrightarrow$

| højde | diameter | farve |
|---|---|---|
| 28 | 7 | grøn |
| 10 | 50 | rød |
| 1 | 30 | rød |
| 5 | 5 | rød |
| 28 | 7 | blå |

Dåse $\cup$ Cylinder $= \emptyset \longrightarrow$ falsk

**. . . og nu en masse operationer beskrevet i Ullman+Widoms bog**

# Eksempel på naturlig join

Kunder:

| Fornavn | Efternavn | Gade/vej | Nr | Postnr |
|---|---|---|---|---|
| Katrine | Petersen | Strandvejen | 1 | 8000 |
| Peter | Jensen | Skolegade | 7 | 8000 |
| Hans | Jensen | Jomfru Ane Gade | 3 | 9000 |
| Børge | Børgesen | Markvej | 1 | 5772 |
| Peter | Hansen | Snevej | 312 | 3900 |
| Petrine | Hansen | Snevej | 312 | 3900 |
| Peter | Hansen | Morbærvej | 8 | 4200 |

Postnr-by:

| Postnr | Bynavn |
|---|---|
| 1050 | København K |
| 1051 | København K |
| ... | ... |
| 2630 | Taastrup |
| ... | ... |
| 3900 | Nuuk |
| ... | ... |
| 4200 | Slagelse |
| ... | ... |
| 5772 | Kværndrup |
| ... | ... |
| 8000 | Århus C |
| ... | ... |
| 9000 | Ålborg |
| ... | ... |

Kunder $\bowtie$ Postnr-by:

| Fornavn | Efternavn | Gade/vej | Nr | Postnr | Bynavn |
|---|---|---|---|---|---|
| Katrine | Petersen | Strandvejen | 1 | 8000 | Århus C |
| Peter | Jensen | Skolegade | 7 | 8000 | Århus C |
| Hans | Jensen | Jomfru Ane Gade | 3 | 9000 | Ålborg |
| Børge | Børgesen | Markvej | 1 | 5772 | Kværndrup |
| Peter | Hansen | Snevej | 312 | 3900 | Nuuk |
| Petrine | Hansen | Snevej | 312 | 3900 | Nuuk |
| Peter | Hansen | Morbærvej | 8 | 4200 | Slagelse |

# Renaming — ombenævning af attributter

NB: Bogens notation forudsætter en "standard rækkefølge på attributter" — praktisk men usmukt!

$$\rho_{Rel'(A_1,A_2,\ldots,A_n)}(Rel)$$

*Rel* skal være relation med $n$ attributter; resultat-relationen har skema svarende til *Rel* men med nye attributter $A_1, A_2, \ldots, A_n$.
I.flg. bogen er $Rel'$ et nyt navn på den resulterende relation — *hvilket grundlæggende og algebraisk set er noget vrøvl!* — relationer er *værdier*.

Eksempelrelationer:

R:

| a | b | z |
|---|---|---|
| 1 | 22 | 777 |
| 3 | 44 | 888 |
| 50 | 40 | blå |

S:

| a | b | c |
|---|---|---|
| 1 | 22 | rød |
| 1 | 33 | blå |

Ombenævne S's b-attribut til bb:

$\rho_{\text{(a,bb,c)}}$(S):

| a | bb | c |
|---|---|---|
| 1 | 22 | rød |
| 1 | 33 | blå |

Eksempler på naturlig join før og efter ombenævning.

R ⋈ S:

| a | b | z | c |
|---|---|---|---|
| 1 | 22 | 777 | rød |

R ⋈ $\rho_{\text{(a,bb,c)}}$(S):

| a | b | z | bb | c |
|---|---|---|---|---|
| 1 | 22 | 777 | 22 | rød |
| 1 | 22 | 888 | 33 | blå |

Hvad er det sjuskeri af et "nyt navn på ny relation"; bogens notation (=gængs) et sørgeligt miskmask af sprog og metasprog, som ødelægger algebraisk forståelse — nyttig måde at skelne forskellige attributnavne i udtryk.