Gagnasafnsfræði

Páll Melsted

30. sept

Uppbrot á töflu

Ef R er tafla með dálka $\bar{A}$ þá getum við brotið R upp í töflur S og T með

1. $\bar{A} = \bar{B} \cup \bar{C}$
2. $S = \pi_B(R)$
3. $T = \pi_C(R)$

Sum uppbrot eru góð, t.d.

- {title,year,length,studionName},
- {title,year,starName}

En önnur eru slæm

- {title,year}
- {year,length,starName,studioName}

BCNF (Boyce-Codd Normal Form)

BCNF lýsir því hvernig má brjóta upp töflur til að losna við frávik (e. anomaly).

Vensl R eru á **BCNF** formi þþaa að ef

$$\bar{A} \to \bar{B}$$

gildir (og eru ófáfengileg, þ.e. $\bar{B} \subsetneq \bar{A}$)
þá er $\bar{A}$ **yfirlykill**.

BCNF Reiknirit

Ef R er á BCNF formi þá skilum við {R}

Finnum einhver BCNF frávik, $X \to Y$

- Reiknum X^+
- Setjum $R_1 = \pi_{X^+}(R)$
- Setjum $R_2 = \pi_L(R)$
- þar sem L er sammengið af X og þeim dálkum í R sem eru ekki í.

Reiknum fallákveður fyrir varpanirnar R_1 og R_2 köllum þær S_1 , S_2

Reiknum endurkvæmt BCNF fyrir og skilum sammenginu af niðurstöðunum

R : title, year, sname, pres, paddr

S : title, year $\rightarrow$ sname — BCNF? ✓

sname $\rightarrow$ pres — ✗

pres $\rightarrow$ paddr — ✗

$\{$~~sname~~ title, year$\}^+ = \{$title, year, sname, pres, paddr$\}$

title, year er yfirlykill svo

title, year $\rightarrow$ sname brýtur ekki BCNF

$\{$sname$\}^+ = \{$sname, pres, paddr$\}$

$\{$sname$\}$ er ekki yfirlykill,

brýtur BCNF

$\{$pres$\}^+ = \{$pres, paddr$\}$

brýtur BCNF

R, S ~~...~~ er ekki á BCNF formi

R : T, Y, N, P, A

S : $T, Y \rightarrow N$ — BCNF? ✓

$N \rightarrow P$ — ✗

$P \rightarrow A$ — ✗

Brjótum upp eftir $P \rightarrow A$

$X = \{P\}$, $X^+ = \{P\}^+ = \{P, A\}$

$R_1 = \pi_{\{P,A\}}(R)$

$R_2 = \pi_{\{T,N,Y,P\}}(R)$

S_1 fallákveður í R_1, finnum út frá S.

S_1 : $\{P\}^+ = \{P, A\}$ svo að

$P \rightarrow A$ $\leftarrow$ eina fallákveðan

$\{A\}^+ = \{A\}$ í S_1.

Er R_1 á BCNF ?

R

S: $T,Y \to N$

$N \to P$

$P \to A$

R_2: T, Y, N, P

S_2: $\{T\}^+ = \{T\}$

$\{Y\}^+ = \{Y\}$

$\{N\}^+ = \{N, P\}$ $\quad N \to P$

$\{P\}^+ = \{P\}$

$\{T,Y\}^+ = \{T, Y, P, N\}$

$T,Y \to P$

$T,Y \to N$ (sleppum tí $N \to P$)

BCNF? $T,Y \to N$? T,Y yvirlykil

á BCNF!

$N \to P$ brýtur BCNF

tí $\{N\}^+ = \{N, P\}$

$\{N\}$ er ikki yvirlykil.

$R_{2,1} = \pi_{\{N,P\}}(R_2)$: $S_{2,1}$: $N \to P$

$R_{2,2} = \pi_{\{T,Y,N\}}(R_2)$, $S_{2,2}$: $T,Y \to N$

Uppbrot

Með því að koma venslum yfir á BCNF þá viljum við

1. Losna við öll frávik
2. Halda sömu upplýsingum
3. Varðveita tengsl, þ.e. fallákveður

BCNF uppfyllir 1 og 2.

Sömu upplýsingar

Þegar við brjótum upp vensl þá er hægt að fá upphaflegu töfluna með náttúrulegri tengingu (natural join).

Stundum er hægt að fá meira en venslin R sem við byrjuðum með

Chase reikniritið notar uppbrotið og fallákveður til að finna hvort gögnin varðveitist nákvæmlega.

Uppbrot á BCNF leiðir ekki endilega til þess að R sé varðveitt.

R

A	B	C
1	2	3
4	2	5

$R_1 = \pi_{\{A,B\}}(R)$ $\qquad$ $R_2 = \pi_{\{B,C\}}(R)$

A	B
1	2
4	2

B	C
2	3
2	5

$R_1 \bowtie R_2$

A	B	C
1	2	3
1	2	5
4	2	3
4	2	5

var ekki í R

$R \subseteq R_1 \bowtie R_2$

Chase reiknirit

Inntak
- R vensl
- $S_1, \dots, S_k$ dálkamengi
- Fallákveður

$R(A,B,C,D,E)$

$S_1 = \{A,B,C\}$

$S_2 = \{B,C,D\}$

$S_3 = \{A,C,E\}$

$A \to D$

$C,D \to E$

$E \to D$

er $R = R_1 \bowtie R_2 \bowtie R_3$?

a) $R \subseteq R_1 \bowtie R_2 \bowtie R_3$

b) Látum $t = (a,b,c,d,e) \in R$

A	B	C	D	E
a	b	c	d_1	e_1
a_2	b	c	d	e_2
a	b_3	c	d_3	e

ein röð f. hverja vörpun

Lágvísar merkja dálka sem eru utan vörpunar.

A	B	C	D	E
a	b	c	d_1	e_1
a_2	b	c	d	e_2
a	b_3	c	d_1	e

Skoðum $A \to D$ verður $d_1 = d_3$

1. og 3. röð eru með eins C, D og $C, D \to E$ svo að $e_1 = e$

fellum kennivísinn burt

A	B	C	D	E
a	b	c	d_1	e
a_2	b	c	d	e_2
a	b_3	c	d_1	e

$A \to D$

$C, D \to E$

$E \to D$

Ef við finnum röð án kennivísa þá er $R = R_1 \bowtie R_2 \bowtie R_3$ annars ekki, ~~kd~~ og taflan er mótdæmið.

A	B	C	D	E
a	b	c	d_1	e
a_2	b	c	d	e_2
a	b_3	c	d_1	e

A	B	C
a	b	c
a_2	b	c
a	b_3	c

B	C	D
b	c	d_1
b	c	d
b_3	c	d_1

A	C	E
a	c	e
a_2	c	e_2

$R_1 \bowtie R_2$

A	B	C	D
a	b	c	d_1
a_2	b	c	d_1
a	b	c	d

var ekki í R!

R {title, Theater, city}

Theater $\rightarrow$ city

title, city $\rightarrow$ theater

{title, city} er lykill

{title, theater} er lykill

Brjótum upp eftir theater $\rightarrow$ city

$\{\text{theater}, \cancel{\text{city}}\}^+ = \{\text{theater}, \text{city}\}$

R_1

theater	city
A	B
A'	B

theater $\rightarrow$ city

R_2

theater	title
A	C
A'	C

$R_1 \bowtie R_2$

theater	city ~~theater~~	title ~~city~~
A	B	C
A'	B	C

title, city $\rightarrow$ theater er brotið.

3-NF

Vensl R eru á 3-NF þþaa ef

$$\bar{A} \to \bar{B}$$

gildir þá gildir annaðhvort

1. $\bar{A}$ er yfirlykill

2. það B_i sem er ekki í $\bar{A}$ er í einhverjum lykli (ekki endilega sama lykli fyrir öll B_i)

3-NF Reiknirit

Inntak: vensl R, fallákveður F

1. Finnum lággrunn G fyrir F

2. Fyrir öll $\bar{A} \to B$ í G, notum $\bar{A} \cup \{B\}$ sem dálka fyrir ný vensl

3. Ef engin af dálkunum í skrefi #2 mynda yfirlykil, þá bætum við við venslum með lykil sem dálka.

3. (15 stig) Eftirfarandi vensl $R(A, B, C, D)$ hafa fallákveður S:

- $C \to B$
- $BC \to A$
- $A \to C$
- $BD \to A$

Finnið alla lykla í R. Teljið upp þær fallákveður í S sem brjóta BCNF skilyrði. Brjótið upp venslin í vensl safn af venslum sem öll eru á BCNF formi. Sýnið að uppbrotið ykkar varðveitir tengsl með chase reikniritinu.

$R(A, B, C, D)$

$C \to B$, $BC \to A$, $A \to C$, $BD \to A$

D kemur aldrei hægra megin, allir lyklar innihalda D.

$\{D\}^+ = \{D\}$, ekki yfirlykill

$\{D, B\}^+ = \{D, B, A, C\}$ er yfirlykill

$\{D, A\}^+ = \{D, A, C, B\}$ —"—

$\{D, C\}^+ = \{D, C, B, A\}$ —"—

ef við tökum einn dálk úr því er þetta ekki yfirlyklar svo AD, BD, CD eru lyklar

$C \to B$ brýtur BCNF því C er ekki yfirlykill

$BC \to A$ —"— BC —"—

$A \to C$ —"— A —"—

$BD \to A$ er í lagi BD yfirlykill.

$C \to B$, $A \to C$, $BC \to A$, $BD \to A$

brjótum upp eftir

$\{A\}^+ = \{A, C, B\}$

$R_1 = \pi_{\{A, B, C\}}(R)$

$R_2 = \pi_{\{A, D\}}$ ← 2 dálkar = BCNF

S_1: $C \to B$, $A \to C$, $BC \to A$

S_2: $\cancel{A \to D}$
$\cancel{D \to A}$

R_1, S_1 brýtur ekki BCNF.

$R_{S_1} = \{A, B, C\}$

$R_{S_2} = \{A, D\}$

A	B	C	D
a	b	c	d_1
a	b_2	c_2	d

$A \to C$ getur $c = c_2$

svo $C \to B$ getur $b = b_2$

svo það varðveitir uppbrot.