

TDT4145 øving4

Aleksander Hansen

16. April 2015

Oppg. 1.

- a) En nøkkel til en relasjon, R , er en mengde attributter $K \subseteq R$ med egenskapene:

1. For to tuppler t_1 og t_2 i R , så er $t_1[K] \neq t_2[K]$
2. $\neg \exists K' \subset K$ med egenskap 1.

Med andre ord, nøkkelen K må være unik og minimal.

En supernøkkel er en nøkkel som mangler egenskap 2.

En funksjonell avhengighet $X \rightarrow Y$ er en føring mellom to mengder attributter, X og Y , hvor alle tuppelpar (t_1, t_2) som har samme verdi for attributtene i X må ha samme verdi for attributtene i Y . i.e $X[t_1] = X[t_2] \Rightarrow Y[t_1] = Y[t_2]$.

- b) Tillukningen X^+ til X er mengden attributter i X samt mengden av attributter som kan deriveres fra de funksjonelle avhengighetene.

Gitt at F inneholder alle funksjonelle avhengigheter, så kan algoritmen beskrives slik (med pytonsk psudokode):

```
X+ = X
oldX+ = null
while oldX+ != X+:
    oldX+ = X+
    for Y → Z i F:
        if Y ⊆ X+:
            X+ = X+ ∪ Z
```

- c) for $F = \{a \rightarrow e, ac \rightarrow d, b \rightarrow c\}$

```
a+ = ae
ab+ = acde
e+ = e
```

- d) En mengde attributter er en supernøkkel for en tabell når tillukningen til attributtene i mengden inneholder alle attributtene i tabellen. Vi kan derfor undersøke om en mengde attributter er en supernøkkel ved å beregne tillukningen og sjekke om den inneholder alle attributtene i tabellen.
- Hvis supernøkkel er minimal er den også en nøkkel. Ved å sjekke om supernøkkel fortsatt er minimal etter vi har fjernet hvert attributt en etter en kan vi finne ut om supernøkkel også er en nøkkel.
- e) Hvis $((R_1 \cap R_2) \rightarrow (R_1 - R_2)) \in F^+$ eller $((R_1 \cap R_2) \rightarrow (R_2 - R_1)) \in F^+$ så har dekomponeringen av en tabell R i to projeksjoner R_1 og R_2 en tapsløs-join-egenskap.
- f)
1. $R_1 \cap R_2 = \{b, c\}$, $R_1 - R_2 = \{a\}$ og $R_2 - R_1 = \{d\}$. Verken $((R_1 \cap R_2) \rightarrow (R_1 - R_2))$ eller $((R_1 \cap R_2) \rightarrow (R_2 - R_1))$ er med i F^+ og dekomponeringen er dermed ikke tapsløs-join.
 2. $R_1 \cap R_2 = \{b, d\}$, $R_1 - R_2 = \{a\}$ og $R_2 - R_1 = \{c\}$. Siden $((R_1 \cap R_2) \rightarrow (R_2 - R_1)) \in F^+$ så er dekomponeringen tapsløs-join.
 3. $R_1 \cap R_2 = \{b\}$, $R_1 - R_2 = \{a, d\}$ og $R_2 - R_1 = \{c\}$. Siden $((R_1 \cap R_2) \rightarrow (R_2 - R_1)) \in F^+$ så er dekomponeringen tapsløs-join.
- g) Fra 15.4.2: A relational schema R is in **third normal form (3NF)** if, whenever a *nontrivial* functional dependency $X \rightarrow A$ holds in R , either (a) X is a superkey of R , or (b) A is a prime attribute of R .
- h) Vi sjekker om dekomponeringen er en god løsning ved å sjekke:
- Attributtbevaring: Siden alle attributtene er med i minst en komponenttabell har vi attributtbevaring.
 - Bevaring av funksjonelle avhengigheter: De funksjonelle avhengighetene som gjelder i R kan utledes fra avhengighetene i komponenttabellene.
 - Tapsløs-join-egenskap: A er en supernøkkel i R_1 , R_1 og R_3 kan da joines tapsløst pga. felles attributt A . C er en supernøkkel i R_2 og pga. felles attributt, C , med joinet over så kan vi joine R_2 tapsløst med tidligere join. Dekomponeringen har altså en tapsløs-join-egenskap.
 - Normalform: De funksjonelle avhengighetene til komponenttabellene har en supernøkkel som venstreside attributt, og er dermed på BCNF. Dekomponeringen er altså en god løsning.

Oppg. 2

Postene vil hashes til følgende blokker:

2369	1
3760	0
4692	4
4871	7
5659	3
1821	5
1074	2
7115	3
1620	4
2428	4 overflow
3943	7
4750	6
6975	7 overflow
4981	5
9208	0

2 av 15 poster er i overflow og vil trenger en ekstra blokkaksess. De andre trenger bare en. Gjennomsnittlig antall blokkaksesser for en tilfeldig aksess er dermed:

$$\frac{13}{15} + 2 \times \frac{2}{15} = \underline{1.133}$$

Oppg. 3

2369	65
3760	48
4692	84
4871	7
5659	27
1821	29
1074	50
7115	75
1620	84
2428	124
3943	103
4750	14
6975	63
4981	117
9208	120

Oppg. 4

Oppg. 5