

LØSNINGSSKISSE TIL EKSAMENSOPPGAVE I FAG TDT4300 – JUNI 2017

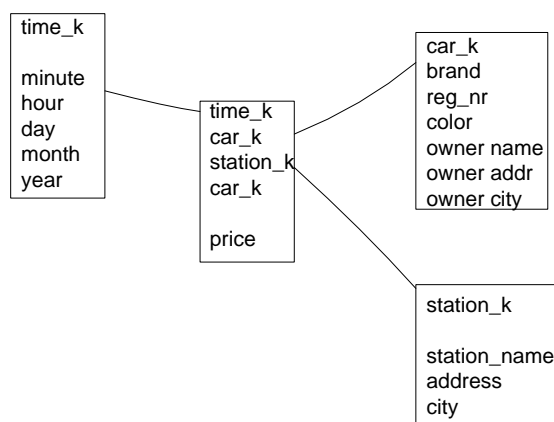
NB! Dette er ikkje fullstendige løysingar på oppgåvene, kun skisse med viktige element, hovudsakleg laga for at vi skal ha oversikt over arbeidsmengda på eksamen, og som huskeliste under sensurering. Det er også viktig å være klar over at det også kan vere andre svar enn dei som er gjeve i skissa som vert rekna som korrekt om ein har god grunngjeving eller dei gjev meining utifrå kva det vert spurd om.

Oppgåve 1

- a) Sidevisning: Samla framstilling av eit sett med Web-objekt som bidrar til ei visning i en nettlesar som følgje av ei enkelt brukarhandling (for eksempel ein "clickthrough")
Konseptuelt: Kvar sidevisning kan verte sett på som ein samling av Web-objekt eller ressursar som representerer ein bestemt "brukarhending"
T.d.: Lese en artikkel, vise ei produktside eller å legge til eit produkt i handlevogna
- b) Klient- eller proxy-caching kan medføre at ein ikkje har alle referansar i loggen. Kan gjerast vha. referent-informasjon i logg i kombinasjon med kunnskap om lenke-struktur.

Oppgåve 2

a) Eit viktig poeng er at fakta-attributt må gje meining utifrå oppgåva, og også gje meining ved aggregering. Med tanke på at oppgåve er relativt triviell er vi tilsvarande strenge, t.d. trekk ved manglande dimensjonstabell, om fakta (pris) er i ein av dimensjonstabellane men ikkje i det heile i faktatabell. Legg merke til at dette skjemaet har meir informasjon enn strengt tatt nødvendig utifrå oppgåveteksta, for car_k er det også OK om dei andre attributtane ikkje er med. Eit viktig poeng er at ein må ha pris og ikkje berre antal i faktatabell. Kan også ha kunde som separat dimensjonstabell.



Oppgåve 3

- a) Roll-up: Roll-up betyr å bevege seg oppover et konsept-hierarki, eller foreta reduksjon av dimensjoner. I praksis aggregering av data, vi går fra detaljerte data til mindre detaljerte.
Drill-down er det motsatte av roll-up: Går et trinn ned i konsepthierarkiet eller inkluderer en dimensjon til. Går fra overordnede data til mer spesifikke data.
- b) Binding:
Hardcover: 1100110
Paperback: 0011001
Language:
English: 1010011
French: 0101100
Et viktig poeng er at dette er indekser, dvs. ikke en del av "data-tabellene".

Oppgave 4

13				P3					
12									
11					I3		P6		
10				P2					
9									
8				P1	I2				
7					P5				
6									
5									
4				I1					
3					P4				
2									
1									
0	1	2	3	4	5	6	7	8	

- 1) $C1=P4, C2=P1, P5, C3=P2, P3, P6$
 2) $S1=(5,3), S2=(4,5,7.5), S3=(5,11.33)$
 3) $C1=P4, C2=P1, P5, C3=P2, P3, P6$
 Ingen endring, dvs. terminert!

Oppgave 5

- a) $A=C3$ (3 nærmeste er $P4/P9/P11$ som hører til hhv. $C1/C3/C3$, flertall for $C3$) evt. $C1$ eller $C2$ om 3 nærmeste er $P4/P9/P6$
 $B=C1, C2$, eller $C3$ (3 nærmeste er $P1/P6/P13$ som hører til hhv. $C1/C2/C3$, ikke flertall for noen). Også korrekt om vektning på distanse er brukt, da $C1$ eller $C2$.
 $C=C3$ (3 nærmeste er $P6/P12/P13$, hhv. $C2/C3/C3$).
 $D=C2$ (3 nærmeste er $P6/P7/P8$, hhv. $C2/C2/C2$).

b)

Gini i rotnode:

$$p(J|Parent) = 4/15 = 0.26667, p(N|Parent) = 11/15 = 0.73333$$

$$GI(J, N) = 1 - 0.26667 * 0.26667 - 0.73333 * 0.73333 = 0.39111$$

1) **Splitting på A:**

$S1="L"$

$$J1=1, N1=6, GI(J1, N1) = GI(1, 6) = 1 - 1/7 * 1/7 - 6/7 * 6/7 = 0.244898$$

$S2="H"$

$$J2=3, N2=5, GI(J2, N2) = GI(3, 5) = 1 - 3/8 * 3/8 - 5/8 * 5/8 = 0.46875$$

$$GAIN(A1) = 0.39111 - 7/15 * 0.244898 - 8/15 * 0.46875 = 0.0268242$$

2) **Splitting på B**

$S1="T"$

$$J1=3, N1=1, GI() = 1 - 3/4 * 3/4 - 1/4 * 1/4 = 0.3750$$

$S2="K"$

$$J2=0, N2=3, GI() = 1 - 3/3 * 3/3 = 0$$

$S3="F"$

$$J3=1, N3=7, GI() = 1 - 1/8 * 1/8 - 7/8 * 7/8 = 0.218750$$

$$GAIN(A2) = 0.39111 - 4/15 * 0.3750 - 3/15 * 0 - 8/15 * 0.218750 = .17444$$

Vi vel attributtet med høyeste GAIN, dvs. **B** vert føretrakt for første splitting av treet.

NB! Viktig å ha med GAIN inkl. $p(J|Parent)$, det kan skje at ein får negativ verdi for begge dei alternative splittingane, som betyr at ein ikkje bør velje nokon av dei.

Oppgave 6

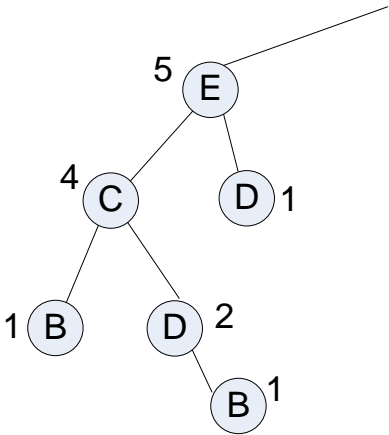
a)

A	3
B	4
C	4
D	6
E	4
F	4
G	5
H	2
BC	2
BD	4
BE	4
BF	0
BG	4
CD	3
CE	2
CF	2
CG	2
DE	4
DF	2
DG	4
EF	0
EG	4
FG	1
BDE	4
BDG	4
BEG	4
DEG	4
BDEG	4

Kun eit 4-elementsett mogleg: BDEG | 4

BDEG	4
------	---

b) Støttetal: A:1, B:2, C:4, D:3, E:5



tid	Itemset	(Ordered) frequent items
T	ACE	EC
T2	BCE	ECB
T3	BCDE	ECDB

T4	CDE	ECD
T5	DE	ED

Item	Conditional sub-database	Conditional FP-tree	Frequent itemsets
B	$\{(EC:1), (ECD:1)\}$	$EC:2$	B, BC, BE, BCE
D	$\{(EC:2), (E:1)\}$	$E3C2$	D, DE, DC, DEC
C	$\{(E:4)\}$	$E4$	C, CE
E	\emptyset	\emptyset	E

(OK om 1-elementsett ikkje er med i tabellen).