Gagnasafnsfræði

Fallákveður og staðalskipulag

Hallgrímur H. Gunnarsson

Hönnun á gagnagrunni

Grf. að við höfum vensl fyrir starfsmenn: R(kt,nafn,launaflokkur,taxti)

Grf. að launaflokkur ákvarði launataxta: launaflokkur $\rightarrow$ launataxti

kt	nafn	launaflokkur	taxti
2004862139	Bjarni Rúnar	A	3000
1210872899	Davíð Fannar	A	3000
0702882989	Einar Skúlason	B	2300
0101852459	Guðmundur Jónsson	B	2300

Er þetta góð hönnun?

Hvaða vandamál geta komið upp?

Þrjár gerðir vandamála (1/3)

Uppfærsluvandamál (e. update anomaly):

Ef taxti breytist þá þarf að breyta upplýsingum á mörgum stöðum

Dæmi:

kt	nafn	launaflokkur	taxti
2004862139	Bjarni Rúnar	A	3000
1210872899	Davíð Fannar	A	3000
0702882989	Einar Skúlason	B	2300
0101852459	Guðmundur Jónsson	B	2300

Þrjár gerðir vandamála (2/3)

Innsetningarvandamál (e. insert anomaly):

Ef enginn starfsmaður er í launaflokki X þá er ekki hægt að setja þær upplýsingar inn í kerfið

Dæmi:

kt	nafn	launaflokkur	taxti
2004862139	Bjarni Rúnar	A	3000
1210872899	Davíð Fannar	A	3000
0702882989	Einar Skúlason	B	2300
0101852459	Guðmundur Jónsson	B	2300
?	?	C	1800

Þrjár gerðir vandamála (3/3)

Eyðingarvandamál (e. delete anomaly)

Ef síðasta starfsmanni í launaflokki X er eytt út þá glatast upplýsingar um þann launaflokk

Dæmi:

kt	nafn	launaflokkur	taxti
2004862139	Bjarni Rúnar	A	3000
1210872899	Davíð Fannar	A	3000
~~0702882989~~	~~Einar Skúlason~~	~~B~~	~~2300~~
~~0101852459~~	~~Guðmundur Jónsson~~	~~B~~	~~2300~~

Fallákveða (e. functional dependency)

Skilgreining:

Fallákveða $X \rightarrow Y$, þar sem X og Y eru mengi dálka, gildir fyrir vensl R ef allar raðir með jöfn gildi í dálkum X hafa jöfn gildi í Y

Skilgreining á ensku:

$X \rightarrow Y$ means that whenever two tuples in R agree on all the attributes of X, they must also agree on all attributes of Y

Háð: Segjum að Y sé háð X ef $X \rightarrow Y$

Dæmi: Kennitala ákvarðar nafn (en nafn ákvarðar ekki kennitölu)

Frumsetningar um falllákveður

Sjálfhverfni (e. reflexivity)

Ef $Y \subseteq X$ þá $X \to Y$

Aukning (e. augmentation):

Ef $X \to Y$ þá $XZ \to YZ$ fyrir eitthvert Z

Gegnvirkni (e. transitivity):

Ef $X \to Y$ og $Y \to Z$ þá $X \to Z$

Frumsetning: Sjálfhverfni

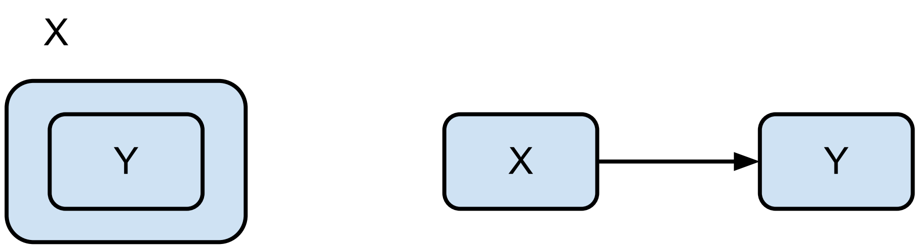
Sjálfhverfni:

Ef $Y \subseteq X$ þá $X \rightarrow Y$

Dæmi:

1. $A \rightarrow A$
2. $AB \rightarrow AB, \quad AB \rightarrow A, \quad AB \rightarrow B$

Myndrænt:



Frumsetning: Aukning

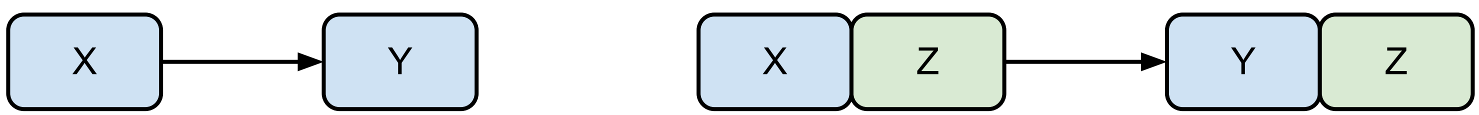
Aukning:

Ef $X \to Y$ þá $XZ \to YZ$ fyrir eitthvert Z

Dæmi:

$\{$ Kennitala $\} \to \{$ Nafn $\} \Rightarrow \{$ Kennitala, Kyn $\} \to \{$ Nafn, Kyn $\}$

Myndrænt:



Frumsetning: Gegnvirkni

Gegnvirkni:

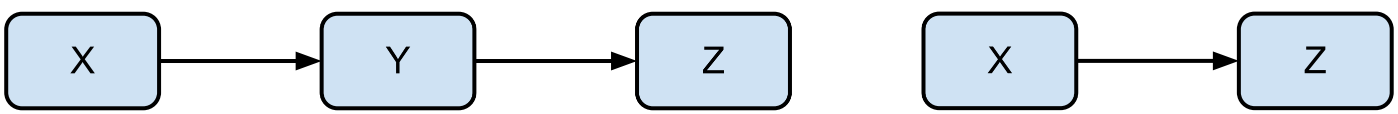
Ef $X \rightarrow Y$ og $Y \rightarrow Z$ þá $X \rightarrow Z$

Dæmi:

Starfsmaður $\rightarrow$ Launaflokkur og Launaflokkur $\rightarrow$ Launataxti

$\Rightarrow$ Starfsmaður $\rightarrow$ Launataxti

Myndrænt:



Reglur um fallákveður

Sammengi (e. union):

Ef $X \to Y$ og $X \to Z$ þá $X \to YZ$

Uppbrot (e. decomposition):

Ef $X \to YZ$ þá $X \to Y$ og $X \to Z$

Aukin gegnvirkni (e. pseudo-transitivity):

Ef $X \to Y$ og $YZ \to A$ þá gildir einnig $XZ \to A$

Lokun fyrir mengi dálka

Skilgreining:

Fyrir mengi dálka Z í R og mengi fallákveða F sem gilda fyrir R þá er *lokun* Z *m.t.t.* F, ritað Z^+, mengi allra dálka sem eru ákvarðanlegir af Z

Dæmi:

$R(A, B, C, D, E, F)$ og $F = \{ A \to B,\ B \to C,\ C \to D,\ E \to F \}$

$\{ C \}^+ = \{ CD \}$

$\{ B \}^+ = \{ BCD \}$

$\{ A \}^+ = \{ ABCD \}$

$\{ E \}^+ = \{ EF \} \qquad \{ CE \}^+ = \{ CDEF \}$

Reiknirit

Reiknirit fyrir lokun dálkamengis:

1. Byrjum með Z
2. Ef $X \to Y$ og X er í lokuninni, þá bætum við Y í lokunina
3. Endurtökum þar til engu er hægt að bæta við

Dæmi:

$R(A, B, C, D, E, F)$ og $F = \{ A \to B,\ B \to C,\ C \to D,\ E \to F \}$

Reiknum $\{ B \}^+$. Byrjum með $S = \{ B \}$.

1. $B \in S$ og $B \to C$. Bætum við C. Þá er $S = \{ B,\ C \}$
2. $C \in S$ og $C \to D$. Bætum við D. Þá er $S = \{ B,\ C,\ D \}$
3. Engar fleiri reglur. Niðurstaða: $\{ B \}^+ = \{ B,\ C,\ D \}$

Lokun fyrir mengi fallákveða

Skilgreining:

Fallákveða f leiðir af safni fallákveða F ef f gildir þegar allar fallákveður í F gilda. Lokun F, ritað F^+, er mengi allra fallákveða sem leiða af F

Dæmi:

$R(A, B, C)$ og $F = \{ A \to B,\ B \to C \}$

$$
\begin{aligned}
F^+ &= \{ A \to B,\ B \to C \} & \\
&\cup \{ A \to C \} & \text{Gegnvirkni} \\
&\cup \{ A \to BC \} & \text{Sammengi} \\
&\cup \ldots &
\end{aligned}
$$

Dæmi

Látum: $F = \{ A \to B, B \to C, CD \to E \}$

Spurning: Er $A \to E$ í F^+ ?

Svar: Finnum A^+, mengi þeirra dálka sem A ákvarðar samkvæmt F^+

1. Reglan um sjálfhverfni gefur okkur $A \in A^+$

2. Höfum $B \in A^+$ því $A \to B$

3. Gegnvirkni: Höfum $C \in A^+$ því $A \to B$ og $B \to C$

Niðurstaða: $A^+ = \{ A, B, C \}$. $A \to E$ leiðir því ekki af F.

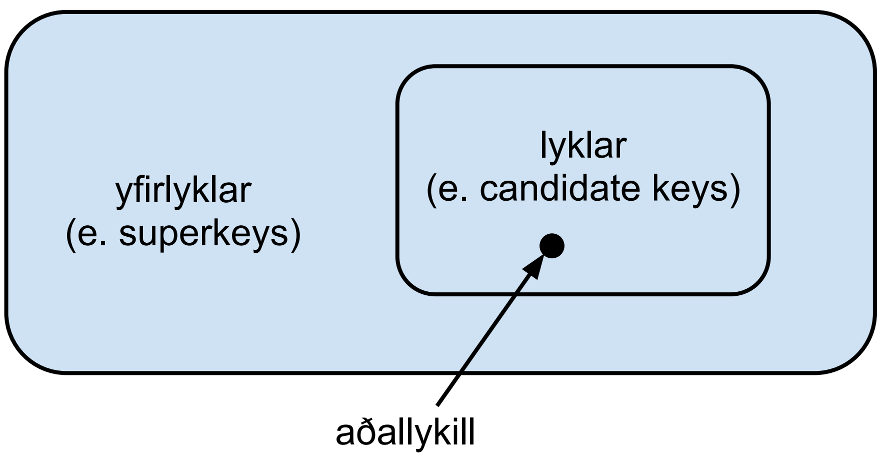
Lyklar

Yfirlykill: Mengi dálka K þar sem $K \to attrs(R)$

önnur skilgr.: Mengi dálka K þar sem $K^+ = attrs(R)$

Lykill: Óminnkanlegur yfirlykill (ekkert hlutmengi í K er yfirlykill)

Tegundir:



Að finna lykla

Gagnlegur eiginleiki:

Ef A er aldrei hægra megin í fallákveðunum þá innihalda allir lyklar A

Ábending:

Prófa lítil mengi fyrst (byrja með staka dálka), svo skoða stærri mengi

Ef $X^+ = attrs(R)$ þá þarf ekki að skoða nein stærri mengi.

Dæmi: $F = \{\ A \rightarrow B$, $B \rightarrow C$, $CD \rightarrow E\ \}$

Spurning: Hver er lykillinn fyrir F ?

Svar: AD

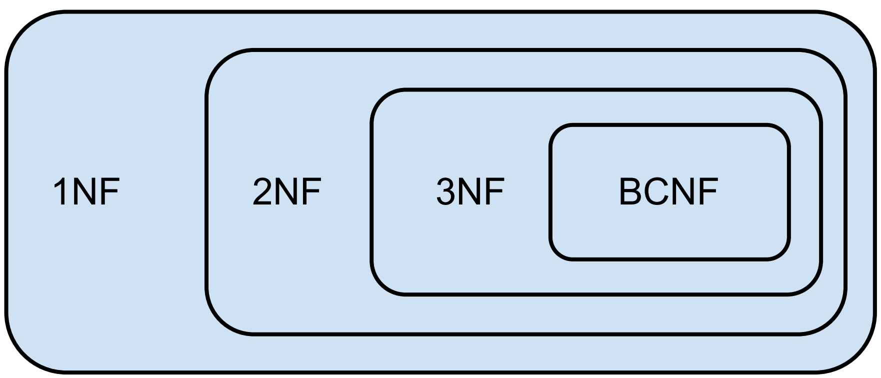
Staðalskipulög

Staðalskipulag (e. normal form)

Staðalskipulag kemur í veg fyrir að ákveðin töfluvandræði geti komið upp.

Til margar tegundir, fyrsta skipulag (1NF), annað skipulag (2NF), o.s.frv. 1NF er veikast, en BCNF er sterkast (gerir ströngustu kröfurnar).

Stigveldi:



Skipulag: 1NF

1NF (e. first normal form)

Sérhver dálkur inniheldur ósamsett (e. atomic) gildi

Dæmi um töflu sem er ekki á 1NF:

kt	nafn	símanúmer
2004862139	Bjarni Rúnar	555-1000
1210872899	Davíð Fannar	555-2000
0702882989	Einar Skúlason	555-3000 555-3001
0101852459	Guðmundur Jónsson	555-4000

Skipulag: 1NF

1NF (e. first normal form)

Sérhver dálkur inniheldur ósamsett (e. atomic) gildi

Dæmi um töflu sem er á 1NF:

kt	nafn	símanúmer
2004862139	Bjarni Rúnar	555-1000
1210872899	Davíð Fannar	555-2000
0702882989	Einar Skúlason	555-3000
0702882989	Einar Skúlason	555-3001
0101852459	Guðmundur Jónsson	555-4000

Skipulag: 2NF

2NF (e. second normal form)

1NF + Enginn dálkur er háður hluta úr lykli (hlutákveða, e. partial dependency).
Til dæmis: Ef við höfum R(A,B,C) og AB er lykill, þá er $B \rightarrow C$ hlutákveða.

Dæmi um töflu sem er ekki á 2NF:

dags	myntnafn	myntheiti	midgengi
2011-09-01	USD	Bandaríkjadalur	117.38
2011-09-01	GBP	Sterlingspund	183.17

Lykillinn er $\{$ dags, myntnafn $\}$, höfum einnig falláákveðuna myntnafn $\rightarrow$ myntheiti

Samsettir lyklar: Ef 1NF tafla hefur enga samsetta lykla þá er hún á 2NF

Skipulag: 2NF

2NF (e. second normal form)

1NF + Enginn dálkur er háður hluta úr lykli (hlutákveða, e. partial dependency).
Til dæmis: Ef við höfum R(A,B,C) og AB er lykill, þá er $B \to C$ hlutákveða.

Taflan lagfærð með uppbroti:

dags	myntnafn	midgengi
2011-09-01	USD	117.38
2011-09-01	GBP	183.17

myntnafn	myntheiti
USD	Bandaríkjadalur
GBP	Sterlingspund

Lykillinn í vinstri töflunni er $\{$ dags, myntnafn $\}$, hægri töflunni er myntnafn

Lykillinn í hægri töflunni er myntnafn

Skipulag: 3NF

3NF (e. third normal form)

2NF + Allir dálkar sem eru ekki hluti af lykli mega bara vera háðir lykli

Önnur skilgreining:

R er á þriðja skipulagi ef fyrir öll $X \to A$ í F^+ gildir amk. eitt af eftirfarandi:

1. $A \in X$, þ.e. X ákvarðar eitthvað í X (trivial FD)
2. X er yfirlykill
3. A er hluti af amk. einum lykli

Brot gegn 3NF:

$X \to A$ brýtur gegn 3NF ef X inniheldur ekki neinn lykil og A er ekki hluti af neinum lykli (og $A \notin X$)

Skipulag: BCNF

BCNF (e. Boyce-Codd normal form)

3NF $+$ Allir dálkar mega bara vera háðir lykli, engu öðru

Önnur skilgreining:

R er á BCNF skipulagi ef fyrir öll $X \to A$ í F^+ gildir amk. eitt af eftirfarandi:

1. $A \in X$, þ.e. X ákvarðar eitthvað í X (trivial FD)
2. X er yfirlykill

Brot gegn BCNF:

$X \to A$ brýtur gegn BCNF ef X inniheldur ekki neinn lykil

Auðvelt að muna

3NF:

"Each non-key attribute must provide a fact about *the key (1NF)*, *the whole key (2NF)* and *nothing but the key (3NF)*"

BCNF:

"Each attribute must provide a fact about *the key (1NF)*, *the whole key (2NF)* and *nothing but the key (3NF)*"

Uppbrot

Uppbrot: (e. decomposition)

Tökum töflu sem er ekki á 3NF/BCNF og brjótum hana upp í smærri töflur sem hver um sig er á 3NF/BCNF.

Ef $X \to Y$ brýtur gegn 3NF/BCNF þá brjótum við R upp í tvö vensl:

Eitt með dálka XY, en hitt með dálka $R - Y$

Taplaust uppbrot: (e. lossless decomposition)

Uppbrot á R í X og Y er taplaust þ.þ.a.a. F^+ innihaldi $X \cap Y \to X$ eða $X \cap Y \to Y$. Þ.e. sameiginlegu dálkarnir $(X \cap Y)$ eru lykill fyrir aðra hvora töfluna

Gagnlegt: Ef $X \to Y$ og $X \cap Y = \emptyset$ þá er uppbrot á R í XY og $R - Y$ taplaust

Dæmi 1

Gefin eru venslin $R(A, B, C)$ með fallákveðunum $A \to B$, $B \to C$

Lyklar: Finnum alla lykla í R

Notfærum okkur að dálkar sem koma aldrei fyrir hægra megin hljóta að vera í lykli. Sjáum að A hlýtur að vera í lykli. En er A lykill? Þeas. getum við ákvarðað alla aðra dálka í R út frá A ? Já, $A \to BC$. A er jafnframt eini lykillinn fyrir R.

Ekki 3NF: Venslin R eru ekki á 3NF, útskýrið hvers vegna

$B \to C$ brýtur gegn 3NF (uppfyllir ekki neitt skilyrði).

$C \notin B$, B er ekki yfirlykill og C er ekki hluti af lykli

Dæmi 2

Gefin eru venslin $R(A, B, C, D)$ með falllákveðunum $C \to D$, $C \to A$, $B \to C$

Lyklar: Finnum alla lykla í R

B kemur aldrei fyrir hægra megin. B getur ákvarðað alla dálkana, $B \to ADC$. B er lykill.

BCNF: Eru venslin R á BCNF?

Nei, $C \to D$ brýtur gegn BCNF (C er ekki yfirlykill)

3NF: Eru venslin R á 3NF?

Nei, $C \to D$ brýtur einnig gegn 3NF

(C er ekki yfirlykill og D er ekki hluti af lykli)

Dæmi 3

$R(A,\ B,\ C,\ D)$ og $F = \{\ A \to B,\ BC \to D,\ D \to BC,\ C \to A\ \}$

Spurning: Finnið alla lykla í R

Svar: ?

Dæmi 3

$R(A,\, B,\, C,\, D)$ og $F = \{\, A \to B,\; BC \to D,\; D \to BC,\; C \to A\,\}$

Spurning: Finnið alla lykla í R

Svar:

D er lykill, $D \to BC$, $C \to A$

C er einnig lykill, $C \to A$, $A \to B$, $BC \to D$

Dæmi 4

Gefin eru venslin $R(A,\ B,\ C)$ með fallákveðunum $AB \to C$ og $B \to C$.

Spurning:

Er hægt að álykta út frá þessu að $A \to C$ gildi líka?

Með öðrum orðum: er $A \to C$ í F^+ ?

Svar: ?

Dæmi 4

Gefin eru venslin $R(A,\ B,\ C)$ með fallákveðunum $AB \to C$ og $B \to C$.

Spurning: Er hægt að álykta út frá þessu að $A \to C$ gildi líka? Með öðrum orðum, er $A \to C$ í F^+ ?

Svar:

$\{\ A\ \}^+ = \{\ A\ \}$. Engu fleiru hægt að bæta við.

Svarið er því nei.

Mótdæmi:

A	B	C
1	2	3
1	4	5