

TÖL303G

Gagnasafnsfræði

Snorri Agnarsson

Vinstri ytri tengingar (left outer join)

- Ef R og S eru vensl með einhver sameiginleg eigindi þá er vinstri ytri tenging R og S (left outer join):

$$R \bowtie S = (R \bowtie S) \cup \left(\left(R - \pi_{r_1, r_2, \dots, r_n}(R \bowtie S) \right) \times \{(\perp, \dots, \perp)\} \right)$$

- þar sem r_1, r_2, \dots, r_n eru eigindi R sem ekki koma fyrir í S og mengið $\{(\perp, \dots, \perp)\}$ er eins staks vensl yfir þau eigindi R sem ekki koma fyrir í S
- Áhrifin eru þau að $R \bowtie S$ inniheldur a.m.k. eina röð fyrir sérhverja röð í R
- Ef röð vantar í $R \bowtie S$ fyrir einhverja tiltekna röð R þá mun $R \bowtie S$ samt innihalda röð sem samsvarar þeirri röð, en í þeirri röð verða eiginleikarnir sem eðlilega kæmu frá S með gildi \perp (samsvarar NULL)

Hægri ytri tengingar (right outer join)

- Ef R og S eru vensl með einhver sameiginleg eigindi þá er hægri ytri tenging R og S (right outer join):

$$R \bowtie S = (R \bowtie S) \cup \left(\{(\perp, \dots, \perp)\} \times \left(S - \pi_{s_1, s_2, \dots, s_m}(R \bowtie S) \right) \right)$$

- þar sem s_1, s_2, \dots, s_m eru eigindi S sem ekki koma fyrir í R og mengið $\{(\perp, \dots, \perp)\}$ er eins staks vensl yfir þau eigindi S sem ekki koma fyrir í R
- Áhrifin eru þau að $R \bowtie S$ inniheldur a.m.k. eina röð fyrir sérhverja röð í S
- Ef röð vantar í $R \bowtie S$ fyrir einhverja tiltekna röð S þá mun $R \bowtie S$ samt innihalda röð sem samsvarar þeirri röð, en í þeirri röð verða eiginleikarnir sem eðlilega kæmu frá R með gildi \perp (samsvarar NULL)

Fullar ytri tengingar (full outer join)

- Ef R og S eru vensl með einhver sameiginleg eigindi þá er full ytri tenging R og S (full outer join):

$$R \bowtie S = (R \bowtie S) \cup (R \bowtie S)$$

- Í Ullman og Widom eru notuð önnur tákn í stað \bowtie , \bowtie og \bowtie
- Þið megið nota þau tákn ef ykkur hentar, en þau eru svona:

\bowtie samsvarar \bowtie_L \bowtie samsvarar \bowtie_R \bowtie samsvarar \bowtie

- Setja má skilyrði með ytri tengingum á svipaðan hátt og með innri tengingum (\bowtie , θ -tengingar)
 - þá er ekki nauðsynlegt að R og S hafi sameiginleg eigindi

Hönnun gagnagrunna – Database Design

- Tengsl milli gagna ákvarða hvernig er „best“ að hanna gagnagrunna
Data relationships decide the best way to design databases
- 1. Hvernig á að skipuleggja gagnagrunn í töflur
How to organize a database with multiple tables/relations
- 2. Hvernig hægt er að bera kennsl á galla í hönnun gagnagrunns
Recognizing design flaws in databases
- 3. Hvernig hægt er að laga lélegan gagnagrunn
How to fix a bad design
- 4. Formlegar aðferðir til að skilgreina tengsl upplýsinga
Formal methods for defining data relationships
- 5. Formlegar aðferðir til að brjóta upp töflur
Formal methods for breaking up tables

Fallákveður (functional dependencies)

- Helsta tækið sem við höfum til að skilgreina tengsl milli gagna eru fallákveður (functional dependencies, FD)
- Eigindi (dálkar, attributes) $\bar{A} = A_1, \dots, A_n$ ákvarða (functionally determine) eigindi $\bar{B} = B_1, \dots, B_m$ í venslum R ef allar n -dir í R sem hafa sömu gildi fyrir A_1, \dots, A_n hafa einnig sömu gildi fyrir B_1, \dots, B_m
- Við táknum þetta með rithættinum (notation)

$$A_1, \dots, A_n \rightarrow B_1, \dots, B_m$$

eða

$$\bar{A} \rightarrow \bar{B}$$

Dæmi

title	year	length	studioName	starName
Star Wars	1977	124	Fox	Carrie Fisher
Star Wars	1977	124	Fox	Harrison Ford
Star Wars	1977	124	Fox	Mark Hamill
Empire Strikes Back	1980	111	Fox	Harrison Ford
Terms of Endearment	1983	132	MGM	Debra Winger
Terms of Endearment	1983	132	MGM	Jack Nicholson
The Usual Suspects	1995	106	MGM	Kevin Spacey

- Hvaða fallákveður gilda um þessi vensl?
What FD's might hold here?

Dæmi

Númer	Heiti	Nafn	Notandanafn	Námsleið	Hópur	Dagur	Tími
TÖL303G	Gagnasafnsfræði	Nonni	abcd	Eðlisfræði	d4	Mið	15:50-17:20
TÖL303G	Gagnasafnsfræði	Sigga	bcde	Stærðfræði	d3	Mið	08:20-09:50
TÖL303G	Gagnasafnsfræði	Sigga	cdef	Tölvunarfræði	d2	Fim	15:50-17:20
TÖL304G	Forritunarmál	Sigga	cdef	Tölvunarfræði	d1	Mán	13:20-14:50

- Hvaða fallákveður gilda um þessi vensl?

What FD's might hold here?

(Mögulegir) Lyklar – (Candidate) Keys

- Eigindi $\bar{A} = A_1, \dots, A_n$ mynda **mögulegan lykil (candidate key)** í venslum R ef

A set $\bar{A} = A_1, \dots, A_n$ of attributes form a **candidate key** in relation R if

1. $\bar{A} \rightarrow B_i$ gildir fyrir alla dálka (öll eigindi) B_i í venslunum
 $\bar{A} \rightarrow B_i$ holds for all columns (all attributes) B_i in the relation

2. Ekkert eiginlegt hlutmengi eigindanna í \bar{A} ákvarðar alla dálka
No proper subset of the attributes in \bar{A} determines all columns

- Safn af dálkum er **yfirlykill (superkey)** ef það inniheldur mögulegan lykil

A collection of columns is a **superkey** if it contains a candidate key

- (Mögulegur) lykill er því áreiðanlega yfirlykill – A (candidate) key is therefore a superkey
- Yfirlykill er **ekki endilega** mögulegur lykill – A superkey is not necessarily a candidate key

Reglur um fallákveður – Rules for FD's

- Gegnvirkni: Fyrir vensl $R(A, B, C)$ gildir:
Transitivity: For a relation $R(A, B, C)$ we have:
Ef $A \rightarrow B$ og $B \rightarrow C$ þá $A \rightarrow C$
If $A \rightarrow B$ and $B \rightarrow C$ then $A \rightarrow C$
- Skipting/sameining: Ef $\bar{B} = \{B_1, \dots, B_m\}$ er safn eiginda þá eru eftirfarandi fullyrðingar jafngildar:
Splitting/Combining: if $\bar{B} = \{B_1, \dots, B_m\}$ is a collection of attributes then the following are equivalent:

$$\begin{array}{l} \bar{A} \rightarrow \bar{B} \\ \text{og/and} \\ \bar{A} \rightarrow B_1 \wedge \dots \wedge \bar{A} \rightarrow B_m \end{array}$$

Fáfengilegar fallákveður – Trivial FD's

- Fallákveða $\bar{A} \rightarrow \bar{B}$ er fáfengileg (trivial) ef $\bar{B} \subseteq \bar{A}$
An FD $\bar{A} \rightarrow \bar{B}$ is trivial if $\bar{B} \subseteq \bar{A}$
- Allar fáfengilegar fallákveður eru sannar (auðvitað – það er fáfengileg fullyrðing)
All trivial FD's are true (that is a trivial statement)
- Fallákveðan $\bar{A} \rightarrow \bar{B}$ er því jafngild fallákveðu $\bar{A} \rightarrow \bar{C}$ þar sem:
The FD $\bar{A} \rightarrow \bar{B}$ is therefore equivalent to an FD $\bar{A} \rightarrow \bar{C}$ where:

$$\bar{C} = \bar{B} - \bar{A}$$

eða, eins og sumir skrifa/or, as some write:

$$\bar{C} = \bar{B} \setminus \bar{A}$$

Lokun eigindamengis – Closure of a set of attributes

- Ef \bar{A} er safn (mengi) eiginda og F er safn af fallákveðum fyrir vensl R , þá er **lokun** \bar{A} mengi þeirra eiginda \bar{B} í venslunum R þannig að F leiðir til þess að $\bar{A} \rightarrow \bar{B}$

If \bar{A} is a collection (set) of attributes and F is a collection of FD's for a relation R , then the **closure** of \bar{A} is the collection of attributes \bar{B} in relation R such that F implies that $\bar{A} \rightarrow \bar{B}$

- Lokun \bar{A} er táknuð með rithættinum \bar{A}^+

The closure of \bar{A} is denoted by \bar{A}^+

Dæmi um útreikning lykils – Computing a key

- Gerum ráð fyrir venslum $R(A, B, C)$ með fallákveður $AB \rightarrow C$ og $C \rightarrow B$
- Eiginleikinn A kemur hvergi fyrir í hægri hlið fallákveðu og hlýtur því að vera hluti allra mögulegra lykla
- Lokunin af A er hins vegar $\{A\}^+ = \{A\}$, sem er ekki allt eiginleikamengið, A er því ekki lykill
- Hins vegar eru lokanirnar af AB og AC allt eigindamengið:
 $\{A, B\}^+ = \{A, C\}^+ = \{A, B, C\}$
- Bæði AB og AC eru því yfirlyklar (superkeys) og mögulegir lyklar (candidate keys) og eru því lyklar fyrir venslin R

Dæmi um útreikning lykils – Computing a key

- Gerum ráð fyrir venslum $R(A, B, C, D, E, F, G, H, I, J)$ með fallákveður $AB \rightarrow C, BD \rightarrow EF, AD \rightarrow GH, A \rightarrow I$ og $H \rightarrow J$
- Eiginleikarnir ABD koma hvergi fyrir í hægri hlið fallákveðu og hljóta því að vera hluti allra mögulegra lykla
- Lokunin af ABD er $\{A, B, D\}^+ = \{A, B, C, D, E, F, G, H, I, J\}$, sem er allt eiginleikamengið
- ABD er því yfirlykill (superkey) og mögulegur lykill (candidate key) og er því lykill – reyndar sá eini sem kemur til greina

Lokun (closure) fallákveðumengis – Closure of a set of FD's

- Ef S er safn (mengi) af fallákveðum fyrir vensl R , þá er **lokun** S mengi þeirra fallákveða S' í venslunum R þannig að sérhvert $\bar{A} \rightarrow \bar{B}$ í S' er afleiðing af fallákveðunum í S

If S is a collection (set) of FD's for a relation R , then the **closure** of S is the set S' of those FD's over R such that every $\bar{A} \rightarrow \bar{B}$ in S' is a consequence of the FD's in S

- Lokun S er táknueð með rithættinum S^+

The closure of S is denoted by S^+

Algrím fyrir lokun eigindamengis

Algorithm for closure of attribute set

Inntak: Vensl R , fallákveður S á R og mengi eiginda \bar{A} í venslum R

Úttak: \bar{A}^+ (þ.e. lokunin á eigindamengi \bar{A})

1. Skiptum upp fallákveðunum í S þannig að aðeins eitt eigindi sé hægra megin
2. Frumstillum $X = \bar{A}$
3. Leitum að fallákveðu $\bar{B} \rightarrow C$ þannig að $\bar{B} \subseteq X$ en $C \notin X$,
 - bætum C við X
 - og endurtökum meðan hægt er
4. Skilum X

Útvíkkuð gegnvirkniregla

Extended transitivity rule

- Ef \bar{A} , \bar{B} og \bar{C} eru mengi eiginda (dálka) í venslum R og $\bar{A} \rightarrow \bar{B}$ og $\bar{B} \rightarrow \bar{C}$ gilda, þá gildir $\bar{A} \rightarrow \bar{C}$
- Til dæmis, fyrir einhvern kvikmyndagagnagrunn gildir $title, year \rightarrow studioName$ og $studioName \rightarrow studioAddress$, sem leiðir til þess að $title, year \rightarrow studioAddress$
- Takið eftir að hér erum við að hugsa okkur kvikmyndagagnagrunninn í heild sem ein stór vensl, ekki safn af minni mismunandi venslum

Grunnur, þekja (basis, cover)

- Tvö söfn af fallákveðum, S og S' eru sögð **jafngild** ef allar fallákveður $\bar{A} \rightarrow B$ sem eru afleiðingar af S eru einnig afleiðingar af S' og öfugt
- Fyrir fallákveður (safn af fallákveðum) S' sem er jafngildar fallákveðum S segjum við einnig að S' sé **þekja eða grunnur** (cover, basis) fyrir S
- Grunnur B fyrir fallákveður S er **lágþekja eða lággrunnur** (minimal cover, minimal basis) ef
 1. Allar fallákveður í B hafa ein eigindi (eitt dálkanafn) hægra megin
 2. Ef fallákveða er fjarlægð úr B þá er B ekki lengur grunnur fyrir S
 3. Ef við fjarlægjum eiginleika úr vinstri hlið í einhverri fallákveðu í B þá er B ekki lengur grunnur fyrir S

Fallákveður og ofanvarp (FD's and projections)

- Látum R vera vensl með fallákveður S
- Látum venslin R_1 vera skilgreind með ofanvarpi, $R_1 = \pi_L(R)$
- Hvaða fallákveður gilda þá um R_1 ?
- Þær fallákveður eru afleiðingar af S
- Þær fallákveður nota einungis eigindi (dálka) í R_1

Algrím fyrir lágþekju (lággrunn, minimal basis, minimal cover) hlutvensla

Algorithm for minimal basis of subrelation (a projection of a relation)

Inntak: Vensl R , fallákveður S og $R_1 = \pi_L(R)$

Úttak: Lágþekja fyrir R_1

1. Látum $T = \emptyset$
2. Fyrir sérhvert hlutmengi $\bar{X} \subseteq L$ af eigindum (dálkamengi) R_1 reiknum við \bar{X}^+
3. Bætum við T öllum fallákveðum $\bar{X} \rightarrow A$ þar sem $A \in \bar{X}^+ \cap L$
4. T er nú grunnur fyrir fallákveður R_1 -- finnum nú lágþekju
 - Ef ein fallákveða í T er afleiðing af hinum þá fjarlægjum við hana úr T
 - Ef til er fallákveða $\bar{Y} \rightarrow B$ í T þar sem \bar{Y} inniheldur a.m.k. tvo dálka og til er ekki tómt $\bar{Z} \subset \bar{Y}$ þannig að $\bar{Y} \rightarrow B$ er afleiðing af $(T - \{\bar{Y} \rightarrow B\}) \cup \{\bar{Z} \rightarrow B\}$, þá gefum við T nýja gildið $(T - \{\bar{Y} \rightarrow B\}) \cup \{\bar{Z} \rightarrow B\}$ --- (fjarlægjum sem sagt $\bar{Y} \rightarrow B$ og setjum $\bar{Z} \rightarrow B$ í staðinn)
 - Endurtökum þetta tvennt þar til ekkert breytist, skilum þá T með áorðnum breytingum

Dæmi um lágþekju (lággrunn) – Example of minimal cover (minimal basis)

- Gerum ráð fyrir heildarvenslum $R(A, B, C)$ með fallákveðum $\{AB \rightarrow C, C \rightarrow B\}$, ásamt hlutvenslum $R_1(B, C)$
Assume a relation $R(A, B, C)$ with FD's $\{AB \rightarrow C, C \rightarrow B\}$, and a projected relation $R_1(B, C)$
- Eina fallákveðan sem verkar inni í R_1 er þá $C \rightarrow B$ og lágþekjan fyrir R_1 er því $\{C \rightarrow B\}$
The only FD that works inside R_1 is then $C \rightarrow B$ and the minimal basis for R_1 is therefore $\{C \rightarrow B\}$

Dæmi um lágþekju (lággrunn) – Example of minimal cover (minimal basis)

- Gerum ráð fyrir heildarvenslum $R(A, B, C, D, E, F, G, H, I, J)$ með fallákveðum

$$\{AB \rightarrow C, BD \rightarrow EF, AD \rightarrow GH, A \rightarrow I, H \rightarrow J\}$$

ásamt hlutvenslum $R_1(A, D, G, H, I, J)$

- Fallákveðurnar sem verka inni í R_1 eru þá

$$\{AD \rightarrow G, AD \rightarrow H, A \rightarrow I, H \rightarrow J\}$$

og þær mynda lágþekju fyrir R_1

Frávik í gagnagrunnum – Database Anomalies

Frávik (anomaly, illbrigði) geta gerst þegar gagnagrunnur er ekki alveg rétt hannaður

1. Endurtekningar (**update anomaly**): Þegar sömu upplýsingar eru endurteknaðar og skrá þarf eða breyta sömu upplýsingum á fleiri en einum stað
2. Eyðingar (**deletion anomaly**): Þegar eytt er upplýsingum hverfa aðrar upplýsingar
3. Innsetningar (**insertion anomaly**): Þegar ekki er hægt að skrá upplýsingar

title	year	length	studioName	starName
-----	----	-----	-----	-----
Star Wars	1977	124	Fox	Carrie Fisher
Star Wars	1977	124	Fox	Harrison Ford
Star Wars	1977	124	Fox	Mark Hamill
Empire Strikes Back	1980	111	Fox	Harrison Ford
Terms of Endearment	1983	132	MGM	Debra Winger
Terms of Endearment	1983	132	MGM	Jack Nicholson
The Usual Suspects	1995	106	MGM	Kevin Spacey

Uppbrot (páttun, decomposition) á töflu

Decomposition of a table (relation)

- Ef tafla (vensl) R hefur 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, til dæmis
 $\{title, year, length, studioName\}, \{title, year, starName\}$
- Önnur eru slæm, til dæmis
 $\{title, year\}, \{year, length, starName, studioName\}$

BCNF staðalsnið (Boyce-Codd Normal Form)

- BCNF segir til um hvernig skipuleggja má töflur til að losna við frávik (illvik, anomaly) – BCNF solves anomalies
- Vensl R eru á BCNF sniði þá og því aðeins að
 - ef fallákveða $\bar{A} \rightarrow \bar{B}$ gildir innan R , og er ófáfengileg (þ.e. $\bar{B} \subseteq \bar{A}$ gildir ekki) þá er \bar{A} yfirlykill
 - Með öðrum orðum, ef \bar{A} ákvarðar meira en sjálft sig innan R þá ákvarðar \bar{A} allt innan R
 - Sometimes this is said: a relation is in BCNF if for any non-trivial FD that works inside the relation the left hand side has attributes that are **the key, the whole key, and nothing but the key (so help me Codd)**
- Það eru til fleiri staðalsnið (1NF, 2NF, **3NF**, **BCNF**, EKNF, 4NF, 5NF, o.fl.)
 - við munum leggja áherslu á BCNF og 3NF

Sömu upplýsingar

- Þegar við þáttum (brjótum upp) vensl á réttan hátt, þá er hægt að fá upphaflegu venslin með náttúrlegri tengingu (natural join)
- Chase algrímið (3.4.2, bls. 92) notar þáttunina og fallákveður til að komast að því hvort gögnin varðveitist nákvæmlega (lossless join)
- Ef R er tafla og R er þáttuð í hlutvensl $R_1 = \pi_{S_1}(R), \dots, R_n = \pi_{S_n}(R)$ þar sem S_1, \dots, S_n eru eigindamengin fyrir hin mismunandi hlutvensl, þá viljum við að þetta gildi:

$$R = R_1 \bowtie \dots \bowtie R_n$$

- Chase algrímið (sjá síðar) staðfestir hvort:

$$R \supseteq R_1 \bowtie \dots \bowtie R_n$$

- Ef svo er þá er þáttunin taplaus (lossless join) því öruggt er að:

$$R \subseteq R_1 \bowtie \dots \bowtie R_n$$

Almenn markmið þáttunar (uppbrots) vensla

1. Viðhald allra upplýsinga (lossless join, taplausar tengingar)

- Er $R = R_1 \bowtie \dots \bowtie R_n$?

2. Viðhald fallákveðna

- Ef sanngildi fallákveðanna er tryggt í sérhverjum af hlutvenslunum R_1, \dots, R_n hverjum fyrir sig, veldur það því að sanngildi fallákveðanna sé tryggt í $R_1 \bowtie \dots \bowtie R_n$?

3. Útrýming frávika (anomalies)

- Það er ekki alltaf hægt að ná öllum markmiðum samtímis
 - Við verðum stundum að velja milli 2 og 3 – við förnum ekki 1 og veljum líklega frekar 2 en 3

Dæmi um chase

- Íhugum $R(A, B, C)$ með fallákveðum $\{AB \rightarrow C, C \rightarrow B\}$
- Þáttum í $R_1(A, C), R_2(B, C)$ (fallákveðan $AB \rightarrow C$ er þá ekki innan neinnar töflu, en veldur það vandræðum? – chase gefur svarið)
- Notum chase algrím á þáttunina, hér er byrjunarstaðan:

A	B	C
a	b_1	c
a_2	b	c

- klárum á töflunni í fyrirlestri

Annað dæmi um chase

- Íhugum $R(A, B, C)$ með fallákveðum $\{AB \rightarrow C, C \rightarrow B\}$
- Þáttum í $R_1(A, B), R_2(B, C)$ (fallákveðan $AB \rightarrow C$ er þá aftur ekki innan neinnar töflu, en veldur það nú vandræðum?)
- Notum chase algrím á þáttunina, hér er byrjunarstaðan:

A	B	C
a	b	c_1
a_2	b	c

- Við getum ekkert gert – engar raðir er hægt að tengja saman – þáttunin er **ekki taplaus**

Uppbrot (þáttun) yfir í BCNF

- Við getum tekið vensl R sem ekki eru á BCNF staðalsniði og þáttað þau í safn af nýjum venslum þannig að
 1. Nýju venslin eru öll á BCNF staðalsniði
 2. Hægt er að endurmynda gögnin úr R með tengingum
 3. Gott markmið (sem ekki alltaf næst) er að til sé lágþekja (lággrunnur) þannig að sérhver fallákveða í lágþekjunni er innan einna þeirra vensla sem út koma úr þáttuninni

BCNF algrím

Inntak: Vensl R , fallákveður S

Úttak: Mengi hlutvensla R sem er BCNF þáttun R miðað við S

1. Ef R er á BCNF sniði þá skilum við $\{R\}$
2. Annars finnum við einhver BCNF frávík, $\bar{X} \rightarrow Y$ innan S^+ , þ.e. ófáfengilega (nontrivial) fallákveðu þannig að \bar{X} er ekki yfirlykill (superkey), þ.e. \bar{X} dugar ekki til að einkvæmt ákvarða röð í R
 - Reiknum \bar{X}^+
 - Setjum $R_1 = \pi_{\bar{X}^+}(R)$
 - Setjum $R_2 = \pi_L(R)$ þar sem L er sammengið af \bar{X} og þeim dálkum í R sem ekki eru í \bar{X}^+
3. Reiknum fallákveður fyrir venslin R_1 og R_2 , köllum þær S_1 og S_2
4. Reiknum endurkvæmt BCNF fyrir R_1, S_1 og R_2, S_2 , skilum sammenginu af niðurstöðunum

Tilgangur BCNF og staðalsniða almennt

- Losna við öll frávik
 - Endurtekningafrávik – margskráning sömu upplýsinga
 - Breytingafrávik – breyta þarf sömu upplýsingunum á mörgum stöðum
 - Eyðingafrávik – eyðing getur valdið eyðingu of mikilla upplýsinga
 - Innsetningafrávik – ekki er hægt að skrá þekktar upplýsingar
- Varðveita tengsl upplýsinga
 - Fallákveðurnar tengja saman upplýsingar

Einfalt dæmi um BCNF þáttun

- Gerum ráð fyrir venslum $R(A, B, C, D, E)$ með $AB \rightarrow C$ og $BC \rightarrow DE$
- Eini mögulegi lykillinn í R er AB
- Fallákveðan $BC \rightarrow D$ brýtur BCNF skilyrði því BC er ekki yfirlykill í R
- Brjótum því R upp í $R_1(B, C, D, E)$ og $R_2(A, B, C)$, sem uppfyllir BCNF skilyrði
- Lykillinn í R_1 er BC , lykillinn í R_2 er AB (sami og í R)

Virkar BCNF alltaf?

- Oftast, en ekki alltaf, sjá í [wikipediu](https://en.wikipedia.org/wiki/Boyce%E2%80%93Codd_normal_form) og kafla 3.4.4 í bókinni, bls. 96-97
- Til dæmis er **ekki til** BCNF staðalsnið fyrir vensl $R(A, B, C)$ með fallákveðum $\{AB \rightarrow C, C \rightarrow B\}$ **sem varðveitir allar fallákveður**
- Þátta má í $R_1(A, C), R_2(C, B)$ en þá er fallákveðan $AB \rightarrow C$ ekki innan neinnar töflu
- Þessi þáttun er taplaus (lossless decomposition), þ.e. $R = R_1 \bowtie R_2$, en gera þarf sérstakar ráðstafanir í gagnagrunninn til að tryggja fallákveðuna $AB \rightarrow C$
 - Tryggja þarf að ekki séu til **mismunandi** n-dir (a, c) og (a, c') í R_1 ásamt (c, b) og (c', b) í R_2
 - Ef svo væri þá væru n-dirnar (a, c, b) og (a, c', b) **báðar** í $R_1 \bowtie R_2$, sem væri þá í mótsögn við fallákveðuna $AB \rightarrow C$
 - Einnig þarf að tryggja að fyrir sérhverja n-d (a, c) í R_1 sé til n-d (c, b) í R_2
- Hér veljum við frekar 3NF staðalsnið, sem gefur þáttunina $R_1(A, B, C), R_2(C, B)$
 - Þægilegt að skorða með því að heimta að öll (b, c) í R_1 séu einnig í R_2 (tvískráning!)
 - Þáttunin kemur beint af augum úr lágþekjunni $\{AB \rightarrow C, C \rightarrow B\}$
 - Athugið samt að BCNF er almennt betra en 3NF (ekki ef fallákveður glatast)

Annað dæmi um chase

- Íhugum $R(A, B, C, D, E)$ með fallákveðum $\{AB \rightarrow C, BC \rightarrow DE\}$
- Þáttum í $R_1(B, C, D, E), R_2(A, B, C)$, sem uppfyllir BCNF án þess að glata fallákveðum
- Notum chase algrím á þáttunina, hér er byrjunarstaðan:

A	B	C	D	E
a_1	b	c	d	e
a	b	c	d_2	e_2

- klárum á töflunni í fyrirlestri

3NF staðalsnið

- 3NF er annað snið til að skipuleggja töflur til að losna við frávik
- Eilítið veikara en BCNF
 - Meira um tvískráningar upplýsinga
- Vensl R eru á 3NF sniði þá og því aðeins að
 - ef fallákveða $\bar{A} \rightarrow \bar{B}$ gildir innan R , og er ófáfengileg (þ.e. $\bar{B} \subseteq \bar{A}$ gildir ekki)
 - þá er annaðhvort \bar{A} yfirlykill fyrir R **eða** sérhvert eigindi í $\bar{B} - \bar{A}$ er hluti af einhverjum mögulegum lykli

3NF algrím

Inntak: Vensl R , fallákveður F

Úttak: Mengi hlutvensla R sem er 3NF þáttun R miðað við S

1. Finnum lágþekju (lággrunn) G sem er jafngild fallákveðusafninu F
2. Fyrir sérhverja fallákveðu $\bar{X} \rightarrow \bar{Y}$ í G búum við til hlutvensl $R_i(\bar{X}\bar{Y})$
 - Til dæmis ef $\bar{X} = ABC$ og $\bar{Y} = DE$ þá búum við til $R_i(A, B, C, D, E)$
3. Ef einhver venslanna R_1, \dots, R_n sem út koma innihalda mögulegan lykil fyrir R (þ.e. eigindi venslanna mynda yfirlykil fyrir R) þá skilum við strax $\{R_1, \dots, R_n\}$
4. Annars búum við til ný hlutvensl $R'(\bar{X})$ þar sem \bar{X} er mögulegur lykill R og skilum síðan $\{R_1, \dots, R_n, R'\}$

Algrím fyrir lágþekju

Inntak: Safn F af fallákveðum

Úttak: Samsvarandi lágþekja G

1. Frumstillum $G = F$
 2. Breytum ákveðum $X \rightarrow YZ$ í tvær eða fleiri ákveður $X \rightarrow Y$ og $X \rightarrow Z$
 3. Breytum ákveðum $XY \rightarrow Z$ í $X \rightarrow Z$ ef G^+ breytist ekki
 4. Fjarlægjum ákveður $X \rightarrow Z$ úr G ef G^+ breytist ekki
- Þegar þessu lýkur (ekkert meira er hægt að gera) er G lágþekja fyrir fallákveðusafnið F
 - Athugið samt að stundum er fleiri en ein möguleg lágþekja

Einfalt dæmi um 3NF þáttun

- Gerum ráð fyrir venslum $R(A, B, C, D, E)$ með $AB \rightarrow C$ og $BC \rightarrow DE$
- Eini mögulegi lykillinn í R er AB
- Lágþekja er $\{AB \rightarrow C, BC \rightarrow DE\}$
- Brjótum því R upp í $R_1(B, C, D, E)$ og $R_2(A, B, C)$, sem uppfyllir 3NF skilyrði (og reyndar einnig BCNF skilyrði, sem er sterkara skilyrði)
- Lykillinn í R_1 er BC , lykillinn í R_2 er AB (sami og í R)
- Sama þáttun og við sáum áður sem BCNF
- Vorum búin að keyra chase og staðfesta taplausar tengingar

Frumeigindi, 3NF og BCNF

- **Skilgreining:** Gerum ráð fyrir venslum R með safni fallákveða F . Eigindi A í R eru þá **frumeigindi** í R ef til er mögulegur lykill sem inniheldur A .
- Munurinn á BCNF og 3NF er sá að í 3NF er **leyfilegt** að fallákveða $\bar{X} \rightarrow Y$ sé til staðar innan vensla R þótt \bar{X} sé ekki yfirlykill R , ef Y er frumeigindi (m.v. að Y sé eitt eigindi)
- Í BCNF, hins vegar, er aðeins leyfilegt að fallákveða $\bar{X} \rightarrow Y$ sé til staðar innan vensla R ef \bar{X} er yfirlykill R

Samband 3NF og BCNF

- Allar þáttanir sem eru BCNF eru 3NF
- Sumar þáttanir sem eru 3NF eru ekki BCNF
- Sem sagt: $BCNF \Rightarrow 3NF$, en ekki öfugt

Yfirlit

- Lokun eigindamengis (dálkasafns)
- Lykill vensla
- Lágþekja (lággrunnur) fallákveðusafns
- Prófun taplausra tenginga
- Þáttun í BCNF og 3NF
- Þáttum helst í BCNF (til að minnka endurtekningar)
 - **með taplausum tengingum og viðhaldi fallákveða**
- Ef það bregst sættumst við á 3NF
 - **með taplausum tengingum og viðhaldi fallákveða**