Adatbázisok elmélete 17. előadás

Csima Judit
Budapesti Műszaki és Gazdaságtudományi Egyetem
Számítástudományi Tsz.
I. B. 136/b
csima@cs.bme.hu

2003. Április 9.

Normalizálás

Tétel. *Tetszőleges (R,F) sémának van hűséges felbontása BCNF relációkra.*

Bizonyítás: Elve:

- Ha (R,F) BCNF $\surd$
- Ha nem, akkor két valódi (kisebb) részre bontjuk hűségesen $\Longrightarrow (R_1,R_2)$
- Ezt ismételjük (R_1,R_2)-re.

Ez véget fog érni, mert ha már csak 2 attribútum marad valamelyikben, azt nem kell tovább bontani.

Hűséges lesz, mert láttuk, hogy ha egy hűséges felbontás egyik részét tovább bontjuk, akkor hűséges marad.

Hogyan bontjuk fel 2 valódi részre, hűségesen?

Keresünk a felbontandó sémában egy olyan $X\to A\in F^+$-t, ami sérti a BCNF tulajdonságot $\Longrightarrow$ A és X része a sémának, $A\notin X$ és X nem szuperkulcs

$$R_1 := XA, \qquad R_2 := R\setminus\{A\}$$

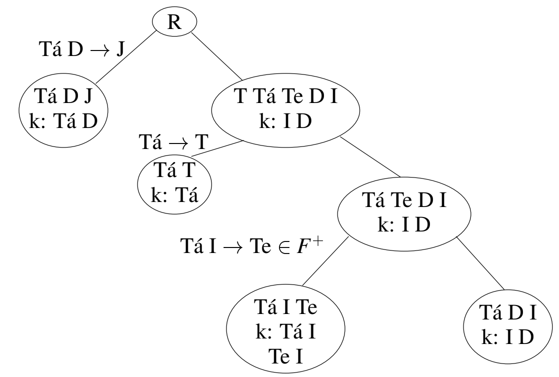
Ezek kisebbek: R_2 nyilván, R_1 pedig azért, mert ha $R_1 = R$ volna, akkor $X\to XA = R$ miatt X szuperkulcs lett volna.

Hűséges a felbontás: kétrészes teszttel $R_1\cap R_2 = X\to A = R_1\setminus R_2$ $\surd$

Miért lesz jobb ez a felbontás? Az $X\to A$ függéssel nem lesz több probléma: R_2-ben nincs A, így nem lehet baj. R_1-ben viszont X szuperkulcs lesz.

Példa:
R(**T**anár, **Tá**rgy, **Te**rem, **D**iák, **J**egy, **I**dő)

$F = \{$Tá $\to$ T; IT $\to$ Te; ID $\to$ Te; ID $\to$ Tá; TáD $\to$ J$\}$ $\Longrightarrow$ kulcs csak ID



Megjegyzések:
Minden felbontás után meg kell nézni, hogy a kapott relációk BCNF-ben vannak-e. Ehhez meg kell konstruálni F_S^+-et, ha S a vizsgált reláció: ez az F_R^+ azon függéseiből áll, amiknek mindkét oldala F-ben van. Ezeket a függéseket úgy kapjuk, hogy minden

$X \subseteq S$ részhalmazra kiszámoljuk $X^+(F)$-et és $X \to Y$ pontosan akkor lesz benne F_S^+-ben, ha $Y \subseteq X^+(F) \cap S$.
Általában nem igaz, hogy elég F-ből kiválogatni azokat, amiknek mindkét oldala S-ben van.

Pl.:
Ha $S =$ Tá Te D I, akkor (csak a nemtrivi függéseket felírva): $F_S^+ = \{$Tá I $\to$ Te; D I $\to$ Te Tá; D I Tá $\to$ Te; D I Te $\to$ Tá$\}$

3 attribútum esetén a BCNF tulajdonság csak úgy sérülhet, ha $X \to Y$, ahol X, Y egy-egy attribútum és X nem kulcs.

Azt is mindig ellenőrizi kell, hogy a kapott relációkban mik a (szuper)kulcsok, hogy egy függésről el tudjuk dönteni, hogy sérti-e a BCNF-et vagy nem. A példában ez viszonylag könnyű lesz, hiszen I és D egyik F-beli függőségben sem szerepel a jobb oldalon, így minden kulcs (amikor I és D szerepel a relációban) tartalmazza I D-t. Csak azt kell megnézni, hogy I D kulcs marad.

Függőség megőrzése

BCNF egy fogyatékossága: nehéz lehet ellenőrizni, hogy teljesülnek-e F függései (pl. beszúráskor). Ilyenkor a költséges $\bowtie$ kell, és ez sokszor előfordulhat.

Kéne egy olyan felbontás, amin könnyen lehet ellenőrizni a függéseket.

Definíció. *Adott (R, F) séma és ennek egy $\rho = (R_1, \ldots, R_k)$ felbontása.*

$$\pi_\rho(F) := \{X \to Y \in F^+ \mid \exists i\ (1 \le i \le k)\ X, Y \subseteq R_i\}^+$$

*az F **függéseinek vetítése a** ρ **felbontásra**. ρ **függőségőrző**, ha $\pi_\rho(F) = F^+$.*

Megjegyzés: $\pi_\rho(F) \subseteq F^+$ persze mindig igaz.

Ha a felbontás függőségőrző, akkor elég a darabokon ellenőrzni valamit, ami garantálja, hogy F minden függése fennmarad az egészen.

Példa: R(**V**áros, **U**tca, **I**rányítószám) $F = \{VU \to I;\ I \to V\}$
Ez nem BCNF $I \to V$ miatt.

Mire jó a függőségőrzés?:
Ha felbontjuk $\Longrightarrow S(V, I), Q(I, U)$
Beszúrunk 2-2 sort:

S	V	I
	Nagykanizsa	8800
	Nagykanizsa	8831

Q	U	I
	Kossuth	8800
	Kossuth	8831

Noha S-ben és Q-ban oké minden, $S \bowtie Q$-ban nem teljesül a $VU \to I$ függés. Ez nem lett volna, ha függőségőrző lenne a felbontás.

Szomorú példa ez: semelyik felbontása sem őrzi meg $VU \to I$-t, mert csak ez olyan függés, aminek jobb oldalán van I, azaz ha egy felbontás függőségőrző lenne, akkor

egy tagjában kéne VUI-nek lennie, de az nem lenne valódi felbontás.
$\Longrightarrow$ ennek nincs függőségőrző valódi felbontása, vagyis van olyan reláció, amit nem lehet függőségőrző módon BCNF-ekre szétszedni
$\Longrightarrow$ felbontás BCNF-be nem feltétlenül függőségőrző.

Kellene egy gyengébb normálforma. Ebben lehet valamennyi redundancia, de legyen függőségőrző.

3NF

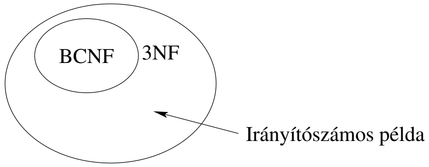
Definíció. *Az (R,F) séma A attribútuma **prím*** (elsődleges)*, ha szerepel valamelyik kulcsban.*

Szuperkulcsban minden szerepel, kulcs helyett szuperkulccsal nem lenne sok értelme az előbbi definíciónak.

Definíció. *Az (R,F) séma **3NF*** (harmadik normálformájú)*, ha tetszőleges nemtriviális $X \to A \in F^+$ függés esetén vagy X szuperkulcs vagy A prímattribútum.*

Következmény. *Minden BCNF séma egyben 3NF is.*

3NF lehet redundáns, de nem nagyon.



Tétel. *Ha (R,F) egy 3NF séma, akkor minden nem prím A attribútumra és $X \subseteq R$ kulcsra igaz, hogy nincs olyan Y, hogy $X \to Y$, $Y \nrightarrow X$, $Y \to A$ és $A \notin Y$. (Nem-elsődleges attribútum nem függ tranzitívan kulcstól.)*

Nem bizonyítjuk, úgy menne, mint BCNF-nél a hasonló állítás.

Tétel. *Tetszőleges (R,F) sémának van hűséges és függőségőrző felbontása 3NF sémákra.*

Definíció. *A G függéshalmaz az F függéshalmaz **fedése** ha $G^+ = F^+$.(Persze ilyenkor F is fedése G-nek.)*

Definíció. *A G függéshalmaz az F függéshalmaz **minimális fedése**, ha egyrészt fedése, másrészt*

(1) a G-beli függések $X \to A$ alakúak, ahol $A \notin X$

(2) G-ből nem hagyható el függés: $(G \setminus \{X \to A\})^+ \subsetneq G^+$

(3) G-beli függések baloldalai minimálisak: $X \to A \in G$, $Y \subsetneq X \implies Y \to A \notin G^+$

Állítás. *Tetszőleges F-nek van minimális fedése.*

Bizonyítás majd a jövő órán lesz.

Lássuk inkább a tétel bizonyítását:

Bizonyítás: (Van jó 3NF felbontás.)
Vegyük F egy minimális fedését: $G = \{X_1 \to A_1, \ldots, X_k \to A_k\}$
Legyen X egy kulcs és $\rho = (X, X_1A_1, \ldots, X_kA_k)$ egy felbontás.
$\implies R_0, R_1, \ldots, R_k$
Állítás: ez függőségőrző lesz, a tagok 3NF-ek és a felbontás hűséges.

ρ **függőségőrző:** $F^+ = G^+$ és minden G-beli $X \to A_i$ függés benne lesz R_i-ben (ott ellenőrizhető).

R_0 **3NF:** R_0-ban nincs nemtriviális függés, mert különben X nem lenne kulcs, csak szuperkulcs $\implies R_0$ BCNF $\implies$ 3NF

Többi R_i is 3NF: tegyük fel, hogy nem az $\implies \exists\, U \to B$ nemtriviális függés, hogy U nem szuperkulcs R_i-ben és B nem prímattribútum R_i-ben.
Ha $B = A_i$, akkor $U \subseteq X_i$, de $U \neq X_i$, hiszen akkor U szuperkulcs lenne R_i-ben. $\implies U \subset X_i \implies X_i \to A_i$ baloldala csökkenthető G-ben U-ra.Ellentmondás, mert akkor

G nem volt minimális fedés.
Ha $B \neq A_i \implies B \in X_i$ és B nem prím R_i-ben $\implies X_i$ nem kulcs R_i-ben (de szuperkulcs) $\implies \exists\, Y \subsetneq X_i$ kulcs R_i-ben $\implies Y \to A_i$ fennáll $\implies X_i \to A_i$ baloldala csökkenthető G-ben Y-ra, megint csak ellentmondás.

ρ **hűséges:** Higgyük el, nem bizonyítjuk.