Adatbázisok elmélete 14. előadás

Katona Gyula Y.
Budapesti Műszaki és Gazdaságtudományi Egyetem
Számítástudományi Tsz.
I. B. 137/b
kiskat@cs.bme.hu
http://www.cs.bme.hu/˜kiskat

2005

Normálformák

Definíció. *Egy $X \to Y$ függés triviális, ha $Y \subseteq X$. (Mert ezek a függések nem hordoznak sok infót, mindig igazak.)*

Definíció (Boyce–Codd normálforma). *Az (R, F) relációs séma BCNF-ben van, ha tetszőleges nemtriviális $X \to A \in F^+$ függés esetén X szuperkulcs.*

Azaz csak olyan függések vannak, hogy a szuperkulcs mindent meghatároz.
Tétel. *Az (R, F) BCNF-ben van pontosan akkor ha tetszőleges $A \in R$-re és $X \subseteq R$ kulcsra igaz, hogy nincs olyan $Y \subseteq R$, amire $X \to Y \in F^+;\ Y \not\to X;\ Y \to A \in F^+$ és $A \notin Y$. (Nincs tranzitív függés kulcstól.)*

Bizonyítás: Ha nincs BCNF-ben a séma, akkor van egy $Y \to A$ függés, ahol Y nem szuperkulcs és $A \notin Y$. Ekkor, tetszőleges X kulccsal: $X \to Y, Y \not\to X, Y \to A$, de $A \notin Y$, ami épp egy kulcstól való tranzitív függés.

Másrészt, ha van tranzitív függés kulcstól, azaz X olyan kulcs, amivel $X \to Y, Y \not\to X, Y \to A$, de $A \notin Y$, akkor $Y \to A$ egy olyan függés, ami sérti a BCNF tulajdonságot, mert Y nem lehet szuperkulcs, ha $Y \not\to X$. √

Miért jó a BCNF séma?

Ha $C \to B;\ B \to A$ teljesülne, de $B \to C$ nem, akkor ugyanaz a B érték több C érték mellett is előfordulhatna, de minden példánynál ugyanazt az A értéket is tároljuk $\Longrightarrow$ **redundancia**.

Állítás. *≤ 2 attribútumos reláció mindig BCNF.*

Bizonyítás: Ha $A \to B \Longrightarrow A$ kulcs. Ha $B \to A \Longrightarrow B$ kulcs.

Mivel F^+-ra követeljük meg a feltételt, nehéz ellenőrizni.
DE:

Tétel. *Ha (R, F) nem BCNF, akkor van olyan $X \to Y \in F$, amely jobboldalának valamely A attribútumára $X \to A$ nemtriviális és X nem szuperkulcs. (Az ilyen $X \to A \in F^+$.)*

Bizonyítás: Ha (R, F) nem BCNF, akkor van $U \to B \in F^+$, hogy U nem szuperkulcs és $B \notin U$.
$\Longrightarrow B \in U^+(F) \Longrightarrow U \subsetneq U^+(F)$
Az algoritmus, ami $U^+(F)$-et számolja, el tud indulni $\Longrightarrow \exists V \to W \in F$, melyre $V \subseteq U,\ W \not\subseteq U$
$\Longrightarrow V \to W$ jó lesz $X \to Y$-nak.
Ugyanis V nem szuperkulcs, hiszen $V \subseteq U$ és U nem szuperkulcs.
$W \not\subseteq U \Longrightarrow \exists A \in W \setminus U \subseteq W \setminus V$, így $V \to W$ nem triviális.

Ez jelentősen könnyíti az ellenőrzést, csak F függőségeit kell végignézni, nem F^+-ét.

Normalizálás

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

Bizonyítás: Elve:

- Ha (R, F) BCNF √
- 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.

Bizonyítás

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.

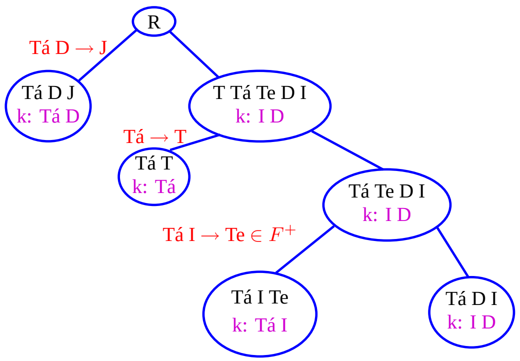
Hűséges a felbontás: kétrészes teszttel $R_1 \cap R_2 = X \to A = R_1 \setminus R_2$ $\surd$

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 S-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.: $F = \{$Tá $\to$ T; IT $\to$ Te; ID $\to$ Te; ID $\to$ Tá; TáD $\to$ J$\}$
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$\}$

$\Longrightarrow$ Az előző algoritmus lehet exponenciális $\Longrightarrow$ Van polinomiális algoritmus is.

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 ellenőrizni 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 ellenőrizni 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

Következmény

Állítás. *Felbontás BCNF-be nem feltétlenül függőségőrző.*

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

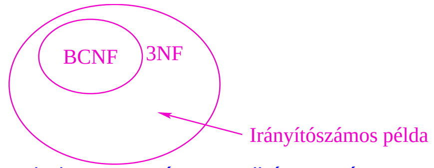
3NF

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 $X \to Y,\ Y \not\to X,\ Y \to A$ és $A \notin Y$. (Nem-elsődleges attribútum nem függ tranzitívan kulcstól.)*

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