

Adatbáziskezelés Normalizálás

Csima Judit

BME, VIK,
Számítástudományi és Információelméleti Tanszék

2017. november 15.

Tétel

Tetszőleges (R, F) sémának van hûséges felbontása BCNF relációkra.

Bizonyítás: Elve:

- Ha (R, F) BCNF, akkor kész.
- Ha nem, akkor két valódi (kisebb) részre bontjuk hûségesen $\implies (R_1, R_2)$
- Ezt ismételjük (R_1, R_2) -re.

Ez véget fog érni, mert ha már csak 2 attribútum marad valamelyikben, azt nem kell tovább bontani. Hûséges lesz, mert láttuk, hogy ha egy hûséges felbontás egyik részét tovább bontjuk, akkor hûséges marad.

Hogyan bontjuk fel 2 valódi részre, hûségesen?

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

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

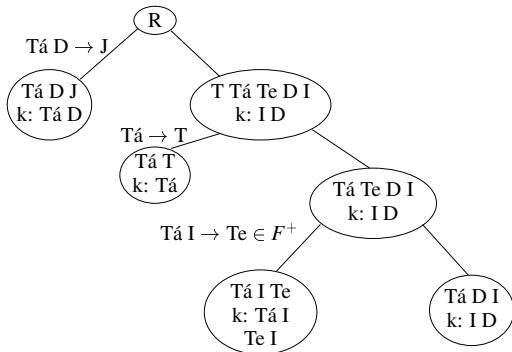
Ezek kisebbek: R_2 nyilván, R_1 pedig azért, mert ha $R_1 = R$ volna, akkor $X \rightarrow XA = R$ miatt X szuperkulcs lett volna.

Hûséges a felbontás: kétrészes teszttel $R_1 \cap R_2 = X \rightarrow A = R_1 \setminus R_2$
Miért lesz jobb ez a felbontás? Az $X \rightarrow A$ függéssel nem lesz több probléma: R_2 -ben nincs A , így nem lehet baj. R_1 -ben viszont X szuperkulcs lesz.

Példa

$R(\mathbf{T}$ anár, \mathbf{T} árgy, \mathbf{T} erem, \mathbf{D} iak, \mathbf{J} egy, \mathbf{I} dő)

$F = \{Tá \rightarrow T; IT \rightarrow Te; ID \rightarrow Te; ID \rightarrow Tá; TáD \rightarrow J\} \implies$ 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

Megjegyzés:

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 $X \rightarrow Y$ 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 $S = T \text{ á } T \text{ e } D \text{ I}$, akkor (csak a nemtrivi függéseket felírva):

$$F_S^+ = \{T \text{ á } T \text{ e}; D \text{ I} \rightarrow T \text{ e } T \text{ á}; D \text{ I } T \text{ á} \rightarrow T \text{ e}; D \text{ I } T \text{ e} \rightarrow T \text{ á}\}$$

Megjegyzések még

- 3 attribútum esetén a BCNF tulajdonság csak úgy sérülhet, ha $X \rightarrow Y$, 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.

Függőség megőrzése

BCNF-ekre bontás egy fogyatékosága: nehéz lehet ellenőrizni, hogy teljesülnek-e F függései (pl. beszúrásokor). Mert lehet, hogy egy $X \rightarrow Y$ függésnél X és Y attribútumai nem ugyanabban a részben vannak a felbontásnál. 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 $\rho = (R_1, \dots, R_k)$ felbontása.

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

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

- $\pi_\rho(F) \subseteq F^+$ persze mindig igaz.
- Ha a felbontás függőségörző, akkor elég a darabokon ellenőrizni az oda eső függéseket és ha itt minden rendben van, akkor F minden függése igaz lesz az egészen.

Példa

$R(\mathbf{Város, Utca, Irányítószám}) \quad F = \{VU \rightarrow I; I \rightarrow V\}$

Ez nem BCNF $I \rightarrow V$ miatt.

Miért kéne a függőségörzés?:

Ha felbontjuk $\implies S(V, I), Q(I, U)$

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 $VU \rightarrow I$ függés. Ez nem lett volna, ha függőségörző lenne a felbontás.

Szomorú példa

Szomorú példa ez: semelyik felbontása sem őrzi meg $VU \rightarrow I$ -t, mert csak ez olyan függés, aminek jobb oldalán van I , azaz ha egy felbontás függőségörző lenne, akkor egy tagjában kéne VUI -nek lennie, de az nem lenne valódi felbontás.

⇒ ennek nincs függőségörző valódi felbontása, vagyis van olyan reláció, amit nem lehet függőségörző módon BCNF-ekre szétszedni

⇒ felbontást BCNF-re nem feltétlenül lehet függőségörző módon megcsinálni.

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

Definíció

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

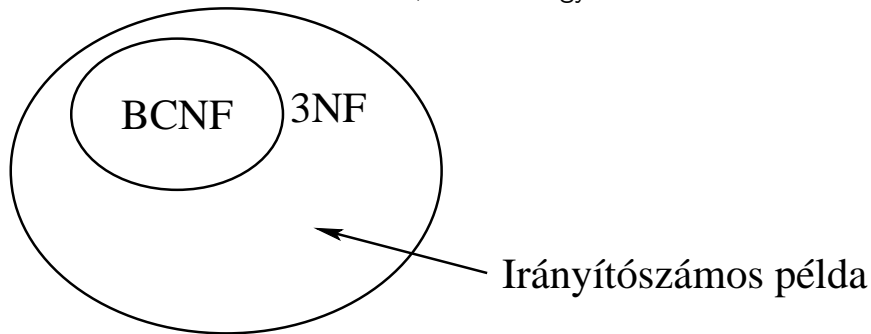
Kulcs helyett superkulccsal nem lenne sok értelme az előbbi definíciónak, mert superkulcsban minden szerepel.

Definíció

Az (R, F) séma **3NF** (harmadik normálformájú), ha tetszőleges nemtriviális $X \rightarrow A \in F^+$ függés esetén vagy X superkulcs vagy A primattribútum.

BCNF vs. 3NF

A definíció következménye: Minden BCNF séma egyben 3NF is. 3NF lehet redundáns, de nem nagyon.



Felbontás 3NF-re

Tétel

Tetszőleges (R, F) sémának van hûséges és függõségõrzõ felbontása 3NF sémákra.

Mindjárt látjuk, de elõbb kell egy új fogalom:

Definíció

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

Minimális fedés

Definíció

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

- (1) a G -beli függések $X \rightarrow A$ alakúak, ahol $A \notin X$ (A egy darab attribútum)
- (2) G -ből nem hagyható el függés: $(G \setminus \{X \rightarrow A\})^+ \subsetneq G^+$
- (3) G -beli függések baloldalai minimálisak: $X \rightarrow A \in G, Y \subsetneq X \implies Y \rightarrow A \notin G^+$

Állítás

Tetszőleges F -nek van minimális fedése.

Bizonyítás majd lesz mindjárt, de inkább lássuk most a főtétel bizonyítását.

Van jó felbontás 3NF-ekre (bizonyítás)

Vegyük F egy minimális fedését: $G = \{X_1 \rightarrow A_1, \dots, X_k \rightarrow A_k\}$

Legyen X egy kulcs és tekintsük a

$\rho = (X, X_1A_1, \dots, X_kA_k) = (R_0, R_1, \dots, R_k)$ felbontást.

Állítás: ez függőségörző lesz, a tagok 3NF-ek és a felbontás hûséges.

ρ függőségörző

$F^+ = G^+$ és minden G -beli $X_i \rightarrow A_i$ függés benne lesz R_i -ben (ott ellenőrizhető).

A tagok 3NF-ek ρ -ban

- R_0 3NF: R_0 -ban nincs nemtriviális függés, mert különben X nem lenne kulcs, csak szuperkulcs $\implies R_0$ még BCNF is, tehát 3NF is
- Többi R_i is 3NF: tegyük fel, hogy nem az, ekkor $\exists U \rightarrow B$ nemtriviális függés R_i -ben, hogy U nem szuperkulcs R_i -ben és B nem prímmattribútum R_i -ben (de U és B is R_i -ben van).
 - Ha $B = A_i$, akkor $U \subseteq X_i$, de $U \neq X_i$, (mert különben U szuperkulcs lenne R_i -ben) és így $X_i \rightarrow A_i$ baloldala csökkenthető lenne 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 prímm R_i -ben $\implies X_i$ nem kulcs R_i -ben (de szuperkulcs) $\implies \exists Y \subsetneq X_i$ kulcs R_i -ben $\implies Y \rightarrow A_i$ fennáll $\implies X_i \rightarrow A_i$ baloldala csökkenthető G -ben Y -ra, megint csak ellentmondás.

ρ hûséges:

Higgyük el, nem bizonyítjuk.

- Előfordulhat, hogy valamelyik $X_i A_i$ már tartalmaz kulcsot. Ilyenkor a $\rho = \{X_1 A_1, \dots, X_k A_k\}$ is jó felbontás már.
- 2NF már nem érdekes, 1NF kicsit érdekes, de nem foglalkozunk vele.

Minimális fedés

Definíció

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

- (1) a G -beli függések $X \rightarrow A$ alakúak, ahol $A \notin X$ (A egy darab attribútum)
- (2) G -ből nem hagyható el függés: $(G \setminus \{X \rightarrow A\})^+ \subsetneq G^+$
- (3) G -beli függések baloldalai minimálisak: $X \rightarrow A \in G, Y \subsetneq X \implies Y \rightarrow A \notin G^+$

Állítás

Tetszőleges F -nek van minimális fedése.

Minimális fedés készítése F -hez

Bizonyítás: Algoritmust adunk, külön gondoskodunk minden pont teljesítéséről.

(1) $X \rightarrow Y \in G$, $Y = A_1 \dots A_k \implies$ minden $X \rightarrow A_i$ -t beveszünk, ha $A_i \notin X$.

(2) Minden $X \rightarrow A \in G$ függésre kiszámoljuk $Y := X^+(G \setminus \{X \rightarrow A\})$ -t. Ha $A \in Y$, akkor $X \rightarrow A$ elhagyható (mert kijön a többiből), különben nem.

(3) Ellenőrizni kell, hogy $X \rightarrow A$ baloldala minimális-e. X minden B elemére kiszámoljuk $Y := (X \setminus \{B\})^+(G)$ -t. Ha $A \in Y$, akkor $X \rightarrow A$ helyett vegyük be $X \setminus \{B\} \rightarrow A$ -t. Ha egyik B -re se lesz ilyen, akkor X minimális.

Megjegyzés: És persze a fenti három lépés során a függéshalmaz lezártja nem változik.

Példa minimális fedés kiszámolására

$R = (A, B, C, D)$ $F = \{AB \rightarrow CD; AC \rightarrow BD; C \rightarrow A; C \rightarrow B\}$

(1) $F' = \{AB \rightarrow C; AB \rightarrow D; AC \rightarrow B; AC \rightarrow D; C \rightarrow A; C \rightarrow B\}$

(2) $C \rightarrow B$ miatt $AC \rightarrow B$ elhagyható és $AB \rightarrow C$ és $AC \rightarrow D$ miatt $AB \rightarrow D$ elhagyható, de más nem, ezt végig lehet nézni.

$F'' = \{AB \rightarrow C; AC \rightarrow D; C \rightarrow A; C \rightarrow B\}$

(3) $C \rightarrow A$ miatt $AC \rightarrow D$ baloldaláról A elhagyható.

$F''' = \{AB \rightarrow C; C \rightarrow D; C \rightarrow A; C \rightarrow B\}$

Ez már minimális fedés.

A minimális fedés nem feltétlenül egyértelmű!

Példa: $R(A, B, C)$ $F = \{AB \rightarrow C; A \rightarrow B; B \rightarrow A\}$ esetén jó minimális fedés lesz

$G_1 = \{B \rightarrow C; A \rightarrow B; B \rightarrow A\}$ és

$G_2 = \{A \rightarrow C; A \rightarrow B; B \rightarrow A\}$ is.

Példa 3NF-re bontásra

$R = (A, B, C, D, E)$ $F = \{AE \rightarrow BC; AC \rightarrow D; CD \rightarrow BE; D \rightarrow E\}$

Ez nem 3NF, mert a kulcsok:

semelyik egyelemû halmaz nem kulcs (csak D lehetne, de az ő lezártja csak DE), viszont kételemûek közül szuperkulcs lesz AC, AD, AE (A-nak benne kell lennie minden kulcsban, mert A nincs jobboldalon), AB viszont nem szuperkulcs.

Ezek kulcsok is lesznek, mert egyik egyelemû se volt kulcs.

Más kulcs nincs is, mert ha lenne legalább háromelemû halmaz, aminek a lezártja az egész, akkor abban A biztos benne van és legalább C vagy D vagy E is benne van, de akkor az már csak szuperkulcs lehet, mert tartalmaz kulcsot.

Innen látszik, hogy a primattribútumok: A, C, D, E, vagyis B nem az.

Tehát a $CD \rightarrow B$ függés rossz a 3NF szempontjából, mert CD nem szuperkulcs és B nem prim, szóval valamit kell

Példa tovább

Csináljunk hát egy 3NF-ekre való függőségörző, hûséges felbontást.

(1) $F' = \{AE \rightarrow B; AE \rightarrow C; AC \rightarrow D; CD \rightarrow B; CD \rightarrow E; D \rightarrow E\}$

(2) $AE \rightarrow C, AC \rightarrow D, CD \rightarrow B$ miatt $AE \rightarrow B$ elhagyható és $D \rightarrow E$ miatt $CD \rightarrow E$ is elhagyható, de más nem, ezt végig lehet nézni

$F'' = \{AE \rightarrow C; AC \rightarrow D; CD \rightarrow B; D \rightarrow E\}$

(3) Semelyik baloldal nem csökkenthető, mert például A lezártjában nincsen benne E, és a többi is ugyanígy látszik. Vagyis F'' már minimális fedés. A minimális fedés alapján a jó felbontás:

(AEC, ACD, CDB, DE) mivel kulcsot nem is kellett hozzávennünk, mert az már benne van az egyik tagban (pl. AE az elsőben).