Adatbáziskezelés
Függőségőrzés, 3NF-re bontás

Csima Judit

BME, VIK,
Számítástudományi és Információelméleti Tanszék

2018. november 9. és 16.

Hol tartunk?

Tétel

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

Függõség megõrzése

BCNF-ekre bontás egy fogyatékossága: nehéz adatmódosításkor ellenõrizni, hogy teljesülnek-e F függései (pl. beszúráskor). Mert lehet, hogy egy $X \to Y$ függésnél X és Y attribútumai nem ugyanabban a részben vannak a felbontásnál. Ilyenkor a költséges $\bowtie$ kell, és ez sokszor elõfordulhat.
Kéne egy olyan felbontás, amin könnyen lehet ellenõrizni a függéseket.

Definíció

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

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

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

Megjegyzések

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

Példa

R(**V**áros, **U**tca, **I**rányítószám) $F = \{VU \to I;\ I \to V\}$
Ez nem BCNF $I \to V$ miatt.
Miért kéne a függõségõrzés?:
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 függõségõrzõ lenne a felbontás.

Szomorú példa

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

$\Longrightarrow$ ennek nincs függõségõrzõ valódi felbontása, vagyis van olyan reláció, amit nem lehet függõségõrzõ módon BCNF-ekre szétszedni

$\Longrightarrow$ felbontást BCNF-re nem feltétlenül lehet függõségõrzõ módon megcsinálni.

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

3NF

Definíció

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

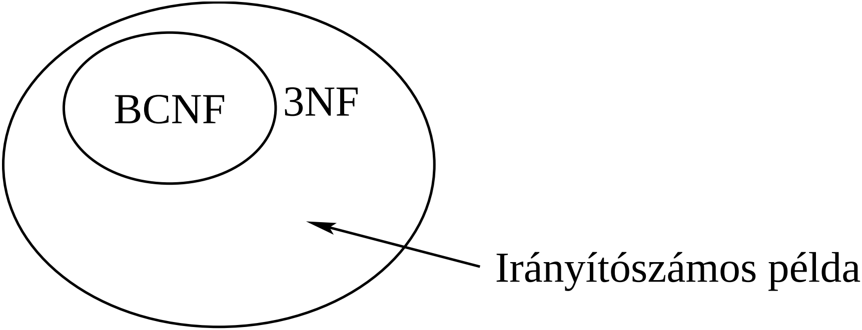
Kulcs helyett szuperkulccsal nem lenne sok értelme az elõbbi definíciónak, mert szuperkulcsban minden szerepel,.

Definíció

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

BCNF vs. 3NF

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



Felbontás 3NF-re

Tétel

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

Mindjárt látjuk, de előbb 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 $X \to A$ alakúak, ahol $A \notin X$ (A egy darab attribútum)*

(2) *G-bõl 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

Tetszõleges 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: $G = \{X_1 \to A_1, \ldots, X_k \to A_k\}$

Legyen X egy kulcs és tekintsük a
$\rho = (X, X_1A_1, \ldots, X_kA_k) = (R_0, R_1, \ldots, R_k)$ felbontást.

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

ρ függõségõrzõ

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

A tagok 3NF-ek ρ-ban

- R_0 3NF: R_0-ban nincs nemtriviális függés, mert különben X nem lenne kulcs, csak szuperkulcs $\Longrightarrow R_0$ még BCNF is, tehát 3NF is
- Többi R_i is 3NF: tegyük fel, hogy nem az, ekkor $\exists\ U \to B$ nemtriviális függés R_i-ben, hogy U nem szuperkulcs R_i-ben és B nem prímattribútum 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 $X_i \to A_i$ baloldala csökkenthetõ lenne G-ben U-ra. Ellentmondás, mert akkor G nem volt minimális fedés.
    - Ha $B \neq A_i \Longrightarrow B \in X_i$ és B nem prím R_i-ben $\Longrightarrow X_i$ nem kulcs R_i-ben (de szuperkulcs) $\Longrightarrow \exists\ Y \subsetneq X_i$ kulcs R_i-ben $\Longrightarrow Y \to A_i$ fennáll $\Longrightarrow X_i \to A_i$ baloldala csökkenthetõ G-ben Y-ra, megint csak ellentmondás.

ρ hûséges:

Higgyük el, nem bizonyítjuk.

Megjegyzések

- Elõfordulhat, hogy valamelyik X_iA_i már tartalmaz kulcsot. Ilyenkor a $\rho = \{X_1A_1, \ldots, X_kA_k\}$ 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* $X \to A$ *alakúak, ahol* $A \notin X$ *(A egy darab attribútum)*

(2) *G-bõl 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

Tetszõleges F-nek van minimális fedése.

Minimális fedés készítése F-hez

Bizonyítás: Algoritmust adunk, külön gondoskodunk minden pont teljesítésérõl.

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

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

(3) Ellenõrizni kell, hogy $X \to A$ baloldala minimális-e. X minden B elemére kiszámoljuk $Y := (X \setminus \{B\})^+(G)$-t. Ha $A \in Y$, akkor $X \to A$ helyett vegyük be $X \setminus \{B\} \to A$-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

$R = (A, B, C, D) \qquad F = \{AB \to CD;\ AC \to BD;\ C \to A;\ C \to B\}$
(1) $F' = \{AB \to C;\ AB \to D;\ AC \to B;\ AC \to D;\ C \to A;\ C \to B\}$
(2) $C \to B$ miatt $AC \to B$ elhagyható és $AB \to C$ és $AC \to D$ miatt $AB \to D$ elhagyható, de más nem, ezt végig lehet nézni.
$F'' = \{AB \to C;\ AC \to D;\ C \to A;\ C \to B\}$
(3) $C \to A$ miatt $AC \to D$ baloldaláról A elhagyható.
$F''' = \{AB \to C;\ C \to D;\ C \to A;\ C \to B\}$
Ez már minimális fedés.

A minimális fedés nem feltétlenül egyértelmû!

Példa: $R(A, B, C)$ $\quad F = \{AB \to C;\ A \to B;\ B \to A\}$ esetén jó minimális fedés lesz
$G_1 = \{B \to C;\ A \to B;\ B \to A\}$ és
$G_2 = \{A \to C;\ A \to B;\ B \to A\}$ is.

Példa 3NF-re bontásra

$R = (A, B, C, D, E)$ $\quad F = \{AE \rightarrow BC;\ AC \rightarrow D;\ CD \rightarrow BE;\ D \rightarrow E\}$
Ez nem 3NF, mert a kulcsok:
semelyik egyelemû halmaz nem kulcs (csak D lehetne, de az õ 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 prímattribútumok: 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 prím, szóval valamit kell

Példa tovább

Csináljunk hát egy 3NF-ekre való függõségõrzõ, 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 csökkenthetõ, 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 elsõben).