Adatbázisok elmélete 15. 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

3NF (emlékeztető)

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

3NF tulajdonság ellenőrzése

Hogyan tudjuk ellenőrizni, hogy egy séma 3NF-ben van-e?

Állítás. *≤ 2 attribútumos reláció mindig 3NF.*
Következmény. *Annak eldöntése, hogy egy séma 3NF-ben van-e, NP-teljes probléma.*

3NF tulajdonság ellenőrzése

Persze olyan algoritmus van, ami legrosszabb esetben exponenciális:

- Meghatározzuk az összes kulcsot.
- Meghatározzuk az összes prímattribútumot.
- Minden F-beli $X \to Y$ függésre nézzük meg:
  - ⋆ Igaz-e, hogy $Y \subseteq X$. Ha igen, a függés triviális. √
  - ⋆ Igaz-e, hogy X kulcs-e. Ha igen, nem sérti a feltételt. √
  - ⋆ Igaz-e, hogy Y-ben csak prímattribútumok vannak. Ha igen, nem sérti a feltételt. √

  Ha egyik sem, akkor van olyan függés, ami sérti a feltételt $\Longrightarrow$ nem 3NF

3NF felbontás

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: $Y \supsetneq X \implies (G \setminus \{X \to A\} \cup \{Y \to A\})^+ \subsetneq G^+$

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

Bizonyítás: Algoritmust adunk, külön gondoskodunk minden pont teljesítéséről.

(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, különben nem.

(3) Ellenőrizni 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 $X - \{B\} \to A$-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

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

(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 egyértelmű!

Példa: $R(A, B, C) \qquad F = \{AB \to C;\ A \to B;\ B \to A\}$ 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.

Bizonyítás

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

Bizonyítás: 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.

Megjegyzés: Előfordulhat, hogy valamelyik X_iA_i már tartalmaz kulcsot. Ilyenkor a $\rho = \{X_1A_1, \ldots, X_kA_k\}$ is jó felbontás már.

Megjegyzés: 2NF már nem érdekes, 1NF kicsit érdekes, de nem foglalkozunk vele.

Példa: 3NF-re bontás

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

Ez nem 3NF, mert a kulcsok:
semelyik egyelemű halmaz nem kulcs (csak D lehetne, de az ő lezártja csak DE), viszont kételeműek 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 egyelemű se volt kulcs.
Más kulcs nincs is, mert ha lenne legalább háromelemű 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.

Példa: 3NF-re bontás (folyt.)

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

Tehát a $CD \to B$ függés rossz a 3NF szempontjából, mert CD nem szuperkulcs és B nem prím.

Csináljunk hát egy 3NF-ekre való függőségőrző, hűséges felbontást.

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

(2) $AE \to C, AC \to D, CD \to B$ miatt $AE \to B$ elhagyható és $D \to E$ miatt $CD \to E$ is elhagyható, de más nem, ezt végig lehet nézni (mert például AE-nek a maradék függésekre vett lezártjában nincsen benne C).
$F'' = \{AE \to C;\ AC \to D;\ CD \to B;\ D \to E\}$

(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).

Többértékű függés

A legfontosabb a funkcionális függés, de vannak másféle függések is.

Motiváló példa: R(Név, Tantárgy, Gyereknév)

Név	Tantárgy	Gyereknév
Katona	Algel	Dani
Katona	Adatbázis	Lilla
Katona	Algel	Lilla
Katona	Adatbázis	Dani

Ez BCNF, de mégis redundáns, mert ha valamelyik tárgynál szerepel egy gyereknév, akkor az összes többinél is szerepelnie kell. (Pl. beszúrni nehéz, mert amikor egy sort beszúrok, figyelni kell arra, hogy egy másikat is beszúrjak.)
Jobb lenne tárolni (Név, Tantárgy) és (Név, Gyereknév) felbontásban.

Ok: a Tantárgy és a Gyereknév független (minden kombinációban előfordulnak) $\Longrightarrow$ ha látjuk az első két sort, tudjuk, hogy a másik kettő is ott van.

Többértékű függés

Definíció. *Az X attribútumhalmaztól* ***többértékűen függ*** *az Y attribútumhalmaz az r relációban (jele: $X \twoheadrightarrow Y$), ha tetszőleges $t_1, t_2 \in r$ sorokra, melyekre $t_1[X] = t_2[X]$, létezik $t_3, t_4 \in r$, melyekre*

- $t_3[XY] = t_1[XY]$
- $t_3[R \setminus XY] = t_2[R \setminus XY]$
- $t_4[XY] = t_2[XY]$
- $t_4[R \setminus XY] = t_1[R \setminus XY]$

	$\overbrace{\ldots\ldots\ldots\ldots}^{X}$	$\overbrace{\ldots\ldots\ldots\ldots}^{Y}$	$\overbrace{\ldots\ldots\ldots\ldots}^{R \setminus XY}$
t_1	$AAAAAAA$	$BBBBBBB$	$CCCCCCC$
t_2	$AAAAAAA$	$DDDDDDD$	$EEEEEEE$
	$\vdots$	$\vdots$	$\vdots$
t_3	$AAAAAAA$	$BBBBBBB$	$EEEEEEE$
t_4	$AAAAAAA$	$DDDDDDD$	$CCCCCCC$

Megjegyzés: A funkcionális függőség **egyenlőséggeneráló**. Ha két dolog egyenlő, akkor másik két dolog is egyenlő lesz. A többértékű függőség **sorgeneráló**. Ha van két sor ami valahol egyenlő, akkor vannak más sorok is.
Az előbbi példában: Név$\twoheadrightarrow$ Tantárgy, Név$\twoheadrightarrow$ Gyereknév

Többértékű függések levezetése

Definíció. ***Triviális többértékű függések*** *(amik mindig igazak):*

- $Y \subseteq X \implies X \twoheadrightarrow Y$, *mert* $t_3 = t_2$ *és* $t_4 = t_1$ *jó lesz.*
- $XY = R \implies X \twoheadrightarrow Y$, *mert* $t_3 = t_1$ *és* $t_4 = t_2$ *jó lesz.*

Ezentúl a többértékű függések is a séma részei lesznek és definiálhatjuk a levezethetőséget ($\vdash$) és a logikai következményt ($\models$) úgy, hogy funkcionális függőségek és többértékű függőségek is vannak F-ben.

Logikai következmény: egy F (funkcionális és többértékű függéseket is tartalmazó) függéshalmaznak logikai következménye egy (funkcionális vagy többértékű) függés, ha minden olyan relációban, amiben F minden függése fennáll, fenn kell hogy álljon a mondott függés is.

Levezetés: Armstrong-axiómák (a funkcionális függésekre) és 5 új axióma, amiben $\rightarrow$ és $\twoheadrightarrow$ is van. Amilyen függés ezekkel előáll F-ből, arra mondjuk, hogy levezethető.

Hasonló elmélet, mint $\rightarrow$-nél $\implies$ belátható, hogy $\vdash \sim \models$ itt is igaz lesz.

Többértékű levezetési szabályok

Két fontos új szabály

- $X \rightarrow Y \vdash X \twoheadrightarrow Y$, mert $t_3 = t_2$ és $t_4 = t_1$ jó lesz.
- $X \twoheadrightarrow Y \vdash X \twoheadrightarrow R \setminus XY$, mert $t_3' = t_4$ és $t_4' = t_3$ jó lesz.
- De pl. $X \twoheadrightarrow AB \nvdash X \twoheadrightarrow A$, nem szétvágható. (Sok minden máshogy van a többértékű függéseknél.)

Tétel. *Legyen* $\rho = (R_1, R_2)$ *az* (R, F) *séma felbontása, ahol* F *most funkcionális és többértékű függéseket is tartalmaz.* ρ *akkor és csak akkor hűséges felbontás, ha* $R_1 \cap R_2 \twoheadrightarrow R_2 \setminus R_1$.

Megjegyzés: Nem kell a „vagy $R_1 \cap R_2 \twoheadrightarrow R_1 \setminus R_2$" a fenti 2. szabály miatt, mert ha $R_1 \cap R_2 \twoheadrightarrow R_1 \setminus R_2$ igaz, akkor $R_1 \cap R_2 \twoheadrightarrow R \setminus (R_1 \setminus R_2)$ is igaz, ebből meg már következik $R_1 \cap R_2 \twoheadrightarrow R_2 \setminus R_1$.

a tétel bizonyítása hasonló, mint a funkcionális függésnél, de nem bizonyítjuk.

4NF

Cél: olyan normálforma, amiben többértékű függés miatt sincs redundancia.

BCNF mintájára:
Definíció. *Az* (R, F) *séma* ***4NF*** *(**negyedik normálformájú**), ha tetszőleges nemtriviális* $X \twoheadrightarrow Y \in F^+$ *esetén* X *szuperkulcs (a szuperkulcsot a régi értelemben, csak funkcionális függőségekkel definiálva).*

Következmény. *Ha egy séma 4NF, akkor BCNF is.*

Bizonyítás: Indirekt tegyük fel, hogy létezik olyan $X \rightarrow A \in F^+$ nemtriviális függés, ahol X nem szuperkulcs. $\implies$ Ekkor $\lightning$, amiatt, hogy $X \rightarrow A$-ból következik, hogy $X \twoheadrightarrow A$.

Megjegyzések:

- Ha F-ben csak funkcionális függőségek vannak, akkor 4NF=BCNF
- 2 attribútumos reláció mindig 4NF, hiszen nincs nemtriviális többértékű függés, azt meg már láttuk, hogy ha csak funkcionális függések vannak, akkor a BCNF-ség rendben van kétattribútumos relációnál.
- Van olyan reláció, ami BCNF, de nem 4NF (a korábbi gyerekes példa, mert ott a Név nem szuperkulcs)