

Adatbázisok elmélete 13. 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

2004

Hűséges felbontás

Tehát az a kérdés, hogy mik azok a felbontásai egy (R, F) sémának, amik esetén tetszőleges (R, F) -re illeszkedő r relációra $r = m_\rho(r)$

Definíció. Adott (R, F) . Ennek ρ felbontása **hűséges (veszteségmentes, lossless)**, ha minden (R, F) -re illeszkedő r relációra $r = m_\rho(r)$.

Példa: Legyen (R, F) a következő: $R(A, B, C)$, $F = \{C \rightarrow A\}$ és legyen r az alábbi reláció.

r	A	B	C
a	c	e	
a	d	f	
b	c	g	
b	d	h	

s	A	B
a	c	
a	d	
b	c	
b	d	

t	B	C
c	e	
d	f	
c	g	
d	h	

s ⋈ t	A	B	C
a	c	e	
a	c	g	
a	d	f	
a	d	h	
b	c	e	
b	c	g	
b	d	f	
b	d	h	

Ez a példa mutatja, hogy $r \neq s(A, B) \bowtie t(B, C)$, azaz ez a felbontás nem hűséges.

De $r = s'(A, C) \bowtie t'(B, C)$, majd látjuk.

Felbontások (emlékeztető)

Cél: Adott (R, F) sémából anomáliát nem tartalmazó olyan felbontás előállítása, amiből ugyanaz az információ nyerhető, mint az eredetiből.

Definíció. $\rho = (R_1, \dots, R_k)$ az (R, F) séma felbontása, ha $R_i \subseteq R$ és $\bigcup_{i=1}^k R_i = R$. Ha r egy (R, F) sémára illeszkedő reláció, akkor legyen $r_i = \pi_{R_i}(r)$ és

$$m_\rho(r) := r_1 \bowtie r_2 \bowtie \dots \bowtie r_k$$

(Megj.: \bowtie asszociatív, így nem kell a zárójelezéssel vesződni)

Kérdés: mikor nyerhető vissza az infó a felbontásból? Mi általában r és $m_\rho(r)$ viszonya?

Tétel.

- (i) $r \subseteq m_\rho(r)$
- (ii) $r_i = \pi_{R_i}(m_\rho(r))$
- (iii) $m_\rho(m_\rho(r)) = m_\rho(r)$

Következmény: ha $r \neq m_\rho(r)$, akkor sehogyan se lehet visszahozni r -t a vetületekből.

Hűséges felbontás két részre

Hogyan biztosíthatja F , hogy a felbontás hűséges legyen?

Tétel. Az (R, F) séma $\rho = (R_1, R_2)$ felbontása hűséges \iff vagy

- (a) $F \models R_1 \cap R_2 \rightarrow R_1 \setminus R_2$, vagy
- (b) $F \models R_1 \cap R_2 \rightarrow R_2 \setminus R_1$.

Példa: $R(\text{TNÉV}, \text{TERMELŐ}, \text{ÁR}, \text{CÍM})$
 $F = \{\text{TERMELŐ} \rightarrow \text{CÍM}; \text{TNÉV}, \text{TERMELŐ} \rightarrow \text{ÁR}\}$

$\rho = (\text{TERMELŐ}, \text{CÍM}; \text{TNÉV}, \text{TERMELŐ}, \text{ÁR})$
 $R_1 \cap R_2 = \{\text{TERMELŐ}\}$, $R_1 \setminus R_2 = \{\text{CÍM}\}$, $R_2 \setminus R_1 = \{\text{TNÉV}, \text{ÁR}\}$
 $\implies (\text{TERMELŐ})^+(F) = \{\text{TERMELŐ}, \text{CÍM}\} \supseteq R_1 \setminus R_2 \implies$ hűséges \checkmark

$\rho = (\text{TNÉV}, \text{TERMELŐ}; \text{TNÉV}, \text{CÍM}, \text{ÁR})$
 $R_1 \cap R_2 = \{\text{TNÉV}\}$, $R_1 \setminus R_2 = \{\text{TERMELŐ}\}$, $R_2 \setminus R_1 = \{\text{CÍM}, \text{ÁR}\}$
 $\implies (\text{TNÉV})^+(F)$ nem tartalmazza egyiket sem \implies nem hűséges \checkmark

Bizonyítás

Tétel. Az (R, F) séma $\rho = (R_1, R_2)$ felbontása hűségese \iff vagy

(a) $F \models R_1 \cap R_2 \rightarrow R_1 \setminus R_2$, vagy

(b) $F \models R_1 \cap R_2 \rightarrow R_2 \setminus R_1$.

Bizonyítás: Tegyük fel, hogy $F \models R_1 \cap R_2 \rightarrow R_1 \setminus R_2$, belátjuk, hogy a felbontás hűséges. (Ha a másik igaz, ugyanígy.)

Legyen r egy tetszőleges reláció, $s = m_\rho(r)$. Elég belátni, hogy $s \subseteq r$, hiszen $r \subseteq s$ mindig igaz. Azaz, lássuk be, hogy ha t sora s -nek, akkor r -nek is.

Ha t sora s -nek, akkor $\exists u_1, u_2$ sorai r -nek, hogy $t[R_1] = u_1[R_1]$ és $t[R_2] = u_2[R_2]$.

$$\implies u_1[R_1 \cap R_2] = t[R_1 \cap R_2] = u_2[R_1 \cap R_2]$$

de ha két sor megegyezik a metszeten, akkor a feltétel miatt $R_1 \setminus R_2$ -n is \implies egyeznek az egész R_1 -en $\implies u_2$ és t egyeznek R_1 -en.

$$\implies t = u_2, \text{ hiszen } R_1\text{-en a fenti miatt, } R_2\text{-n a feltevés miatt egyeznek. } \checkmark$$

Másik irány: Belátjuk, hogy ha $R_1 \setminus R_2 \not\subseteq (R_1 \cap R_2)^+(F)$ és $R_2 \setminus R_1 \not\subseteq (R_1 \cap R_2)^+(F)$, akkor ρ nem hűséges.

Legyen r a következő reláció:

r	$\overbrace{\hspace{12em}}^{R_1}$												
	$\overbrace{\hspace{6em}}^{R_2}$						$\overbrace{\hspace{6em}}^{(R_1 \cap R_2)^+(F)}$						
$\overbrace{\hspace{12em}}^{R_1 \cap R_2}$													
t_1	0	0	0	1	1	1	1	1	1	1	1	1
t_2	1	1	1	1	1	1	1	1	1	0	0	0

A feltétel miatt a két szélső rész **nem üres**, ott nem egyezik meg a két sor.

r -ben igazak F függőségei (mint a teljességi tételnél).

Viszont $m_\rho(r) \not\supseteq r$, hiszen $m_\rho(r)$ -ben a csupa 1 sor is benne van. ζ

Hűségesség ellenőrzése általában

Adott (R, F) és $\rho = (R_1, \dots, R_k)$, ahol $R = A_1, \dots, A_n$.

Hogyan tudjuk eldönteni, hogy hűségese-e a felbontás?

Készítünk egy $k \times n$ -es táblázatot:

	A_1	...	A_j	...	$A_{j'}$...	A_n
R_1							
\vdots							
R_i			a_j		$b_{ij'}$		
\vdots							
R_k							

- Kezdetben az (i, j) helyre a_j -t írunk, ha $A_j \in R_i$ és b_{ij} -t, ha $A_j \notin R_i$.
- Veszünk egy tetszőleges $X \rightarrow Y \in F$ függést.
Ha két sor megegyezik X -en, akkor egyenlővé tesszük Y -on is az alábbi módon:
 - ★ Ha valahol a_j és b_{ij} van, akkor a b_{ij} -t a_j -ra cseréljük.
 - ★ Ha b_{kj} és b_{lj} van, akkor az egyiket átírjuk a másikra.
- Ezt minden függésre megcsináljuk tetszőleges sorrendben, szükség esetén többször is.

Jó a módszer?

Tétel. ρ pontosan akkor hűséges ha a végén lesz csupa a sor.

Nem bizonyítjuk.

Példa a táblázatos tesztre

$R(ABCD)$ $F = \{A \rightarrow C; C \rightarrow B; B \rightarrow D\}$ $\rho = (AB, BC, ACD)$

	A	B	C	D
R_1	a_1	a_2	$b_{13} \rightarrow a_3^*$	b_{14}
R_2	b_{21}	a_2	a_3	b_{24}
R_3	a_1	$b_{32} \rightarrow a_2^{**}$	a_3	a_4

* $A \rightarrow C$ miatt

** $C \rightarrow B$ miatt

Lett csupa a sor \Rightarrow hűségese felbontás

Hűségese felbontás

Tétel. Adott (R, F) , $\rho = (R_1, \dots, R_k)$ az R hűségese felbontása és $\sigma = (S_1, \dots, S_m)$ az R_1 hűségese felbontása (azaz R_1 -et tovább bontjuk). Ekkor $\tau = (S_1, \dots, S_m, R_2, \dots, R_k)$ hűségese felbontása R -nek.

Bizonyítás: Legyen r egy R -re illeszkedő reláció és ennek R_1 -re eső vetülete legyen r_1 .

Tovább bontva r_1 -et σ szerint kapjuk az s_1, s_2, \dots, s_m vetületeket. Mivel σ hűségese, ezért

$$s_1 \bowtie s_2 \bowtie \dots \bowtie s_m = m_\sigma(r_1) = r_1.$$

Mivel ρ is hűségese, ezért $r = m_\rho(r) = r_1 \bowtie r_2 \bowtie \dots \bowtie r_k$. Ebbe beírva r_1 helyére a σ hűségességből kapott egyenlőséget, kapjuk, hogy $r = s_1 \bowtie s_2 \bowtie \dots \bowtie s_m \bowtie r_2 \bowtie \dots \bowtie r_k = m_\tau(r)$, azaz τ is hűségese. Itt persze használtuk \bowtie asszociativitását.

Tétel. Ha ρ hűségese és $\sigma \supseteq \rho$ (σ -ban több komponens van), akkor σ is hűségese.

Bizonyítás: $r \subseteq m_\sigma(r) \subseteq m_\rho(r) = r$

A középső tartalmazás azért igaz, mert a keresztszorzatból szigorúbb feltételek szerint válogatunk.

$$\Rightarrow r = m_\sigma(r) \quad \checkmark$$

Normálformák

Definíció. Egy $X \rightarrow Y$ 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 $X \rightarrow A \in F^+$ függés esetén X superkulcs.

Azaz csak olyan függések vannak, hogy a superkulcs 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 $X \rightarrow Y \in F^+$; $Y \rightarrow X$; $Y \rightarrow A \in F^+$ és $A \notin Y$. (Nincs tranzitív függés kulcstól.)

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

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

Miért jó a BCNF séma?

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

Állítás. ≤ 2 attribútumos reláció mindig BCNF.

Bizonyítás: Ha $A \rightarrow B \Rightarrow A$ kulcs. Ha $B \rightarrow A \Rightarrow B$ kulcs.

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

DE:

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

Bizonyítás: Ha (R, F) nem BCNF, akkor van $U \rightarrow B \in F^+$, hogy U nem superkulcs és $B \notin U$. $\Rightarrow B \in U^+(F) \Rightarrow U \subseteq U^+(F)$

Az algoritmus, ami $U^+(F)$ -et számolja, el tud indulni $\Rightarrow \exists V \rightarrow W \in F$, melyre $V \subseteq U$, $W \notin U$ $\Rightarrow V \rightarrow W$ jó lesz $X \rightarrow Y$ -nak.

Ugyanis V nem superkulcs, hiszen $V \subseteq U$ és U nem superkulcs.

$W \notin U \Rightarrow \exists A \in W \setminus U \subseteq W \setminus V$, így $V \rightarrow W$ nem triviális.

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