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

2005

$X^+(F)$ kiszámítása

Algoritmus:

$X_0 = X$,

⋮

$X_i = \dots$,

$X_{i+1} = X_i \cup \{A \in R \mid$ van olyan $U \to V \in F$, hogy $U \subseteq X_i$ és $A \in V\}$,

⋮

$X^+(F) = X_{\text{utolsó}}$, (amikor már nem nő)

Állítás. $X_{utolsó} \subseteq X^+(F)$ *(azaz, a fontos lemma miatt* $F \vdash X \to X_{utolsó}$*)*

Bizonyítás: Indukcióval i-re belátjuk, hogy $F \vdash X \to X_i$, innen már következik az állítás.

Bizonyítás (folyt.)

$i = 0$: $F \vdash X \to X_0 = X$, reflexivitás.

$i \rightsquigarrow i + 1$

1. $U \to V \in F,\ U \subseteq X_i,\ A \in V$
2. $X_i \to U$, reflexivitás
3. $X \to U$, tranzitivitás + indukciós felt. $F \vdash X \to X_i$ + (2. sor)
4. $X \to V$, tranzitivitás (1. és 3. sor)
5. $V \to A$, reflexivitás, (1. sor)
6. $X \to A$, tranzitivitás, (4. és 5. sor)
7. $F \vdash X \to A$, minden $A \in X_{i+1}$
8. $F \vdash X \to X_{i+1}$

$\Longrightarrow F \vdash X \to X_{\text{utolsó}} \Longrightarrow$ (Lemma) $X_{\text{utolsó}} \subseteq X^+(F)$. $\surd$

Példa:

$R(A, B, C, D)$, $F = \{A \to B,\ B \to C,\ BC \to D\}$, $A^+(F) = ?$

$X_0 = \{A\}$, $X_1 = \{A, B\}$, $X_2 = \{A, B, C\}$, $X_3 = \{A, B, C, D\} = X_{\text{utolsó}}$

Bizonyítás másik iránya

Állítás. $X^+(F) \subseteq X_{utolsó}$

Bizonyítás: Tekintsük az alábbi kétsoros r relációt: a két sor $X_{\text{utolsó}}$-n egyenlő, azon kívül eltérnek.

r							$X_{\text{utolsó}}$						
	A_1	…	…				X				…	…	A_n
t_1	1	1	1	1	1	1	1 1	1	1	1	1	1	1
t_2	0	0	0	1	1	1	1 1	1	1	1	0	0	0

r-ben egyrészt minden F-beli függés teljesül. (Bizonyítás, hasonlóan, mint a teljességi tételnél csak most az kell, hogy ha egy $W \to S$ függés esetén $W \subseteq X_{\text{utolsó}}$, akkor $S \subseteq X_{\text{utolsó}}$ is igaz, az algoritmus működése miatt.)
r segítségével azt látjuk be, hogy ha $A \notin X_{\text{utolsó}}$, akkor $A \notin X^+(F)$, ami éppen a kívánt állítás.

Ha $A \notin X_{\text{utolsó}} \Longrightarrow r$ olyan reláció, amiben $X \not\twoheadrightarrow A$
ezért $F \nvDash X \to A$, hiszen r-ben F minden függése teljesül.
$\Longrightarrow$ (az igazság tétel miatt) $F \nvdash X \to A \Longrightarrow$ (fontos lemma) $A \notin X^+(F)$ $\surd$

Következmények

Következmény. $X^+(F) = X_{\text{utolsó}}$, *azaz tényleg jó az algoritmus.*

Következmény. *Adott X-ről el lehet dönteni, hogy (szuper)kulcs-e.*

Megnézzük, hogy $X^+(F) = R$ igaz-e. Ha igen, akkor szuperkulcs. Ha minden $X - A$-ra már nem szuperkulcsot kapunk, akkor X kulcs.
Gyorsan implementálható.

Felbontások

Cél: *Adott (R, F) sémából anomáliát nem tartalmazó olyan felbontás előállítása, amiből ugyanaz az információ nyerhető, mint az eredetiből.*
Definíció. $\rho = (R_1, \ldots, R_k)$ *az* (R, F) *séma felbontása, ha* $R_i \subseteq R$ *és* $\cup_{i=1}^k R_i = R$.
Ha r egy (R, F) sémára illeszkedő reláció, akkor legyen $r_i = \pi_{R_i}(r)$ *és*

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

(*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) $m_\rho\left(m_\rho(r)\right) = m_\rho(r)$

Bizonyítás

Bizonyítás: Ha t egy sor, akkor $\pi_{R_i}(t)$ helyett $t[R_i]$-t írunk.

(i) $r \subseteq m_\rho(r)$:
Ha t egy sor r-ben, akkor t minden vetülete benne van a megfelelő $t[R_i]$-ben, ezek össze is illenek, így $m_\rho(r)$-ben is szerepelni fog t.

(ii) $r_i = \pi_{R_i}(m_\rho(r))$:
$r \subseteq m_\rho(r) \implies r_i = \pi_{R_i}(r) \subseteq \pi_{R_i}\left(m_\rho(r)\right)$.
Ha $t \in m_\rho(r)$, akkor ez természetes illesztéssel jött létre, r_i-beli sorokból, így levetítve R_i-re épp r_i egy sorát kapjuk.

(iii) $m_\rho\left(m_\rho(r)\right) = m_\rho(r)$:
$m_\rho(r) = \bowtie_{i=1}^k r_i = \bowtie_{i=1}^k \pi_{R_i}(r)$
$m_\rho\left(m_\rho(r)\right) = \bowtie_{i=1}^k \pi_{R_i}\left(m_\rho(r)\right) \stackrel{(ii)}{=} \bowtie_{i=1}^k r_i = m_\rho(r)$

Megjegyzés: (i) szerint a szétszedés és összerakás után vagy pont r-t kapom meg, vagy többet kapok, kevesebb sor nem lehet. Ha $r \neq m_\rho(r)$, akkor ez nem egy túl hasznos felbontás.
De ennél több is igaz: ebben az esetben teljesen reménytelen a felbontásból visszaszerezni r-t: mivel (ii) szerint r és $m_\rho(r)$ (függőleges) vetületei ugyanazok, ezért ha $r \neq m_\rho(r)$, akkor *van két olyan reláció (r és $m_\rho(r)$), aminek a vetületei ugyanazok $\implies$ a vetületekből nem lehet visszaállítani r-et (nem lehet eldönteni, hogy r vagy $m_\rho(r)$ volt).*

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

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

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) $F \models R_1 \cap R_2 \to R_1 \setminus R_2$, *vagy*

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

Példa: R(TNÉV, TERMELŐ, ÁR, CÍM)

F = {TERMELŐ → CÍM; TNÉV, TERMELŐ → ÁR}

ρ = (TERMELŐ, CÍM ; TNÉV, TERMELŐ, ÁR)

$R_1 \cap R_2$ = {TERMELŐ}, $R_1 \setminus R_2$ = {CÍM}, $R_2 \setminus R_1$ = {TNÉV, ÁR}

$\Longrightarrow$ (TERMELŐ)$^+(F)$ = {TERMELŐ, CÍM} $\supseteq R_1 \setminus R_2$ $\Longrightarrow$ hűséges $\surd$

ρ = (TNÉV, TERMELŐ; TNÉV, CÍM, ÁR)

$R_1 \cap R_2$ = {TNÉV}, $R_1 \setminus R_2$ = {TERMELŐ}, $R_2 \setminus R_1$ = {CÍM, ÁR}

$\Longrightarrow$ (TNÉV)$^+(F)$ nem tartalmazza egyiket sem $\Longrightarrow$ nem hűséges ↯

Bizonyítás

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

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

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

Bizonyítás: Tegyük fel, hogy $F \models R_1 \cap R_2 \to R_1 \setminus R_2$, 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 $t[R_2] = u_2[R_2]$.

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

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

$\Longrightarrow t = u_2$, hiszen R_1-en a fenti miatt, R_2-n a feltevés miatt egyeznek. $\surd$

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_1, R_2, $(R_1 \cap R_2)^+(F)$, $R_1 \cap R_2$

r													
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 $m_\rho(r) \supsetneq r$, hiszen $m_\rho(r)$-ben a csupa $\mathbf{1}$ sor is benne van. ↯

Hűségesség ellenőrzése általában

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

Hogyan tudjuk eldönteni, hogy hűséges-e a felbontás?

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

	A_1	$\ldots$	A_j	$\ldots$	$A_{j'}$	$\ldots$	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 $X \to Y \in F$ 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

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

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

* $A \to C$ miatt

** $C \to B$ miatt

Lett csupa a sor $\Longrightarrow$ hűséges felbontás

Hűséges felbontás

Tétel. *Adott (R, F), $\rho = (R_1, \ldots, R_k)$ az R hűséges felbontása és $\sigma = (S_1, \ldots, S_m)$ az R_1 hűséges felbontása (azaz R_1-et tovább bontjuk). Ekkor $\tau = (S_1, \ldots, S_m, R_2, \ldots, R_k)$ hűséges 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 $s_1, s_2, \ldots, s_m$ vetületeket. Mivel σ hűséges, ezért $s_1 \bowtie s_2 \bowtie \cdots s_m = m_\sigma(r_1) = r_1$.
Mivel ρ is hűséges, ezért $r = m_\rho(r) = r_1 \bowtie r_2 \bowtie \cdots r_k$. Ebbe beírva r_1 helyére a σ hűségességéből kapott egyenlőséget, kapjuk, hogy $r = s_1 \bowtie s_2 \bowtie \cdots s_m \bowtie r_2 \bowtie \cdots r_k = m_\tau(r)$, azaz τ is hűséges. Itt persze használtuk $\bowtie$ asszociativitását.

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

Bizonyítás: $r \subseteq m_\sigma(r) \subseteq m_\rho(r) = r$
A középső tartalmazás azért igaz, mert a keresztszorzatból szigorúbb feltételek szerint válogatunk.
$\Longrightarrow r = m_\sigma(r)$ $\surd$

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.