Adatbázisok elmélete 15. előadás

Csima Judit
Budapesti Műszaki és Gazdaságtudományi Egyetem
Számítástudományi Tsz.
I. B. 136/b
csima@cs.bme.hu

2003. Április 2.

Teljességi tétel

Tétel (Teljességi tétel). *Ha $F \models X \to Y$, akkor $F \vdash X \to Y$.*

Bizonyítás: Tegyük fel indirekt, hogy van olyan $X \to Y$ függés és F függéshalmaz, hogy $X \to Y$ nem vezethető le F-ből ($F \nvdash X \to Y$), noha logikai következménye neki ($F \models X \to Y$).
Ez utóbbi azt jelenti, hogy minden olyan relációban, amiben F függőségei teljesülnek, ha X-en megegyezik két sor, akkor azok megegyeznek Y-on is.
Úgy jutunk ellentmondásra, hogy konstruálunk egy olyan r relációt, ahol F függőségei teljesülnek, de $X \not\to Y$, ami ellentmond $F \models X \to Y$-nak.

r							$X^+(F)$						
							X						
	A_1	…	…				....				…	…	A_n
t_1	1	1	1	1	1	1	....	1	1	1	1	1	1
t_2	0	0	0	1	1	1	....	1	1	1	0	0	0

Tehát r két soros reláció, $X^+(F)$-en megegyezik a két sor, a többin különbözik.
Állítás. *r-ben teljesülnek F függései.*

Bizonyítás: Legyen $U \to V \in F$.
Ha $U \nsubseteq X^+(F) \implies U \to V$ igaz, hiszen nincs olyan két sor, ami U-n megegyezik.

Ha $U \subseteq X^+(F)$, akkor lemma miatt $F \vdash X \to U$.
Tranzitivitás miatt $F \vdash X \to V$. Lemma $\implies V \subseteq X^+(F)$, V-n megegyezik a két sor. $\surd$

Állítás. *r-ben nem igaz $X \to Y$.*

Bizonyítás: Mivel $F \nvdash X \to Y$, ezért a fontos lemma miatt $Y \nsubseteq X^+(F)$, azaz Y kilóg $X^+(F)$-ből, abból a részből, ahol a két sor egyenlő. Vagyis a két sor egyenlő X-en, de nem egyenlő Y-on, így $X \to Y$ nem igaz r-ben.

Tehát r tényleg olyan, hogy benne F minden függése fennáll, de $X \to Y$ nem, ami bizonyítja, hogy $F \not\models X \to Y$.

Következmény. *$\vdash$ és $\models$ felcserélhető.*

Kulcs

Definíció. *$X \subseteq R$ **szuperkulcsa** az (R, F) sémának, ha $F \vdash X \to R$. Másképpen, ha $R = X^+(F)$.*
*$X \subseteq R$ **kulcsa** az (R, F) sémának, ha szuperkulcs és nincs olyan valódi részhalmaza, ami szuperkulcs.*

Példa:
$F = \{$TERMELŐ, TERMÉKNÉV $\to$ ÁR; TERMELŐ $\to$ CÍM$\}$
$X =$ TERMELŐ, TERMÉKNÉV
$\implies X^+(F) =$ TERMELŐ, TERMÉKNÉV, ÁR, CÍM
TERMELŐ$^+(F) =$ TERMELŐ, CÍM
TERMÉKNÉV$^+(F) =$ TERMÉKNÉV
$\implies X$ kulcs

$X^+(F)$ kiszámítása

Algoritmus:

$X_0 = X$,

$\vdots$

$X_i = \dots$,

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

$\vdots$

$X^+(F) = X_{\text{utolsó}}$, (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.
$i = 0$: $F \vdash X \to X_0 = X$, 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) $X_{\text{utolsó}} \subseteq X^+(F)$.

Példa:

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

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

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				$X_{\text{utolsó}}$									
							X						
	A_1	…	…				....				…	…	A_n
t_1	1	1	1	1	1	1	....	1	1	1	1	1	1
t_2	0	0	0	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 $F \not\models X \to A$, hiszen r-ben F minden függése teljesül.

$\Longrightarrow$ (az igazság tétel miatt) $F \nvdash X \to A \Longrightarrow$ (fontos lemma) $A \notin X^+(F)$ $\surd$

Következmény. *$X^+(F) = X_{utolsó}$, 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 (lásd mintavizsga utolsó példája).

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

(i) 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) $r \subseteq m_\rho(r) \implies r_i = \pi_{R_i}(r) \subseteq \pi_{R_i}\left(m_\rho(r)\right)$.
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) $m_\rho(r) = \bowtie_{i=1}^k r_i = \bowtie_{i=1}^k \pi_{R_i}(r)$
$m_\rho\left(m_\rho(r)\right) = \bowtie_{i=1}^k \pi_{R_i}\left(m_\rho(r)\right) \overset{(ii)}{=} \bowtie_{i=1}^k r_i = m_\rho(r)$

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

$s \bowtie t$	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.