

Adatbázisok zárthelyi dolgozat – 2003. 11. 21. – A csoport – megoldás

Darvas Dániel (darvas@db.bme.hu), Szárnyas Gábor (szarnyas@db.bme.hu)

A normál betűvel szedett részek teljes értékű megoldásokat közölnek, de természetesen más érvelés is tökéletes lehet. Az apró betűvel szedett részletes magyarázatok megadása a *maximális pontszám eléréséhez sem* szükséges.

1. feladat

Az R és S sémák közös attribútuma B , ennek mentén történik a természetes illesztés. Az illesztésként kapott reláció miatt az r relációnak mindenképpen van (a,b) és (f,c) sora, az s relációnak mindenképpen van (b,a) és (c,g) sora. Így a két reláció az alábbi elemeket biztosan tartalmazza:

$r(R)$		$s(S)$	
A	B	B	C
a	b	b	a
f	c	c	g

Az uniójuk tehát biztosan tartalmazni fogja az (a,b) , (f,c) , (b,a) és (c,g) elemeket. Ahhoz, hogy az uniónak a (d,e) is eleme lehessen, a (d,e) sor lehet akár r , akár s vagy akár mindkettő eleme is, egyik esetben sem kerül bele az illesztésként kapott relációba. Ezen kívül mindkét relációnak tetszőlegesen sok olyan további sora lehet, amelyben a B attribútum értéke nem b vagy c , illetve a (d,e) sorral (sorokkal) sem illeszthető.

Részletesebb magyarázat. Mivel a (d,e) sor szerepel $r \cup s$ -ben, ezért nyilván legalább az egyik relációban is szerepelnie kell. Mindegy azonban, hogy melyiknek eleme (akár mindkettőnek is lehet), a természetes illesztésben ez a sor nem fog megjelenni, ha a relációknak csak a fenti sorai vannak ezen kívül. Ha (d,e) az $r(R)$ reláció eleme, akkor ahhoz, hogy megjelenjen $r \bowtie s$ -ben, szükséges lenne egy $(e,*)$ sor s relációban (a $*$ itt egy tetszőleges elemet jelöl).

Vegyük észre továbbá, hogy a két reláció egyértelműen nem határozható meg a feladatban megadott adatok alapján. Például a fenti sorok mellett s -hez hozzávehető lenne egy (x,y) sor, ez sem jelenne meg a természetes illesztésben, amennyiben r -ben nem található egy $(*,x)$ sor.

2. feladat

A séma kulcsai: AB és AC , tehát A, B és C elsődleges, D másodlagos attribútum.

A séma 1NF, mert minden attribútum atomi. A séma nem 2NF, mert a D másodlagos attribútum részben függ az AB kulcstól a $B \rightarrow D$ függőség miatt. Tehát a séma legmagasabb normál formája az 1NF.

Részletesebb magyarázat. Az A attribútum csak függőség bal oldalán szerepel érdemileg, ezért mindenképpen eleme a kulcsnak. (Bár van F -ben egy $AB \rightarrow AC$ függőség, ennek jobb oldaláról A elhagyható az $AB \rightarrow A$ triviális függőség miatt.) Az A attribútum önmagában nem kulcs, mert nem határozza meg a teljes sémát. A lehetséges kételemű kulcsokat megvizsgálva azt kapjuk, hogy csak AB és AC kulcs.

$$(AB)^+(F) = ABDC = R, (AC)^+(F) = ACBD = R, (AD)^+(F) = AD \neq R$$

Szükségtelen vizsgálni, hiszen a tanult tételekből következik, de a séma természetesen nem csak a 2NF tulajdonságokat, hanem a 3NF és BCNF tulajdonságokat is sérti. Könnyen látható ez például a $B \rightarrow D$ függésen: a függés bal oldalán nem szuperkulcs áll (B csak része egy kulcsnak), a függés jobb oldalán pedig nem elsődleges attribútum található.

3. feladat

Üres függéshalmaz esetén a lezártban csak triviális függőségek szerepelnek. A 3 elemű attribútumhalmazon értelmezett triviális függőségek a bal oldali attribútumhalmaz elemszáma szerint csoportosítva:

- 3 elemű (8 db): $ABC \rightarrow ABC; AB, AC, BC; A, B, C; \emptyset$

- 2 elemű (12 db):
$$\begin{cases} AB \rightarrow AB; A, B; \emptyset \\ AC \rightarrow AC; A, C; \emptyset \\ BC \rightarrow BC; B, C; \emptyset \end{cases}$$

- 1 elemű (6 db):
$$\begin{cases} A \rightarrow A; \emptyset \\ B \rightarrow B; \emptyset \\ C \rightarrow C; \emptyset \end{cases}$$

- 0 elemű (1 db): $\emptyset \rightarrow \emptyset$

F lezártja $8+12+6+1=27$ elemű. Ha valaki következetesen nem számolt az üres halmazzal, az ő megoldása is lehetett teljes értékű.

Érdekesség. n attribútum esetén a triviális függőségek száma az alábbi módon számítható ki. A függőségek bal oldalát tetszőlegesen megválaszthatjuk: egy k -elemű attribútumhalmaz $\binom{n}{k}$ -féleképpen választható ki. A függőség jobb oldalára a bal oldal tetszőleges részhalmazát írhatjuk, egy k -elemű halmaznak 2^k részhalmaza van. Összegezve $k=0$ -tól n -ig: $\sum_{k=0}^n \binom{n}{k} 2^k$.

Ha az összeadás törzsében lévő kifejezést $1=1^{n-k}$ -nal szorozzuk, alkalmazhatjuk a binomiális tételt: $\sum_{k=0}^n \binom{n}{k} 2^k 1^{n-k} = (2+1)^n = 3^n$. A fenti esetben $n=3$, a triviális függőségek száma $3^3 = 27$.

4. feladat

Egy blokkba $f_r = \left\lfloor \frac{b}{s_r} \right\rfloor = \left\lfloor \frac{4000}{850} \right\rfloor = 4$ adatrekord fér. Egy kulcs–mutató párból álló indexrekord $s_i = 50 + 18 = 68$

bájt, egy blokkba $f_i = \left\lfloor \frac{b}{s_i} \right\rfloor = 58$ indexrekord fér.

a) Az $n_r = 25\,000$ adatrekord tárolásához legkevesebb $b_r = \left\lceil \frac{n_r}{f_r} \right\rceil = \left\lceil \frac{25\,000}{4} \right\rceil = 6250$ adatblokk szükséges. A

ritka index minden blokkra tartalmaz mutatót ($n_i = b_r$), ez minimum $b_i = \left\lceil \frac{n_i}{f_i} \right\rceil = \left\lceil \frac{6250}{58} \right\rceil = 108$ indexblokk,

tehát összesen legalább 6 358 blokkra van szükség.

b) 6000 bájtön csak 1 blokkot tudunk a memóriában tartatni. A ritka indexen történő bináris kereséshez legfeljebb $1 + \lfloor \log_2 108 \rfloor = 7$ blokkműveletre van szükség, az adatblokk kiolvasásához még egyre, azaz egy rekord tartalmának kiolvasása legfeljebb $8 \cdot 0,5 \text{ ms} = 4 \text{ ms}$ -ig tart.

c) A többletmemóriát az indexstruktúra gyorsítótáraként érdemes felhasználni.

10-szer annyi szabad memória 60 000 bájt, amivel 15 blokkot tudunk a memóriában tárolni. Célszerű a bináris keresés első 3 lépése során elérhető $1 + 2 + 4 = 7$ indexblokkot folyamatosan a memóriában tartani, ezzel 3 blokkműveletet tudunk megtakarítani, tehát a keresés ekkor csak 2,5 ms ideig tart.

100-szor annyi szabad memória: 600 000 bájt hely, 150 blokkot tudunk a memóriában tartani. A teljes indexstruktúra befér a memóriába, így csak az adatblokkot kell 1 blokkművelettel a háttértárról beolvasnunk, tehát a keresés 0,5 ms alatt elvégezhető.

Részletesebb magyarázat. A c) feladatban 10-szer annyi memória esetén a 4. lépéshez tartozó további 8 blokk memóriában tartása nem gyorsítaná a rekord tartalmának kiolvasását minden esetben, mert ezzel már az indexblokkok teljesen elfoglalnák az összes szabad memóriát. Így minden adatblokk beolvasása után újra be kellene olvasnunk egy indexblokkot.

5. feladat

A rekordhossz $s_r = |A| + |B| + |C| + |D| = 10$ bájt, egy blokkba $f_r = \left\lfloor \frac{b}{s_r} \right\rfloor = \left\lfloor \frac{1000}{10} \right\rfloor = 100$ rekord fér. Az adatstruktú-

ra tárolásához minimum $b_r = \left\lceil \frac{n_r}{f_r} \right\rceil = \left\lceil \frac{1\,000\,000}{100} \right\rceil = 10\,000$ blokk szükséges. A heap szervezés miatt legrosszabb

esetben az összes blokkot végig kell olvasnunk ahhoz, hogy a módosítandó rekordot megtaláljuk, ha pedig nem lehetünk benne biztosak, hogy csak egyetlen rekordot kell módosítani, akkor az összes blokkot végig kell olvasni, és eközben biztosan megtaláljuk az összes módosítandó rekordot, akárhány is van. Induljunk ki tehát abból, hogy a módosítást megelőző kereséshez 10 000 blokkműveletre (azaz 5 s-re) van szükség, és a továbbiakban csak a módosítások időigényére összpontosítsunk.

Ehhez azt kell megvizsgáljunk, hogy az írás a funkcionális függőségek miatt hány rekordot érinthet.

a) A C attribútum teljesen függ az AB kulcstól, ezért nem fordulhat elő két olyan sor a relációban, amely az AB attribútumokon megegyezik. Így mindig elég egyetlen rekordot írni, amihez egy blokkműveletre van

szükség. A módosítás időigénye – a rekord megkeresésén felül – az egyetlen blokkműveletnek megfelelő 0,5 ms.

- b) A $B \rightarrow D$ függőség miatt a D attribútum módosítása esetén a B attribútumon megegyező valamennyi sorban is módosítanunk kell a D attribútumot. Legrosszabb esetben akár az összes sort módosítanunk kell. Ehhez 10000 blokk írására lehet szükség, azaz a módosítás időigénye – a keresés időigényén felül – 5000 ms = 5 s.

Részletesebb magyarázat. A heap szervezés nem használ mutatókat, ezért azok mérete irreleváns. A b) feladatrész legrosszabb esetére példa: az alábbi reláció esetén bármelyik sorban megváltoztatjuk D értékét, az összes sorban meg kell változtatnunk a B attribútumon felvett azonos b értékek miatt. A $B \rightarrow D$ függőségben a D másodlagos attribútum a kulcs részétől függ, ezért a függőség redundanciát okozhat (és sérti a 2NF definícióját).

A	B	C	D
a1	b	c1	d
a2	b	c2	d
a3	b	c3	d

Megjegyzendő, hogy a megoldás során kimondatlanul feltételeztük, hogy a rekordok a lehető legkevesebb blokkot foglalják el. Előfordulhat azonban, hogy (törlések miatt) ez nem így van, legrosszabb esetben akár az is előfordulhat, hogy minden blokkban csak 1-1 rekord található. Ekkor természetesen az egyes műveletek blokkművelet-igénye jelentősen megnő. (Ezt a teljes értékű megoldáshoz sem kell figyelembe venni.)

6. feladat

A relációs sémák:

- $A(\underline{A1}, A2, A3, \underline{B1})$, $B(\underline{B1}, B2, \underline{C1})$, $C(\underline{C1}, C2, \underline{D1}, \underline{D2}, \underline{D3})$, $D(\underline{D1}, \underline{D2}, \underline{D3}, \underline{S.A1}, \underline{T.A1})$
- $U(\underline{D1}, \underline{D2}, \underline{D3}, \underline{A1})$

Részletesebb magyarázat. A több-egy kapcsolattípusok leképezhetők úgy, hogy a „több” oldalon álló entitáshalmaznak megfelelői relációs sémába felvesszük a megfelelő attribútumo(ka)t. A több-több kapcsolathoz külön sémát kell felvennünk.