

Hálózatba kapcsolt adatbázisok

Lekérdezés-feldolgozás
Erős Levente, 2016.

Bevezető

- Relációk töredékekben tárolódnak
 - Felhasználó elől rejtve
 - Felhasználó teljes relációkat lát
 - Cél: Teljes relációkra vonatkozó lekérdezések elvégzése a töredékeken
- Lépések
 - 1. Lekérdezés-dekompozíció
 - 2. Adatlokalizáció
 - 3. Optimalizálás (majd végrehajtás)

1. Lekérdezés-dekompozíció

- Feladata
 - 1.1 Lekérdezés-normalizálás
 - 1.2 (Típus-, szemantikai) megfelelés ellenőrzése
 - 1.3 Redundancia megszüntetése
 - 1.4 Lekérdezés-átírás

1.1 Lekérdezés-normalizálás

- Lekérdezés logikai kifejezéseinek (szelekció) ekvivalens átírása valamely normálformába
- Cél: későbbi feldolgozás megkönnyítése
- Konjunktív normálforma
 - VAGY-ok ÉS kapcsolata
 - $(p_{11} \vee \dots \vee p_{1n}) \wedge \dots \wedge (p_{m1} \vee \dots \vee p_{mn})$
- Diszjunktív normálforma
 - ÉS-ek VAGY kapcsolata
 - $(p_{11} \wedge \dots \wedge p_{1n}) \vee \dots \vee (p_{m1} \wedge \dots \wedge p_{mn})$

1.1 Lekérdezés-normalizálás

- Ekvivalencia-szabályok az átalakításhoz (referencia)

1) $p1 \wedge p2 = p2 \wedge p1$

2) $p1 \vee p2 = p2 \vee p1$

3) $p1 \wedge (p2 \wedge p3) = (p1 \wedge p2) \wedge p3$

4) $p1 \vee (p2 \vee p3) = (p1 \vee p2) \vee p3$

5) $p1 \wedge (p2 \vee p3) = (p1 \wedge p2) \vee (p1 \wedge p3)$

6) $p1 \vee (p2 \wedge p3) = (p1 \vee p2) \wedge (p1 \vee p3)$

7) $\neg(p1 \wedge p2) = \neg p1 \vee \neg p2$ (De Morgan)

8) $\neg(p1 \vee p2) = \neg p1 \wedge \neg p2$ (De Morgan)

9) $\neg(\neg p) = p$

1.1 Lekérdezés-normalizálás

- Mikor melyik normálforma előnyös?
 - Diszjunktív normálforma esetén
 - A lekérdezés független (párhuzamosítható) lekérdezésekkel helyettesíthető
 - Ezek uniója az eredeti lekérdezés eredménye
 - Konjunktív normálforma esetén
 - Könnyebb az átalakítás, mivel a szelekciós feltételekben általában több ÉS van
 - Diszjunktív normálformára való alakításkor predikátumok ismétlődhetnek (5) szabály)

1.2 Megfelelőség ellenőrzése

- Típusmegfelelőség
 - Típus-inkorrekt a lekérdezés, pl. ha nem megfelelő típusúak az operandusok (1+”alma” – van ahol ez is értelmezett, nyelvfüggő)
 - Szemantikailag inkorrekt, ha van olyan komponens, ami nem járul hozzá a végeredmény előállításához
 - V és \neg műveletet nem tartalmazó relációalgebrai kifejezés esetén eldönthető
 - **Lekérdezési gráf (nem lekérdezési fa)**
 - Csomópontok – relációk, eredmény
 - Élek
 - » Relációk között – join (szelekció + Descartes-szorzat)
 - » Reláció, önmagával – szelekció, önillesztés
 - » Reláció és eredmény között – projekció
 - Ha nem összefüggő a gráf, akkor és csak akkor **szemantikailag inkorrekt** a lekérdezés
 - Példa → Tábla

1.3 Redundancia megszüntetése

- Ha feleslegesen bonyolult a lekérdezés
- Szabályok (referencia)
 - 1) $p \wedge p = p$
 - 2) $p \vee p = p$
 - 3) $p \wedge \text{igaz} = p$
 - 4) $p \wedge \text{hamis} = \text{hamis}$
 - 5) $p \vee \text{hamis} = p$
 - 6) $p \vee \text{igaz} = \text{igaz}$
 - 7) $p \wedge \neg p = \text{hamis}$
 - 8) $p \vee \neg p = \text{igaz}$
 - 9) $p_1 \wedge (p_1 \vee p_2) = p_1$
 - 10) $p_1 \vee (p_1 \wedge p_2) = p_1$

1.4 Lekérdezés átírása

- Részben optimalizáció (ököl szabályok mentén)
- Lekérdezésből – **lekérdezési fa**
 - Levelek – relációk
 - Élek – operandusok
 - Csomópontok \ Levelek – műveletek
- Lekérdezési fa ekvivalencia-szabályok szerint alakítható
 - 1) Bináris operátorok kommutatívak
 $r \times s = s \times r, r \bowtie s = s \bowtie r$
 - 2) Bináris operátorok asszociatívak
 $(r \times s) \times t = r \times (s \times t), (r \bowtie s) \bowtie t = r \bowtie (s \bowtie t)$
 - 3) Unáris operátorok csoportosíthatósága
 $\pi_A(\pi_B r) = \pi_A r, \text{ ha } A \subseteq B$

1.4 Lekérdezés átírása

- Lekérdezési fa ekvivalencia-szabályok szerint alakítható

1) Szelekció és projekció felcserélhetősége

$$\pi_{A_1, \dots, A_n}(\sigma_{p(A_p)} r) = \pi_{A_1, \dots, A_n}(\sigma_{p(A_p)}(\pi_{A_1, \dots, A_n, A_p} r))$$

2) Szelekció és bináris operátorok felcserélhetősége

$$\sigma_{p(A_p)}(r \bowtie s) = (\sigma_{p(A_p)} r) \bowtie s,$$

$$\sigma_{p(A_p)}(r \times s) = (\sigma_{p(A_p)} r) \times s,$$

ha A_p r sémájának (R -nek) attribútuma

hasonlóan, ha A_p s attribútuma (is)

$$\sigma_p(r \cup t) = (\sigma_p r) \cup (\sigma_p t)$$

3) Projekció és bináris operátorok felcserélhetősége

$$\pi_C(r \bowtie s) = \pi_A r \bowtie \pi_B s,$$

$$\pi_C(r \times s) = \pi_A r \times \pi_B s,$$

ha $C = A \cup B$, $A \subseteq R$, $B \subseteq S$

$$\pi_C(r \cup s) = \pi_C r \cup \pi_C s$$

1.4 Lekérdezés átírása

- Átírás szempontjai, lehetőségei
 - Unáris operátorok különválasztása
 - Külön relációkra külön végezzük
 - Egy reláció unáris operátorainak csoportosítása, süllyesztése
- Cél
 - Köztes (részeredmény-) relációk méretének csökkentése
 - Költséges műveletek (Descartes-szorzat, illesztés) operandusai minél kisebb relációk legyenek
- Példa → Tábla

2. Adatlokalizáció

- Feladata

- 1) Globális relációkra vonatkozó lekérdezés lokális relációkra (töredékekre) vonatkozó lekérdezésekké való átalakítása

- Eredménye: **Lokalizált lekérdezés**

- Globális relációkat helyettesítjük a töredékek **uniójával** (horizontális felosztás esetén) vagy **természetes illesztésével** (vertikális felosztás esetén)

- 2) Lokalizált lekérdezés egyszerűsítése, redukció

- Eredménye: **Redukált lekérdezés**

2. Adatlokalizáció

- Redukció horizontális felosztás esetén
 - Szelekció esetén
 - Ha a szelekciós feltétel (p_i) ellentmond a töredék (r_i) felosztási szabályának (p_j), a töredéken elvégzett szelekció üres relációt ad vissza.
 $\forall x \in r_i \neg(p_i(x) \wedge p_j(x))$, akkor $\sigma_{p_i}(r_j) = \emptyset$ ($r_j = \sigma_{p_j} r$)
 - Példa \rightarrow Tábla
 - Illesztés esetén
 - Feltétel: az illesztett relációk a join-attribútum szerint fragmentáltak
 - Első lépés: Join süllyesztése
 $r \bowtie s = (r_1 \cup r_2) \bowtie (s_1 \cup s_2) = (r_1 \bowtie s) \cup (r_2 \bowtie s) = (r_1 \bowtie s_1) \cup (r_1 \bowtie s_2) \cup (r_2 \bowtie s_1) \cup (r_2 \bowtie s_2)$, ahol $r_i = \sigma_{p_i} r$, $s_i = \sigma_{q_i} s$
 - Ha a két illesztett töredék felosztási feltételei ellentmondóak, az adott rész-illesztés elhagyható.
 - Példa \rightarrow Tábla

2. Adatlokalizáció

- Vertikális felosztás esetén
 - Ha a lekérdezésben projekciót végzünk, az üres relációt eredményez azon töredékeken, amely felosztási szabályának nincs közös attribútuma a projekció attribútumhalmazával
 - Nem üres, de haszontalan relációt eredményez azon töredékeken, amely felosztási szabályának a projekció attribútumhalmazával vett egyetlen közös attribútuma a join attribútum*
 - *ha csak a join attribútumra is projektálunk, egy tetszőleges töredéket megtartunk
 - Példa → Tábla

3. Optimalizálás

- Mire optimalizálunk?
 - Helyi feldolgozás ideje
 - Kommunikációhoz szükséges idő
 - Kulcsfontosságú
 - **Join-ok végrehajtása** optimális módon
 - Nagy lépésszám
 - Join stratégiák → Tábla
 - Teljes reláció (vagy töredék) átvitele
 - Igény szerinti átvitel

3. Optimalizálás

- Mikor melyik módszer jó?
 - Teljes reláció (vagy töredék) átvitele
 - Kevesebb üzenet
 - Nagyobb üzenetméret
 - Lokálisan 1 hozzáférés a relációhoz
 - Igény szerinti átvitel
 - Több üzenet
 - Kisebb üzenetméret
 - Lokálisan több hozzáférés a relációhoz (projekció, join)
 - Kis relációk → teljes reláció átvitele
 - Nagy relációk → igény szerinti átvitel