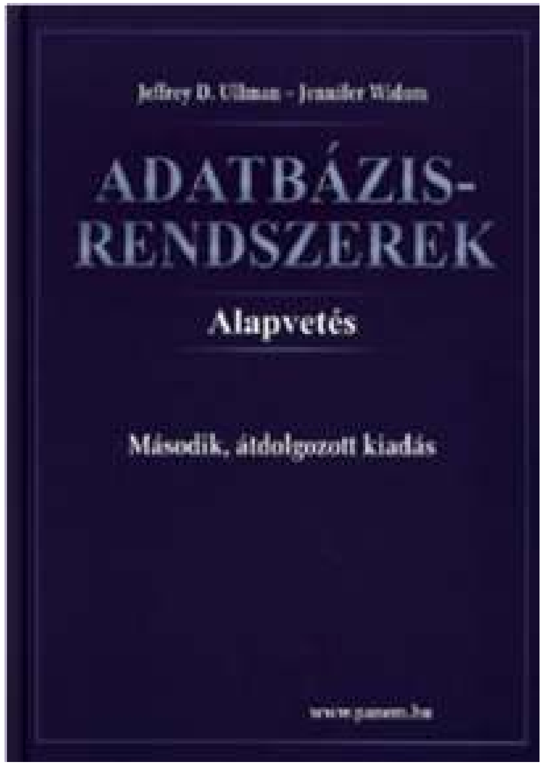
Relációs adatbázisok tervezése 2.rész (dekompozíció)

Tankönyv: Ullman-Widom:
Adatbázisrendszerek Alapvetés
Második, átdolgozott kiadás,
Panem, 2009



3.3. Relációs adatbázissémák tervezése, relációk felbontása
3.4. Dekompozíció tulajdonságai, veszteségmentes felbontás, függőségőrző felbontás.

Bevezető példda

- Több tábla -> vegyük a táblák összekapcsolását
  Melyik séma jobb?

Sörivó(<u>név</u>, cím, <u>kedveltSörök</u>, gyártó, kedvencSör)

név	cím	kedveltSörök	gyártó	kedvencS
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	???	WickedAle	Pete's	???
Spock	Enterprise	Bud	???	Bud

- Redundáns adat, a ??? helyén mi szerepel, ha név -> cím, kedvencSör és kedveltSörök -> gyártó funkcionális függőségek (def.később) érvényesek

Hibás tervezés problémái

A rosszul tervezettség anomáliákat is eredményez

Sörivó(<u>név</u>, cím, <u>kedveltSörök</u>, gyártó, kedvencSör)

név	cím	kedveltSörök	gyártó	kedvencS
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	???	WickedAle	Pete's	???
Spock	Enterprise	Bud	???	Bud

- Módosítási anomália: ha Janeway-t Jane-re módosítjuk, megtesszük-e ezt minden sornál?
- Törlési anomália: Ha senki sem szereti a Bud sört, azzal töröljük azt az infót is, hogy ki gyártotta.
- Beillesztési anomália: és felvinni ilyen gyártót?

Relációs sémák tervezése

- Cél: az anomáliák és a redundancia megszüntetése.
  - Módosítási anomália : egy adat egy előfordulását megváltoztatjuk, más előfordulásait azonban nem.
  - Törlési anomália : törléskor olyan adatot is elveszítünk, amit nem szeretnénk.
  - Beillesztési anomália : megszorítás, trigger kell, hogy ellenőrizni tudjuk (pl. a kulcsfüggőséget)
  - Redundancia (többszörös tárolás feleslegesen)

Dekomponálás (felbontás)

A fenti problémáktól dekomponálással (felbontással) tudunk megszabadulni:

- Definíció:
  $d=\{R_1,...,R_k\}$ az (R,F) dekompozíciója, ha nem marad ki attribútum, azaz $R_1 \cup ... \cup R_k = R$. Az adattábla felbontását projekcióval végezzük.
- Példa:
  R=ABCDE, d={AD,BCE,ABE}
  3 tagú dekompozíció, ahol
  $R_1=AD$, $R_2=BCE$, $R_3=ABE$,

Felbontásra vonatkozó elvárások

- Elvárások

(1) Veszteségmentes legyen a felbontás, vagyis ne legyen információvesztés (VM def.később)

(2) A vetületek legyenek jó tulajdonságúak, és a vetületi függőségi rendszere egyszerű legyen (normálformák: BCNF, 3NF def. később)

(3) Függőségek megőrzése a vetületekben (FŐ)

- Tételek (ezekre nézünk majd algoritmusokat)
  - Mindig van VM BCNF-ra való felbontás
  - Mindig van VM FŐ 3NF-ra való felbontás

Veszteségmentes szétvágás ---1

- A fenti jelölésekkel: ha $r = \Pi_{R_1}(r) \bowtie \ldots \bowtie \Pi_{R_k}(r)$ teljesül, akkor az előbbi összekapcsolásra azt mondjuk, hogy veszteségmentes. Itt r egy R sémájú reláció-előfordulást jelöl.

R

A	B	C
a	b	c
d	e	f
c	b	c

R_1

A	B
a	b
d	e
c	b

R_2

B	C
b	c
e	f

Veszteségmentes szétvágás ---2

- Megjegyzés: Könnyen belátható, hogy $r \subseteq \Pi_{R_1}(r) \bowtie \ldots \bowtie \Pi_{R_k}(r)$ mindig teljesül.
- Példa: itt a szétvágás után keletkező relációk összekapcsolása nem veszteségmentes:

R

A	B	C
a	b	c
c	b	e

R_1

A	B
a	b
c	b

R_2

B	C
b	c
b	e

Chase-teszt VM ellenőrzése ---1

- Példa: adott R(A, B, C, D), F = { $A \rightarrow B$, $B \rightarrow C$, $CD \rightarrow A$ } és az $R_1(A, D)$, $R_2(A, C)$, $R_3(B, C, D)$ felbontás. Kérdés veszteségmentes-e a felbontás?
- Vegyük $R_1 \bowtie R_2 \bowtie R_3$ egy t = (a, b, c, d) sorát. Bizonyítani kell, hogy t R egy sora. A következő tablót készítjük:

A	B	C	D
a	b_1	c_1	d
a	b_2	c	d_2
a_3	b	c	d

Itt pl. az (a, b_1, c_1, d) sor azt jelzi, hogy R-nek van olyan sora, aminek R_1-re való levetítése (a, d), ám ennek a B és C attribútumokhoz tartozó értéke ismeretlen, így egyáltalán nem biztos, hogy a t sorról van szó.

Chase-teszt VM ellenőrzése ---2

- Az F-beli függőségeket használva egyenlővé tesszük azokat a szimbólumokat, amelyeknek ugyanazoknak kell lennie, hogy valamelyik függőség ne sérüljön.
  - Ha a két egyenlővé teendő szimbólum közül az egyik index nélküli, akkor a másik is ezt az értéket kapja.
  - Két indexes szimbólum esetén a kisebbik indexű értéket kapja meg a másik.
  - A szimbólumok minden előfordulását helyettesíteni kell az új értékkel.
- Az algoritmus véget ér, ha valamelyik sor t-vel lesz egyenlő, vagy több szimbólumot már nem tudunk egyenlővé tenni.

Chase-teszt VM ellenőrzése ---3

A	B	C	D
a	b_1	c_1	d
a	b_2	c	d_2
a_3	b	c	d

$A \rightarrow B$

A	B	C	D
a	b_1	c_1	d
a	b_1	c	d_2
a_3	b	c	d

$B \rightarrow C$

A	B	C	D
a	b_1	c	d
a	b_1	c	d_2
a_3	b	c	d

$CD \rightarrow A$

A	B	C	D
a	b_1	c	d
a	b_1	c	d_2
a	b	c	d

Chase-teszt VM ellenőrzése ---5

- Ha t szerepel a tablóban, akkor valóban R-nek egy sora, s mivel t-t tetszőlegesen választottuk, ezért a felbontás veszteségmentes.
- Ha nem kapjuk meg t-t, akkor viszont a felbontás nem veszteségmentes.
- Példa: R(A, B, C, D), F = { $B \rightarrow AD$ }, a felbontás: R_1(A, B), R_2(B, C), R_3(C, D).

A	B	C	D
a	b	c_1	d_1
a_2	b	c	d_2
a_3	b_3	c	d

$B \rightarrow AD$ ⇒

A	B	C	D
a	b	c_1	d_1
a	b	c	d_1
a_3	b_3	c	d

Itt az eredmény jó ellenpélda, hiszen az összekapcsolásban szerepel t = (a, b, c, d), míg az eredeti relációban nem.

Másik tul.: Függőségek megőrzése

- A dekompozíciókban érvényes függőségekből következzen az eredeti sémára kirótt összes függőség.
- Milyen függőségek lesznek érvényesek a dekompozíció sémáiban?
- Példa: R=ABC, F= {$A\rightarrow B$, $B\rightarrow C$, $C\rightarrow A$} vajon a d= (AB, BC) felbontás megőrzi-e a $C\rightarrow A$ függőséget? (függőségek vetülete, definíció a következő lapon)

Függőségek megőrzése (definíció)

- Definíció: Függőségek vetülete

  Adott (R,F), és $R_i \subseteq R$ esetén:

  $\Pi_{R_i}(F) := \{ X \rightarrow Y \mid F \vdash X \rightarrow Y \text{ , } XY \subseteq R_i \}$

- Definíció: Adott (R,F) esetén $d=(R_1,\ldots,R_k)$ függőségőrző dekompozíció akkor és csak akkor, ha minden F-beli függőség levezethető a vetületi függőségekből:

  minden $X \rightarrow Y \in F$ esetén

  $\Pi_{R_1}(F) \cup \ldots \cup \Pi_{R_k}(F) \vdash X \rightarrow Y$

Példa: függőségek vetülete

- ABC, $A \to B$ és $B \to C$ FF-kel.
  Nézzük meg az AC-re való vetületet:
  - $A^+ = ABC$; ebből $A \to B$, $A \to C$.
    - Nem kell kiszámítani AB^+ és AC^+ lezárásokat.
  - $B^+ = BC$; ebből $B \to C$.
  - $C^+ = C$; semmit nem ad.
  - $BC^+ = BC$; semmit nem ad.
- A kapott FF-ek: $A \to B$, $A \to C$ és $B \to C$.
- AC -re projekció: $A \to C$.

Függőségek megőrzése (tételek)

- A függőségőrzésből nem következik a veszteségmentesség:

  $R=ABCD$, $F=\{A \rightarrow B, C \rightarrow D\}$, $d=\{AB, CD\}$ függőségőrző, de nem veszteségmentes.

- A veszteségmentességből nem következik a függőségőrzés

  $R=ABC$, $F=\{AB \rightarrow C, C \rightarrow A\}$, $d=\{AC, BC\}$ veszteségmentes, de nem függőségőrző.