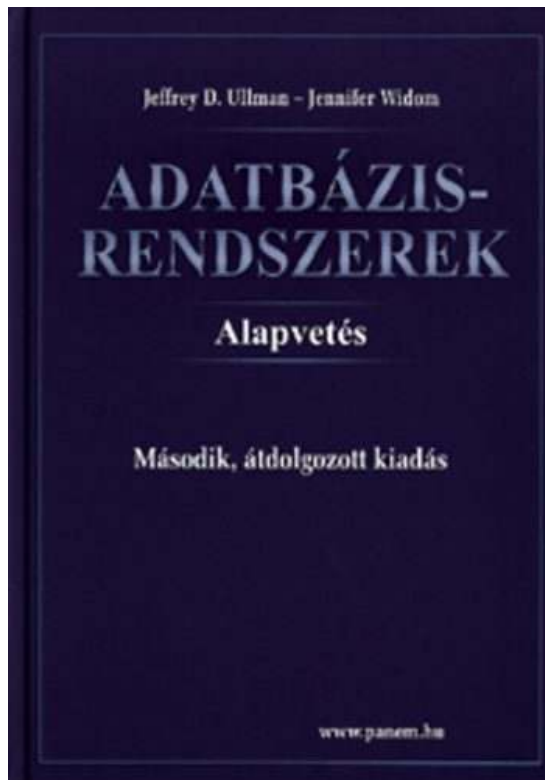


Kiterjesztett relációs algebra



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

5.1. Relációs műveletek
multihalmazokon

5.2. Kiterjesztett műveletek
a relációs algebrában

Relációs algebra kiterjesztése

- Eddig tanult műveletek: **projekció** (Π), **szelekció** (σ), **direkt/Descartes-szorzat** ($\times$), **unió** ($\cup$), **különbség** ($-$), **természetes összekapcsolás** ($\bowtie$), **átnevezés** (ρ), ...
- Ezeket most multihalmazok fölött értelmezzük, azaz egy reláció nem sorok **halmazából**, hanem **multihalmazából** áll, ahol **megengedettek az ismétlődések**.
- Ezeken kívül bevezetjük még:
 - ❑ **az ismétlődések kiküszöbölésére** új műveletet (δ),
 - ❑ **csoportosítás és összesítést** (γ_L)
 - ❑ **a projekcióknak egy kiterjesztett változatát.**
 - ❑ **a rendezést** (τ_L)
 - ❑ **a külső összekapcsolásokat** ($\overset{0}{\bowtie}$)

Multihalmazok egyesítése, metszete, különbsége

- **Unió:** $R \cup S$ -ben egy t sor annyiszor fordul elő ahányszor előfordul R -ben, plusz ahányszor előfordul S -ben.
- **Metszet:** $R \cap S$ -ben egy t sor annyiszor fordul elő, amennyi az R -ben és S -ben lévő előfordulások minimuma.
- **Különbség:** $R - S$ -ben egy t sor annyiszor fordul elő, mint az R -beli előfordulások mínusz az S -beli előfordulások száma.

R			S			A		B	
A	B		A	B		A		B	
1	3	$\cup$	1	3	$\Rightarrow$	1		3	
1	2		2	5		1		2	
						1		3	
						2		5	

A „többi művelet” multihalmazok fölött

- A projekció, szelekció, Descartes-szorzat, természetes összekapcsolás és Théta-összekapcsolás végrehajtása során nem küszöböljük ki az ismétlődéseket.

R			$\Pi_A(R)$	
A	B		A	
1	2	➔	1	
1	5		1	
2	3		2	

Ismétlődések megszüntetése

- **Ismétlődések kiküszöbölése:** $R1 := \delta(R2)$
- A művelet jelentése: R2 multihalmazból R1 halmazt állít elő, vagyis az R2-ben egyszer vagy többször előforduló sorok csak egyszer szerepelnek az R1-ben.
- A DISTINCT reprezentálására szolgál

$R = ($

A	B
1	2
3	4
1	2

$\delta(R) =$

A	B
1	2
3	4

Kiterjesztett projekció

- $\Pi_L(R)$ L listája az SQL **SELECT** záradékához hasonlóan tartalmazhatja:
 - R egy attribútumát,
 - $x \rightarrow y$, ahol x, y attribútumnevek, s itt x-et y-ra nevezzük át,
 - $E \rightarrow y$, ahol E R attribútumait, konstansokat, aritmetikai operátorokat és karakterlanc operátorokat tartalmazhat például: $A + 5$, $C \parallel \text{'nevű emberek'}$.

R			$\Pi_{A, B+C \rightarrow x}(R)$	
A	B	C	A	x
0	1	3	0	4
4	9	1	4	10
5	5	7	5	12

Rendezés

- **Rendezés:** $\tau_{A_1, \dots, A_n}(R)$.
- Először A_1 attribútum szerint rendezzük R sorait. Majd azokat a sorokat, amelyek értéke megegyezik az A_1 attribútumon, A_2 szerint, és így tovább.
- Hasonlóan az **ORDER BY** működéséhez.
- Ez az egyetlen olyan művelet, amelynek az eredménye se nem halmaz se nem multihalmaz.

$R =$

(A	B)
1	2
3	4
5	2

$$\tau_B(R) = [(5,2), (1,2), (3,4)]$$

Aggregáció, összesítési műveletek

R =

A	B
1	3
3	4
3	2

SUM(A) = 7

COUNT(A) = 3

MAX(B) = 4

AVG(B) = 3

Csoportosítás és összesítés $\gamma_L(R)$

- A művelet a **csoportosítást** (**GROUP BY**), a csoportokon végezhető **összesítő függvényeket** (AVG, SUM, COUNT, MIN, MAX) reprezentálja.
- A művelet jele: $\gamma_L(R)$.
- Itt az **L** lista valamennyi eleme a következők egyike:
 - R egy attribútuma: ez az attribútum egyike a csoportosító attribútumoknak (a GROUP BY után jelenik meg).
 - a reláció egyik attribútumára (ez az összesítési attribútumra) alkalmazott összesítő operátor, ha az összesítés eredményére névvel szeretnénk hivatkozni, akkor nyilat és új nevet használunk.

Csoportosítás és összesítés $\gamma_L(R)$

- Osszuk R sorait csoportokba. Egy csoport azokat a sorokat tartalmazza, amelyek az L listán szereplő csoportosítási attribútumokhoz tartozó értékei megegyeznek
 - Vagyis ezen attribútumok minden egyes különböző értéke egy csoportot alkot.
- Minden egyes csoporthoz számoljuk ki az L lista összesítési attribútumaira vonatkozó összesítéseket
- Az eredmény minden egyes csoportra egy sor:
 1. A csoportosítási attribútumok és
 2. Az összesítési attribútumra vonatkozó összesítések (az adott csoport összes sorára)

Példa: Csoportosításra és aggregációra

R =

A	B	C
1	2	3
4	5	6
1	2	5

$\gamma_{A,B,AVG(C)}(R) = ??$

Először csoportosítunk

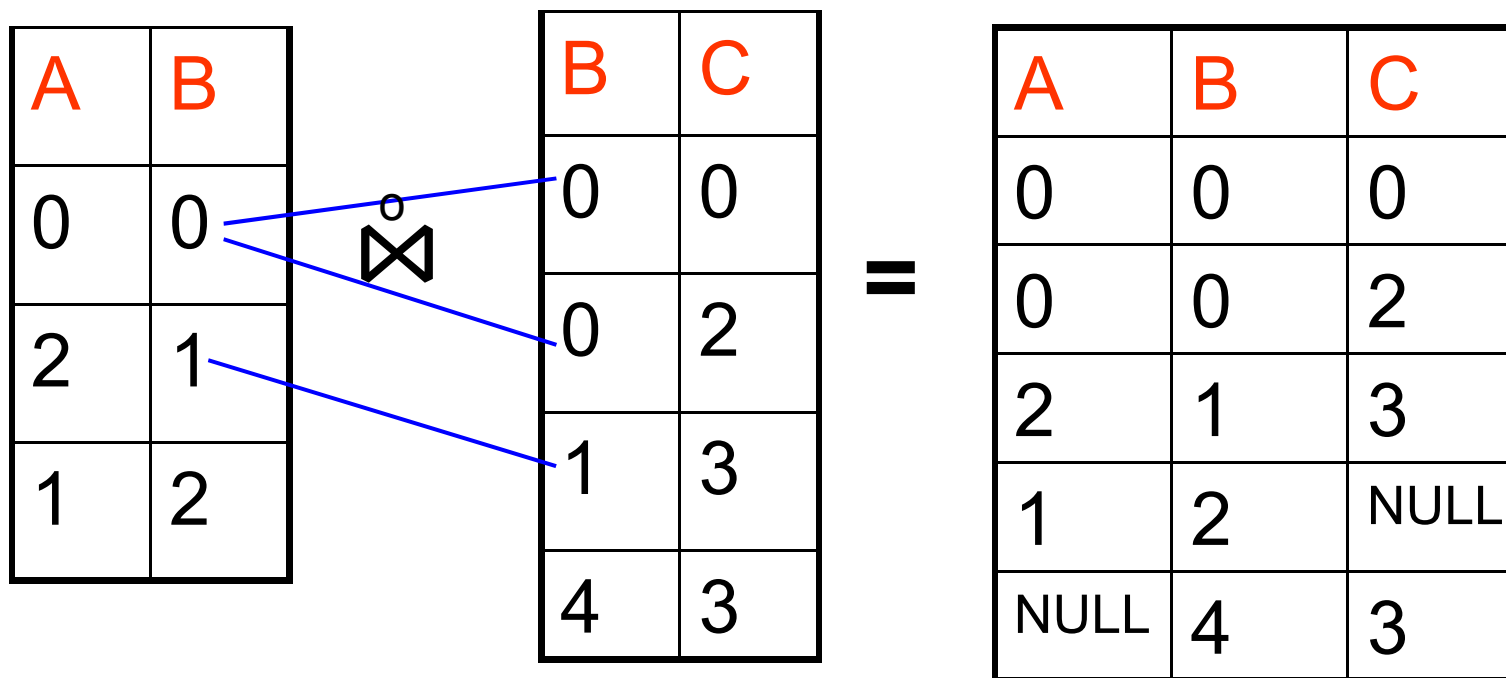
A	B	C
1	2	3
1	2	5
4	5	6

Csoportonként AGG:

A	B	X
1	2	4
4	5	6

Külső összekapcsolások

- Ez **nem relációs algebrai művelet**, mert kilép a modellből, lehet baloldali, jobboldali, és teljes külső összekapcsolás.
- R, S sémái $R(A_1, \dots, A_n, B_1, \dots, B_k)$, illetve $S(B_1, \dots, B_k, C_1, \dots, C_m)$
- $R \bowtie S = R \bowtie S$ relációt kiegészítjük az R és S soraival, a hiányzó helyekre NULL értéket írva megőrzi a „lógó sorokat”



Kiterjesztett relációs algebrai műveletek

- **Unió:** $R \cup S$ (előfordulások **összege**)
- **Metszet:** $R \cap S$ (előfordulások **minimuma**)
- **Különbség:** $R - S$ (előfordulások **különbsége**)
- **Kiválasztás:** $\sigma_{C(R)}$ (ahol **C** egy feltétel, AND, OR, NOT)
- **Vetítés:** $\pi_{L(R)}$ (L-ben **E $\rightarrow$ z** megengedett, ahol E egy kifejezés, és z ennek az új neve)
- **Szorzat:** $R \times S$
- **Ismétlődések kiküszöbölése:** $\delta(R)$
- **Csoportosítás és összesítés:** $\gamma_{L(R)}$ (L-ben attribútumok és összesítő fv-ek lehetnek valamint ezek átnevezése)
- **Rendezés:** $\tau_{L(R)}$ L-ben attribútumok listája szerepelhet) (algebrai kifejezésben csak utolsó operátorként van értelme beszélni a τ operátorról)
- **Külső összekapcsolások:** (outer join $R \bowtie S$)

Vissza az SQL-hez

- A kiterjesztett relációs algebrai műveletek mindegyikének van megfelelője az SQL-ben
- Lásd a következő AB1_04B_SQL4joins.pdf a Tankönyv 6.3.6- Összekapcsolások SQL-ben valamint AB1_04C_SQL5csop.pdf bemutatókat a Tankönyv 6.4. Relációkra vonatkozó műveletek című fejezetei alapján
- A Tankönyv 5.fejezet további 5.3-5.4. Datalog szakaszaira később, az SQL (6-8.fejezetek) és a PSM (9.3-9.4.szakaszok) tárgyalása után a 10.2 Rekurzió az SQL-ben előtt visszatérünk.