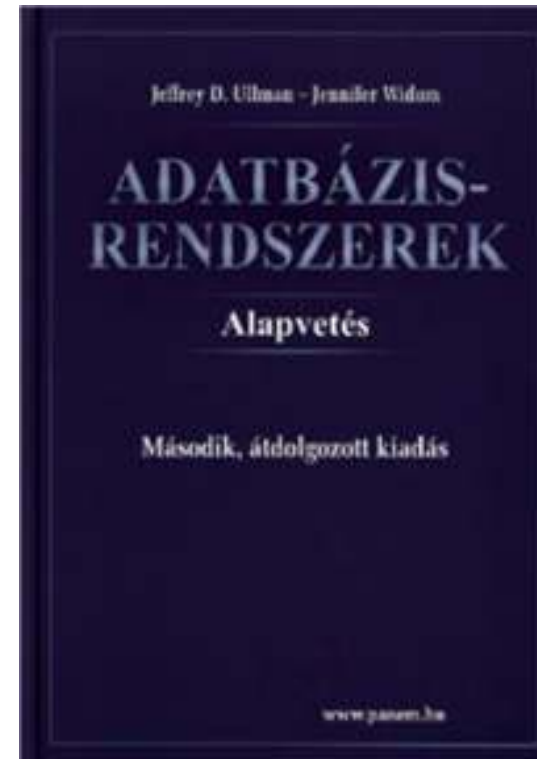


Relációs algebra 3.rész műveletek kiterjesztése multihalmazokra

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



- 5.1. Relációs műveletek
multihalmazokon
- 5.2. Kiterjesztett vetítés

Folyt.köv. 04A_RelAlg5kiterj.pdf :
további kiterjesztett műveletek a relációs algebrában

Relációs algebra kibővítése

- Az eddig tanult műveleteket: **vetítés** (Π), **kiválasztás** (σ), **halmazműveletek**: **unió** (\cup), **különbség** ($-$), **metszet** (\cap), **szorzás**: **természetes összekapcsolás** (\bowtie), **direkt-szorzat** (\times), stb. **multihalmazok fölött értelmezzük, mint az SQL-ben**, egy reláció nem sorok halmazából, hanem **multihalmazából áll**, vagyis megengedett a sorok ismétlődése.
- Ezekon kívül a **SELECT kiegészítéseinek és záradékainak** megfelelően **új műveletekkel is kibővítjük** a rel. algebrát:
 - **Ismétlődések megszüntetése** (δ) - **select distinct ..**
 - **Vetítési művelet kiterjesztése** (Π_{lista}) - **select kif [as onev]..**
 - **Rendezési művelet** (τ_{lista}) - **order by..**
 - **Összesítő műveletek és csoportosítás** (γ_{lista}) - **group by..**
 - **Külső összekapcsolások** ($\overset{\circ}{\bowtie}$) - **[left | right | full] outer join**

Multihalmazok egyesítése, különbsége

- **Unió:** $R \cup S$ -ben egy t sor annyiszor fordul elő ahányszor előfordul R -ben, plusz ahányszor előfordul S -ben: $n+m$
- **Metszet:** $R \cap S$ -ben egy t sor annyiszor fordul elő, amennyi az R -ben és S -ben lévő előfordulások minimuma: $\min[n, m]$
- **Különbség:** $R - S$ -ben egy t sor annyiszor fordul elő, mint az R -beli előfordulások mínusz az S -beli előfordulások száma, ha ez pozitív, egyébként pedig 0, vagyis $\max[0, n-m]$
- $(R \cup S) - T =? (R - S) \cup (S - T)$

R		S		A	B
	A		A	1	3
	B		B	1	2
1	3	\cup	1	3	
1	2		2	5	
				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 (duplikátumok kiszűrése)

- **Ismétlődések megszüntetése:** $R1 := \delta(R2)$
- 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
- A δ speciális esete lesz az általánosabb γ műveletnek

$R =$ (

A	B
1	2
3	4
1	2

$\delta(R) =$

A	B
1	2
3	4

Általánosított vetítés

- $\Pi_L(R)$ 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 karakterlánc 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

Általánosított vetítés

- $\Pi_L(R)$ L listája az SQL **SELECT** záradékához hasonlóan tartalmazhat az attribútumnevek helyén függvényeket.
- SQL leggyakrabban használt sorfüggvények:
- Numerikus, dátum, konverziós, stb... folyt.kov....

Kiválasztott sorok rendezése

- **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

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