

Relációs algebra

Relációs sémák tervezése



Takács Gábor

mérnök informatikus, okl. mérnöktanár

takacsg@sze.hu

<http://rs1.sze.hu/~takacsg/>

Bevezető

- Relációs adatbázisséma:

Relációsémák egy halmaza és az integritási megszorítások összessége

Mitől lesz ez használható?

Az adatbázisban

- Nem csak adatokat akarunk tárolni, hanem
- információkat is szeretnénk kinyerni, ehhez
- műveletek szükségesek, amivel manipulációt hajthatunk végre a relációkon.
- Lekérdezéseket specifikálunk
- A műveletek eredménye újabb reláció

PROBLÉMÁK HIBÁS RELÁCIÓSÉMÁK ESETÉN

Ha túl sok információt (attribútumot) helyezünk el egy relációban, akkor problémák, **anomáliák** fordulhatnak elő.

Alapvető anomáliák (szabálytalanságok):

➤ **Tárolási anomália:**

Mivel egy adat több sorban is szerepelhet, feleslegesen pazaroljuk rá a tárat (háttértár).

➤ **Módosítási problémák:**

Ugyan azt az adatot több helyen kell módosítani. Ha valamelyik kimarad az veszélyes

➤ **Törlési problémák:**

Törlések után egy értékhalmoz üressé válhat (csak egész sorokat törölhetünk), vele együtt egyéb információ is eltűnhet a relációból.

Pl.: Töröljük a tanszékvezetőt, és vele együtt törlődik a tanszék.

Ezek megszüntetésére alkalmas megoldás a relációk felbontása, azaz a dekompozíció.

Absztrakt lekérdező nyelvek:

- Relációalgebra

Egy eljárást adunk meg, (a hogyan) a kinyerni kívánt információ előállítására.

Imperatív (Procedurális) nyelv

- Relációkalkulus

Dekleratív kifejezésben adjuk meg (a mit) a kinyerni kívánt információt.

Nem procedurális nyelv

- Rekord alapú

- tartományalapú



Relációalgebra

A relációkkal kapcsolatban a matematika egy külön ága fejlődött ki.
Ez a relációs algebra...

- Matematikai halmazelméleten alapuló lekérdező nyelv
- A lekérdezés egy kifejezés, amelyben az
 - operátorok a relációalgebrai műveletek,
 - az operandusok pedig a relációk.
- A lekérdezés eredménye szintén egy reláció

RELÁCIÓS ALGEBRA (RELÁCIÓS ADATMODELL)

Ismétlés:

- Az adatok **relációkban**, táblázatokban jelennek meg
- Táblázat **oszlopai** = egyed attribútumai, tulajdonságai
- Táblázat **sorai** = egy konkrét egyed-előfordulás értékei

rendszer	alvázszám	gjFajta	gyártmány	gjSzín
ABC123	WVWZZZ3GH3S456987	személygépkocsi	Volkswagen	fehér
XYZ987	SZKAFF2JG5L123456	személygépkocsi	Suzuki	piros
ACY456	RENKKZ4GH6Q321456	haszonjármű	Renault	fekete

Pl.: Gépjármű relációra

RELÁCIÓS ADATMODELL

Alapfogalom: **RELÁCIÓSÉMA**

Reláció nevének, és attribútum halmazának megadását jelenti.

Reláció név: GÉPJÁRMŰ

Reláció attribútumok: Rendszám, Alvászsám, gjFajta, gyártmány, gjSzín

Reláció séma

→ **Reláció név** (Reláció attributum1, Ra2 , Ra3, Ra4, , RaN)

Megjelenítése:

rendsám	alvászám	gjFajta	gyártmány	gjSzín
ABC123	WVWZZZ3GH3S456987	személygépkocsi	Volkswagen	fehér
XYZ987	SZKAFF2JG5L123456	személygépkocsi	Suzuki	piros
ACY456	RENKKZ4GH6Q321456	haszonjármű	Renault	fekete

A relációséma attribútumai halmazt alkotnak, nem listát.

Reláció foka: attribútumok száma | Példánkban: 5-öd fokú reláció.

RELÁCIÓS ADATMODELL

Alapfogalom: **SOROK**

- ▶ A reláció azon sorait, amelyek különböznek az attribútumokból álló fejléc sorától, soroknak nevezzük. A sorban egy adott attribútumhoz kapcsolódó érték, a **komponens**.

Példa reláció séma:

GÉPJÁRMŰ (Rendszám, Alvázszám, gjFajta, gyártmány, gjSzín)

Komponens példa:

BBB-222, SZKAWFF2GKTKK34452958, személygépkocsi, Toyota, zöld

Komponenseket mindig ugyan olyan sorrendben írjuk le, mint ahogyan az attribútumokat felsoroltuk a relációsémában

RELÁCIÓS ADATMODELL

Alapfogalom: **ÉRTÉKTARTOMÁNY**

- **Attribútumok típusa csak atomi, azaz elemi lehet.**
Pl.: valamilyen adattípus, Pl.: egész szám, vagy karaktorsor
(Olyan attribútum érték nem lehet, ami kisebb egységekre még felbontható)
- **A komponensek csak az *attribútumnak megfelelő értéktartományból* vehetnek fel értéket.**

Alapfogalom: **RELÁCIÓ ELŐFORDULÁS**

Adott reláció sorainak halmaza a *reláció előfordulás*
De, időben a sorok változnak (törlés, módosítás, stb.)

Éppen létező állapot: **aktuális előfordulás**

RELÁCIÓS ADATMODELL

Alapfogalom: **RELÁCIÓS ALGEBRA**

- A relációs algebrai kifejezések alapjait a relációk képezik, mint operandusok. Egy reláció megadható a nevével (pl. **R** vagy **S**) vagy közvetlenül, sorainak egy listájával.

ABC123	WVWZZZ3GH3S456987	személygépkocsi	Volkswagen	fehér
XYZ987	SZKAFF2JG5L123456	személygépkocsi	Suzuki	piros
ACY456	RENKKZ4GH6Q321456	haszonjármű	Renault	fekete

- A reláció sémája az attribútumok felsorolását tartalmazza pl.: $R(A, B, C, D, E)$.

Gépjármű (rendszer, alvázszám, gjFajta, gyártmány, gjSzín)

- **(HALMAZ)MŰVELEZEK**: unió, metszet, különbség, vetítés, kiválasztás, szorzat, természetes összekapcsolás, bal/jobbs fél illesztés, théta összekapcsolás

HALMAZMŰVELETEK

UNIO (egyesítés) \rightarrow **Jelölés:** \cup .

► Relációink: $\mathbf{R, S} \rightarrow \mathbf{R \cup S}$

eredményül olyan reláció ad, amely tartalmazza az \mathbf{R} és \mathbf{S} sorait. Az ismétlődő sorokat elhagyjuk.

A művelet elvégzéséhez a két relációnak *azonos fokúnak* kell lenni. Legyen az \mathbf{R} és \mathbf{S} relációk sémája: (A, B, C) !

R	A	B	C	és	S	A	B	C
	2	-2	3			1	2	4
	4	1	4			2	-2	1
	1	2	4			4	1	4
						4	1	2

HALMAZMŰVELETEK

UNIO (egyesítés) → **Jelölés:** \cup .

R	A	B	C	és	S	A	B	C
	2	-2	3			1	2	4
	4	1	4			2	-2	1
	1	2	4			4	1	4
						4	1	2

Megoldás: Először az egyesített reláció részeként vesszük az **R** reláció sorait, majd az **S** relációból hozzáírjuk azokat a sorokat, amelyek nem fordultak elő az **R**-ben.

Eredményül az alábbi reláció adódik:

$R \cup S$	A	B	C
	2	-2	3
	4	1	4
	1	2	4
	2	-2	1
	4	1	2

HALMAZMŰVELETEK

METSZET \rightarrow **Jelölés:** \cap .

Az **R** és **S** relációk metszete eredményül olyan relációt ad, amely azokat a sorokat tartalmazza, amelyek az **R** relációban és az **S** relációban is előfordulnak.

R	A	B	C	és	S	A	B	C
	2	-2	3			1	2	4
	4	1	4			2	-2	1
	1	2	4			4	1	4
						4	1	2

Példa: Végezzük el az előző relációkon a metszet műveletet!

Megoldás: a relációk metszete:

$R \cap S$	A	B	C
	4	1	4
	1	2	4

HALMAZMŰVELETEK

KÜLÖNBSÉG → **Jelölés:** \setminus .

- Az **R** és **S** relációk különbsége eredményül olyan relációt ad, amely tartalmazza az R azon sorait, amelyek az S relációban nem fordulnak elő.

R	A	B	C	és	S	A	B	C
	2	-2	3			1	2	4
	4	1	4			2	-2	1
	1	2	4			4	1	4
						4	1	2

R \ S	A	B	C
	2	-2	3

➤ **S \ R ?**

RELÁCIÓS MŰVELETEK

VETÍTÉS → **Jelölés**: $\pi_L(R)$, ahol **L** attribútum listát jelent

Az **R** reláció vetítésével olyan relációt hozunk létre, amely az **R** bizonyos oszlopait tartalmazza valamilyen sorrendben.

R	A	B	C
	-2	-2	3
	4	6	4
	1	2	4
	3	1	5
	-1	7	2

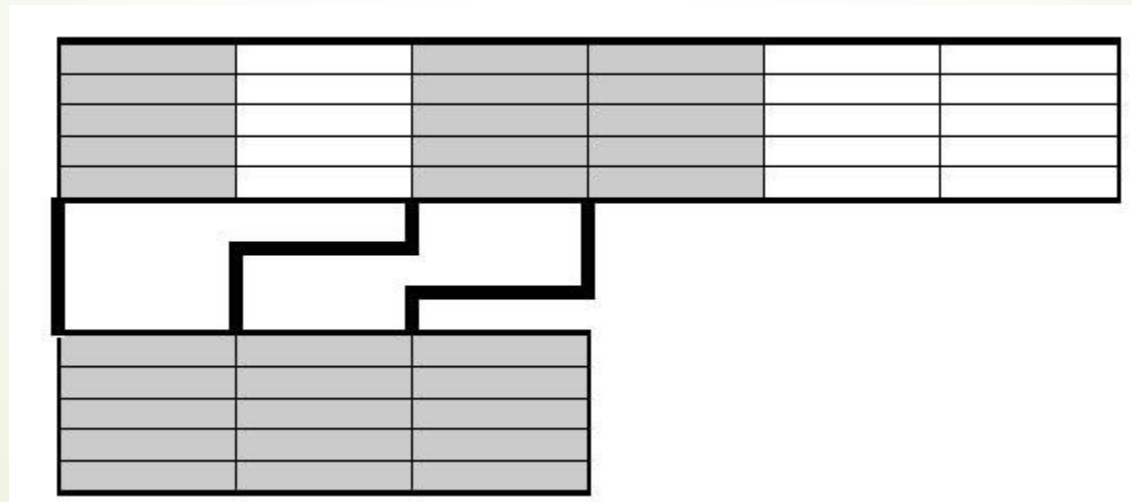
Megoldás: a vetítés eredménye:

$\pi_{B,C}(R)$	B	C
	-2	3
	6	4
	2	4
	1	5
	7	2


A projekció

megvalósítása SQL-ben:

```
SELECT attribútumlista FROM R;
```



A projekció során egy reláció oszlopai közül csak bizonyosakat őrzünk meg az eredmény relációban.



Kiválasztás (szelekció)

Kiválasztjuk a rekordoknak egy olyan részhalmazát, amely eleget tesz egy szelekciós feltételnek.

Tekinthetünk a szelekcióra úgy is, mint egy szűrésre, amely csak azokat a rekordokat tartja meg, melyek kielégítenek egy minősítő feltételt.

Két halmazként is nézhetjük: Lesznek rekordok amik kielégítik a feltételt, és lesznek melyek nem

RELÁCIÓS MŰVELETEK

KIVÁLASZTÁS (szelekció) →

Jelölés (kis szigma): $\sigma_F(R)$,
ahol **F** jelenti a feltételt.

- Az **R** relációra alkalmazott kiválasztás művelet olyan relációt eredményez, amely az **R** sorainak részhalmazát adja. A sorok kiválasztásához feltételt (logikai kifejezést) kell megfogalmaznunk.
- A feltételben: konstansok, attribútum nevek, operátorok (pl.: +, -, AND), theta-relációk (<, >, =, ≤, ≥, ≠) fordulhatnak elő.
- Például **F** lehet: **A > 1 AND B < 4**, ahol A és B attribútumokat jelent. A σ művelet az **R** reláció minden sorára megvizsgálja, hogy az **F** feltétel teljesül-e (behelyettesíti a kifejezésbe az attribútumok soron levő értékeit). Ha az adott soron teljesül a feltétel, akkor az bekerül az eredmény relációba, egyébként nem.

R	A	B	C
	-2	-2	3
	4	6	4
	1	2	4
	3	1	5
	-1	7	2
	3	3	3

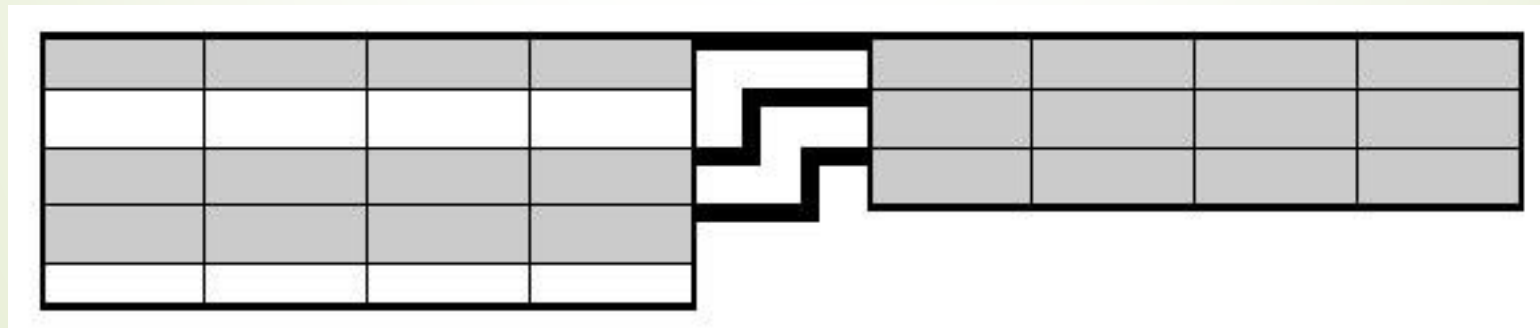
Megoldás →

$\sigma_{A > 1 \text{ AND } B < 4}(R)$	A	B	C
	3	1	5
	3	3	3

A KIVÁLASZTÁS (szelekció)

megvalósítása SQL-ben:

```
SELECT * FROM R WHERE szelekciós feltétel;
```



A szelekció művelete során egy relációból csak egy adott feltételt kielégítő sorokat őrizzük meg az eredmény relációban.



Tehát, vegyük észre

- **A projekció:** bizonyos oszlopokat választ ki a táblából, míg a többi oszlopot elveti.

```
SELECT attribútumlista FROM R;
```

- **A szelekció:** néhány sort választ ki a táblából, míg a többi sort elveti

```
SELECT * FROM R WHERE szelekciós feltétel;
```


RELÁCIÓS MŰVELETEK

Descartes-szorzat (egyszerű szorzat) → **Jelölés**: $R \times S$.

- Az **R** és **S** reláció **Descartes-szorzata** (egyszerűen szorzata) eredményül olyan relációt ad, amely az összes lehetséges módon párosított **R**-beli és **S**-beli sorokat tartalmazza.

Az eredmény reláció sémája:

- az **R** és **S** sémájának egyesítése.
- Az **R** attribútumai megelőzik az **S** attribútumait.
- Azonos attribútum nevek esetén vagy átnevezünk, vagy ún. minősített nevet használunk.
pl. **R.A** jelöli az **R**-beli A attribútumot, **S.A** az **S**-belit.

Descartes-szorzat

- ▶ A Descartes szorzat két reláció sorait minden kombinációban egymás mellé teszi az eredmény relációban.

1		
2		
3		
4		

1			
2			
3			

1			1			
2			2			
3			3			
4			1			
1			2			
2			3			
3			1			
4			2			
1			3			
2			1			
3			2			
4			3			

RELÁCIÓS MŰVELETEK


Descartes-szorzat (egyszerű szorzat) → **Jelölés:** $R \times S$.

Példa: Adottak az alábbi R és S relációk előfordulásai. Határozzuk meg az R és S relációk $R \times S$ szorzatát!

R	A	B	C	és	S	A	D	E
	2	a	3			1	2	e
	4	b	4			2	-2	f
	1	c	4			4	1	g

Megoldás: Az A attribútum előfordul mindegyik relációban, ezért minősítve az eredmény relációban az R.A és S.A neveket használjuk:

$R \times S$	R.A	B	C	S.A	D	E
	2	a	3	1	2	e
	2	a	3	2	-2	f
	2	a	3	4	1	g
	4	b	4	1	2	e
	4	b	4	2	-2	f
	4	b	4	4	1	g
	1	c	4	1	2	e
	1	c	4	2	-2	f
	1	c	4	4	1	g



Descartes-szorzat megvalósítása SQL-ben

```
SELECT * FROM auto, ember;
```

vagy:

```
SELECT * FROM auto CROSS JOIN ember;
```

SZÁRMAZTATOTT MŰVELETEK

Természetes összekapcsolás \rightarrow Jelölés: \bowtie

- ▶ Gyakorlatiasabb a logikailag összetartozó sorok kezelése
- ▶ A **természetes összekapcsolás** műveletében az **R** és **S** relációk azon sorait illesszük egy sorba, amelyek **megegyeznek** az **R** és **S** sémájának összes **közös** attribútumán
Legyenek **R** és **S** közös attribútumai: A és B.
A művelet elvégzéséhez az alábbi lépéseket kell végrehajtani:
 1. Elkészítjük **R** és **S** **Descartes-szorzatát**.
 2. Végrehajtjuk a (kis szigma) $\sigma_{R.A=S.A}$ ($R \times S$) műveletet (**kiválasztás** adott feltétellel: csak a logikailag összetartozó **R**-beli és **S**-beli sorok maradnak).
 3. Végrehajtjuk a $\pi_{A, B, \dots, H, K, \dots}$ ($\sigma_{R.A=S.A} (R \times S)$) **vetítés** műveletet (az **R** sémája (A, B, C, ...), az **S** sémája (A, B, H, K, ...)). Az azonos attribútumú oszlopok közül elhagyjuk az **S**-belieket.

SZÁRMAZTATOTT MŰVELETEK

Természetes összekapcsolás \rightarrow Jelölés: \bowtie

Példa: Adottak az alábbi R és S relációk előfordulásai. Határozzuk meg az R és S relációk $R \bowtie S$ természetes összekapcsolását!

R	A	B	C	és	S	A	D	E
	2	a	3			1	2	e
	4	b	4			2	-2	f
	1	c	4			4	1	g
						4	1	d

1. Elkészítjük **R** és **S** Descartes-szorzatát.
(minősítjük az A oszlopokat: R.A, S.A)

$R \times S$	R.A	B	C	S.A	D	E
	2	a	3	1	2	e
	2	a	3	2	-2	f
	2	a	3	4	1	g
	2	a	3	4	1	d
	4	b	4	1	2	e
	4	b	4	2	-2	f
	4	b	4	4	1	g
	4	b	4	4	1	d
	1	c	4	1	2	e
	1	c	4	2	-2	f
	1	c	4	4	1	g
	1	c	4	4	1	d

SZÁRMAZTATOTT MŰVELETEK

Természetes összekapcsolás \rightarrow Jelölés: \bowtie

$R \times S$	R.A	B	C	S.A	D	E
2	a	3	1	2	e	
2	a	3	2	-2	f	
2	a	3	4	1	g	
2	a	3	4	1	d	
4	b	4	1	2	e	
4	b	4	2	-2	f	
4	b	4	4	1	g	
4	b	4	4	1	d	
1	c	4	1	2	e	
1	c	4	2	-2	f	
1	c	4	4	1	g	
1	c	4	4	1	d	

1. Elkészítjük R és S Descartes-szorzatát.
(minősítjük az A oszlopokat: R.A, S.A)

2. Végrehajtjuk a (kis szigma) $\sigma_{R.A=S.A} (R \times S)$
(Kiválasztjuk a logikailag összetartozó sorokat.)

$\sigma(R \times S)$	R.A	B	C	S.A	D	E
2	a	3	2	-2	f	
4	b	4	4	1	g	
4	b	4	4	1	d	
1	c	4	1	2	e	

SZÁRMAZTATOTT MŰVELETEK

Természetes összekapcsolás \rightarrow Jelölés: \bowtie

$\sigma(R \times S)$	R.A	B	C	S.A	D	E
	2	a	3	2	-2	f
	4	b	4	4	1	g
	4	b	4	4	1	d
	1	c	4	1	2	e

1. Elkészítjük **R** és **S** Descartes-szorzatát.
(minősítjük az A oszlopokat: R.A, S.A)
2. Végrehajtjuk a $\sigma_{R.A=S.A}$ ($R \times S$) műveletet.
(Kiválasztjuk a logikailag összetartozó sorokat.)
3. Végrehajtjuk az alábbi műveletet
 $\pi_{A, B, \dots, H, K, \dots} (\sigma_{R.A=S.A} (R \times S))$
Vetítéssel elhagyjuk az S.A oszlopot

A feladat képletesen:

$$R \bowtie S = \pi_{A, B, C, D, E} (\sigma_{R.A=S.A} (R \times S))$$

R \bowtie S	A	B	C	D	E
	2	a	3	-2	f
	4	b	4	1	g
	4	b	4	1	d
	1	c	4	2	e

SZÁRMAZTATOTT MŰVELETEK

Példa: Adottak az alábbi R és S relációk előfordulásai. Határozzuk meg az R és S relációk $R \bowtie S$ természetes összekapcsolását!

R	A	B	C	és	S	A	D	E
	2	a	3			1	2	e
	4	b	4			2	-2	f
	1	c	4			4	1	g
						4	1	d

A feladat képletesen: $R \bowtie S = \pi_{A,B,C,D,E} (\sigma_{R.A=S.A} (R \times S))$

$R \bowtie S$	A	B	C	D	E
	2	a	3	-2	f
	4	b	4	1	g
	4	b	4	1	d
	1	c	4	2	e

D-szorzat + kiválasztás + vetítés

SZÁRMAZTATOTT MŰVELETEK

Théta-összekapcsolás \rightarrow Jelölés: $\mathbf{R} \bowtie_{\mathbf{F}} \mathbf{S}$

Néha a *természetes összekapcsolásnál* „általánosabb” illesztésre is szükségünk van. Azaz nemcsak az „=” relációt engedjük meg, hanem bármely Théta-operátort (<, >, =, \leq , \geq , \neq) az \mathbf{R} -beli és \mathbf{S} -beli attribútumok között.

Az \mathbf{R} és \mathbf{S} relációk *théta-összekapcsolása* az $\mathbf{R} \times \mathbf{S}$ relációból kiválasztja azokat a sorokat, amelyek megfelelnek egy Θ (théta) feltételnek. A Θ helyett a könnyebb írásmód miatt \mathbf{F} -et használunk. Jelölés: $\mathbf{R} \bowtie_{\mathbf{F}} \mathbf{S}$, ahol az \mathbf{F} feltétel konstansokat, attribútum neveket, théta-operátorokat, logikai műveleteket tartalmazhat.

A természetes összekapcsolás képletesen:

$$\mathbf{R} \bowtie \mathbf{S} = \pi_{A,B,C,D,E} \left(\sigma_{R.A=S.A} (\mathbf{R} \times \mathbf{S}) \right)$$



Théta-összekapcsolás megvalósítása SQL-ben

```
SELECT * FROM R, S  
WHERE összekapcsolási feltétel;
```

Házi feladat:

Adottak az R és S relációk alábbi előfordulásai.

R	A	B	C	és	S	A	D	E
	2	a	3			1	2	e
	4	b	4			2	-2	b
	1	c	4			4	1	g
						4	1	c

Feladat: Határozzuk meg az R és S feladat Théta-összekapcsolását ($R \bowtie_F S$), ahol
 $F = R.A < S.A \text{ OR } B=E$

MULTIHALMAZOK - halmazműveletek

SQL-ben a relációk multihalmazok, azaz egy sor többször is megjelenhet egy SQL-relációban (lekérdezés eredménytáblában), ezért kiegészíthetjük a lekérdezést új műveletekkel:

- csoportosítás: GROUP BY
- összesítések: SUM, AVG, MIN, MAX

Példa: A megadott R reláción végezzük el az alábbi műveletet:

(gamma) $\gamma_{A, \text{SUM}(B \cdot C) \rightarrow X}(R)$

Megoldás:

$\gamma_{A, \text{SUM}(B \cdot C) \rightarrow X}(R)$	A	X
	1	32
	2	24
	4	32

R	A	B	C	D
	2	3	2	e
	2	3	4	f
	2	3	1	g
	2	3	1	d
	4	4	2	e
	4	4	4	f
	4	4	1	g
	4	4	1	d
	1	4	2	e
	1	4	4	f
	1	4	1	g
	1	4	1	d

KULCSOK

Az attribútumok egy halmaza egy kulcsot alkot egy relációra nézve, ha a reláció előfordulásaiban nincs két olyan sor, amelyek a kulcs összes attribútumának értékein (komponensein) megegyeznének.

► Kulcsok: elsődleges kulcs (PK), idegen kulcs (FK)

- a kulcs a relációnak egy és csakis egy sorát határozza meg.
- a kulcsnak nincs olyan részhalmaza, amely szintén kulcs lenne.
- a kulcs tulajdonságok nem lehetnek nulla-értékűek.
- minden relációnak van kulcsa.
- egy relációnak több kulcsa is lehet.

GÉPJÁRMŰ (Rendszám, Alvászsám, gjMárkaID, gyártmány, gjSzín)

Vagy

GJMÁRKA (gjMárkaID, gjMárkaNév)

vagy

FILMEK (Filmcím, Év, Hossz, Műfaj, Szereplők, Rendező)

FUNKCIONÁLIS FÜGGŐSÉGEK

- ▶ Jelölés: Adott $R(A_1, \dots, A_n)$ reláció.

X, Y attribútum halmazok $\Rightarrow X, Y \subseteq R$

$X = (A_{i_1}, A_{i_2}, \dots, A_{i_k})$ helyett $X = A_{i_1} A_{i_2} \dots A_{i_k}$

$Y = (A_{j_1}, A_{j_2}, \dots, A_{j_l})$ helyett $Y = A_{j_1} A_{j_2} \dots A_{j_l}$

- ▶ **Definíció:**

$Y \subseteq R$ **funkcionálisan függ** $X \subseteq R$ -től, (jelölés: $X \rightarrow Y$), ha R bármely két sorára igaz, hogy ha megegyeznek X -en, akkor Y -on is megegyeznek ($t_1[X]=t_2[X] \Rightarrow t_1[Y]=t_2[Y]$).

Más szavakkal a funkcionális függőség azt jelenti, hogy a tulajdonságok által meghatározott rendszerben **egy tulajdonság egyértelműen meghatároz egy vagy több másik tulajdonságot.**

Példa:

- ▶ Legyen az R reláció: R (Név, Irányítószám, Település, Utca, Telefon)
- ▶ Adjunk meg funkcionális függéseket!
 - ▶ Település \rightarrow Irányítószám
 - ▶ Irányítószám \rightarrow Település
 - ▶ (vonalas)Telefon \rightarrow Név, Cím
 - ▶ Számlaszám \rightarrow vevői és eladói adatok

FUNKCIONÁLIS FÜGGŐSÉGEK

Armstrong három axiómája a funkcionális függőségről:

Reflexivitás:

- Ha ugyanazon R reláción $X \subseteq Y$, akkor $Y \rightarrow X$. Szokás triviális függőségnek nevezni.

Tranzitivitás:

- Ha ugyanazon R reláción $X \rightarrow Y$ és $Y \rightarrow Z$, akkor $X \rightarrow Z$.

Bővíthetőség:

- Ha ugyanazon R reláción $X \rightarrow Y$, akkor $XZ \rightarrow YZ$



Köszönöm a figyelmet!