NAGY ISTVÁN

Relációalgebra

Áttekintés

2010

Tartalomjegyzék

Relációalgebrai műveletek

Az alábbiakban – a multiműveletek és redukáló műveletek alkalmazásaként, esetenként kiterjesztéseként – bemutatjuk a relációk gyakorlatban megvalósított műveleteit. E műveletek segítségével egy, vagy két relációból egy újat tudunk létrehozni. Ezek tehát relációtranszformációk, melyek az úgynevezett *paraméter-relációkat* leképezik (általában) egy úgynevezett *eredmény-relációba*. Azokat a műveleteket, melyek egy paraméter-relációt tartalmaznak *unárisnak*, amelyek kettőt, azokat *binárisnak* nevezzük.

Ezek egyik típusa (redukálás, metszetképzés, különbségképzés, egyesítés, Descartes-szorzás, osztás) a relációkon, mint halmazokon végez alapvetően halmazalgebrai műveleteket, egy másik típusba tartozók (projekció, dekompozíció, szelekció) a reláción, mint táblán hajtanak végre valamilyen (sor, vagy oszlop) kiválasztási transzformációt. Harmadik típust (természetes és általános összekapcsolás) azok a relációs műveletek alkotják, amelyek olyan, összetett relációs műveleteket végeznek, melyek az előző két típus tulajdonságát ötvözik, azaz két relációból úgy hoznak létre egy harmadikat, hogy a kiinduló tábláknak csak bizonyos feltételeket kielégítő sorait veszik bele a műveletbe. Ez utóbbi műveletek már rendkívül összetett kiválasztási és lekérdezési lehetőségeket biztosítanak, a gyakorlati adatbázis-kezelő (például az SQL nyelvet megvalósító) szoftverek alapvető eszközkészletét alkotják. Végül a negyedik típust az úgynevezett osztályozó műveletek alkotják, melyek a paraméter-reláció bizonyos tulajdonságú rekordjait helyettesítik egyetlen reprezentáns rekorddal, miközben bizonyos, jellemzően aritmetikai függvényeket határoznak meg.

E műveletek közül érdemes külön kiemelnünk a multiunió a projekció és a dekompozíció műveletét, mint *multihalmazképző műveletet*. Ez alatt azt értjük, hogy e műveletek eredménye (ellentétben a metszettel, a különbség képzéssel, vagy például a szelekcióval) akkor is lehet multihalmaz, ha a művelet paraméterei monohalmazok.

A gyakorlatban azért fontos számon tartanunk egy műveletről, hogy multihalmazképző művelet-e, mert az adatbázisok konzisztenciájának (ellentmondás-mentességének) fenntartása érdekében a lehető legalacsonyabb szinten szeretjük tartani az adatredundanciát (adatismétlődést), márpedig ez egyértelműen növeli azt. A teljes irredundanciát természetesen egyéb szempontok miatt általában nem tudjuk biztosítani (sőt egyes esetekben – például normalizálások esetén – bizonyos típusú redundanciát mi magunk növelünk), mégis legalább azt elvárnánk, hogy egy adattáblában ugyanaz a rekord (sor) ne ismétlődjön. Adatfelvitel esetén (mint már említettük) ezt a fejlesztők által írott programok megakadályozzák, azonban bizonyos adatlekérdezési tevékenységek – mégpedig éppen a multihalmazképző művelet – során a keletkezett táblák között lehetnek multihalmazok. Mit tegyünk ezekkel?

Az egyik lehetőség, hogy megakadályozzuk az ilyen táblák létrejöttét azáltal, hogy a táblák keletkezésekor megtiltjuk a sorok többszöröződését. Erre az SQL-ben például a `SELECT` parancs `DISTINCT` paramétere ad lehetőséget. Ennek használata azonban rendkívül lelassítja a műveletvégzést, mivel minden egyes új sor generálásakor össze kell azt hasonlítani az összes már előállított sorral. Ám, ha benne hagyjuk a sorismétlődéseket, akkor ez hibát is okozhat, vagy megtévesztheti a felhasználót még akkor is, ha egy rendezés révén az azonos sorok egymás alá kerülnek (mivel esetleg nem veszi észre)!

A másik lehetőség, hogy meghagyjuk a feldolgozás során a sorismétlődéseket, és csak akkor töröljük, amikor az ismétlődés már feldolgozási hibát okozna (például leszámlálás esetén), vagy amikor a végső felhasználói listát állítjuk elő. Igaz a többlet sorok némileg megnövelik a műveletvégzési időt, de közel sem annyira, mintha minden elemi művelet után törölnénk őket. Ráadásul a különböző műveletek eredménytáblái általában kisebbek az operandusok tábláinál, és így a többszörös elemek kiszűrése már eleve gyorsabb, mint azokban. (Persze van kivétel, például a táblák egyesítése, vagy szorzata, ahol az eredménytábla a nagyobb, ám ezeket a műveleteket elég ritkán használjuk.)

A fentiek tanulsága tehát az, hogy ne féljünk a sor-többszörözéstől, ám tartsuk számon, hogy mikor léphet fel: multihalmazképző műveletek esetén. Ezek közül a legfontosabb a projekció, mivel ezt szinte minden lekérdezésben használjuk. E műveletnél esetenként nem is az lesz a legnagyobb gondunk, hogy azonos sorok keletkeznek, hanem az, hogy nem tudjuk azonosítani a keletkezett sorokat az eredeti tábla soraival. Ennek érdekében olyan sorazonosító adatra kell majd hivatkoznunk, melyet sem az algebrai leírás, sem az SQL nem tartalmaz, azonban a gyakorlatban minden adatbázis-kezelő szoftver használ. Az Oracle-ben ezt a sorazonosítót `ROWID`-nek nevezik.

Az alábbi tárgyalás során a relációalgebrai műveleteket aszerint fogjuk csoportosítani, hogy a paraméter-relációk adattábláin milyen jellegű változtatásokat hajtanak végre. Eszerint beszélünk szerkezettartó, szerkezetmódosító és osztályozó műveletekről.

Végül még két megjegyzés. 1.) A „relációalgebrai műveletek” nem mindegyike művelet algebrai értelemben, ezeket elvégezhető számítássorozatként, algoritmusként értelmezzük. 2.) Az egyes műveletek definiálásakor nem foglalkozunk az úgynevezett NULL-értékekkel, ezek a III.rész gyakorlati tárgyalásában fognak majd szerepelni.

Szerkezettartó műveletek

Az ebbe a csoportba tartozó relációs műveletek a paraméter-relációk szerkezetét nem változtatják meg, a bináris műveletek esetén azonban úgynevezett *kompatibilitási követelményeket* támasztanak azokkal szemben.

Ez a programozási gyakorlatban egyrészt azt jelenti, hogy a paraméter-relációkban meg kell egyezni az attribútumok számának, valamint a paraméter-relációk attribútumvektoraiban páronként az egyes attribútumok neveinek és adattípusainak is (mivel ez az attribútumvektor „öröklődik” át az eredményrelációba). Az esetleges név-eltérést az átnevező függvénnyel, a típuseltérést pedig az adatkonvertáló függvény alkalmazásával lehet kiküszöbölni. Az alábbiakban feltételezzük, hogy ezek a kompatibilitási követelmények teljesülnek.

Reláció átnevezése, alias név, másodlagos attribútumnév

A relációkat esetenként úgynevezett *alias-névvel* látjuk el. Ez nem jelenti azt, hogy a relációt reprezentáló tábla „fizikailag tárolt” nevét megváltoztatjuk. Egy alias-név érvényességi köre általában nem is terjed túl egy reláció-műveleten, és a legfontosabb funkciója az egy műveletben szereplő paraméter-relációk érthetőbb megnevezése, vagy jobb megkülönböztetése. Hasonló okokból látjuk el egy reláció egyes attribútumneveit úgynevezett másodlagos attribútumnévvel. Ebben az esetben is a reláció „fizikai” attribútumneve nem változik meg, csak éppen az adott műveleteken belül lehetőség van e másodlagos néven történő hivatkozásra. Az alábbiakban ezeket az átnevezéseket formálisan is definiáljuk.

Legyen $R\langle \underline{A} \rangle$ egy reláció az A attribútumhalmaz felett. Ennek $S\langle \underline{B} \rangle$ alakra való átnevezését a $Ren_{S\langle \underline{B} \rangle}(R\langle \underline{A} \rangle)$ *átnevezőfüggvény* végzi el. Ha csak a reláció nevét akarjuk megváltoztatni, akkor a $Ren_S(R)$, ha csak az attribútumokét, akkor a $Ren_{R\langle \underline{B}\rangle}(R\langle \underline{A}\rangle)$ alakot használjuk. A *Ren* függvény tehát mindössze egy átjelölést végez el.

Ha egy $R\langle A_1,\ldots, A_n\rangle$ relációban valamely $A_i \in I_n$ attribútumot akarunk átnevezni B_i alakra, akkor használhatjuk az „**AS**” *átnevező-operátort* is A_i **AS** B_i módon. Ha csak az attribútumokat, vagy azok közül is csak néhányat akarunk átnevezni, akkor az átnevező-operátor használata egyszerűbb.

Reláció redukálása

Legyen $R\langle \underline{A}\rangle$ egy reláció az A attribútumhalmaz felett. Ha biztosítani akarjuk, hogy $R\langle \underline{A}\rangle$ ne tartalmazzon rekordismétlést, akkor alkalmazzuk rá a multihalmazoknál bevezetett *redukáló függvényt*. A $Reduc(R\langle \underline{A}\rangle)$ már nyilván monoreláció.

Relációk metszete, különbsége és egyesítése

Legyen $R\langle \underline{A}\rangle$ és $S\langle \underline{A}\rangle$ két reláció ugyanazon A attribútumhalmaz felett. (Ha $S = S\langle \underline{B}\rangle$ és $B \neq A$, de $\mu(B) = \mu(A)$, akkor műveletvégzés előtt $S = Ren_{S\langle \underline{A}\rangle}(S\langle \underline{B}\rangle)$ módon át kell nevezni az attribútumokat. A gyakorlatban természetesen oda kell figyelni arra is, hogy az összetartozó A_i és B_i attribútumok azonos típusúak legyenek (lásd kompatibilitási követelmények).

Ekkor értelmezhetjük az $R\langle \underline{A}\rangle \cap^{\mathrm{R}} S\langle \underline{A}\rangle$ relációmetszetet, mint multihalmazok redukált metszetét, az $R\langle \underline{A}\rangle \setminus^{\mathrm{R}} S\langle \underline{A}\rangle$ relációkülönbséget, mint multihalmazok redukált különbségét, valamint az $R\langle \underline{A}\rangle \cup^{\mathrm{R}} S\langle \underline{A}\rangle$, illetve az $R\langle \underline{A}\rangle \cup^{\mathrm{M}} S\langle \underline{A}\rangle$ relációegyesítéseket, mint multihalmazok redukált egyesítését, illetve multiunióját.

Példa

Metszet	X	Y		X	Y		X	Y
	1	2	$\cap^{\mathrm{R}}$	1	2	=	1	2
	3	4		7	8			
	1	2		1	2			

Különbség	X	Y		X	Y		X	Y
	1	2	$\setminus^{\mathrm{R}}$	1	2	=	1	2
	3	4		7	8		3	4
	1	2						

Redukált	X	Y		X	Y		X	Y		X	Y
egyesítés	1	2	$\cup^{\mathrm{R}}$	1	2	$\cup^{\mathrm{R}}$	1	2	=	1	2
	3	4		5	6					3	4
										5	6

Multiunió $\begin{array}{cc} X & Y \\ \hline 1 & 2 \\ 3 & 4 \end{array} \uplus \begin{array}{cc} X & Y \\ \hline 1 & 2 \\ 5 & 6 \end{array} \uplus \begin{array}{cc} X & Y \\ \hline 1 & 2 \end{array} = \begin{array}{cc} X & Y \\ \hline 1 & 2 \\ 3 & 4 \\ 1 & 2 \\ 5 & 6 \\ 1 & 2 \end{array}$

Láthatóan a multiunió a monorelációkból ezúttal valódi multirelációt állított elő.

A multiunió kivételével a többi halmazművelet láthatóan kiszűrte a rekordismétlődéseket.

Szelekció (kiválasztás, szűrés)

Egy A attribútumhalmaz felett *értelmezett* $R\langle \underline{A}\rangle$ relációnak ugyanezen A halmazon értelmezett T_A logikai függvényre vonatkozó *szelekciója* (más néven *kiválasztása*, vagy *szűrése*) az alábbi reláció:

$$\sigma_{T_A}(R\langle \underline{A}\rangle) \triangleq \{ \underline{r} \mid \underline{r}\in R\langle \underline{A}\rangle,\ T_A \} .$$

Megjegyzés

A $\sigma_{T_A}(R\langle \underline{A}\rangle)$ szelekció tehát a T_A logikai feltételnek megfelelő sorokat választja ki az $R_{\underline{A}}$ táblából (illetőleg a nem megfelelőeket kiszűri).

Példa

$R\langle X, Y\rangle$: $\begin{array}{cc} X & Y \\ \hline 1 & 2 \\ 8 & 7 \\ 1 & 2 \end{array}$ $\quad \sigma_{(Y>X)}(R\langle X, Y\rangle)$: $\begin{array}{cc} X & Y \\ \hline 1 & 2 \\ 1 & 2 \end{array}$

A szelekció csak akkor eredményezhet valódi multirelációt, ha az operandusa is az.

Szerkezetmódosító műveletek

A szerkezetmódosító műveletek esetén az eredmény-reláció felépítése eltér a paraméter-reláció(k)étól, ilyenmódon *algebrai értelemben nem is tekinthetők műveletnek*. Ezek közül a legegyszerűbb a projekció, mely voltaképpen a paraméter-reláció kijelölt attribútumait (a tábla kijelölt oszlopait) tartja meg, esetleg rendezi át. Jelentőségét az adja, hogy az egyik legtipikusabb kiválasztási feladat, a listakészítés ezen alapszik. Mivel multihalmazt hoz létre, így alkalmazásakor fokozott jelentősége van a redukálási lehetőségnek.

Projekció (vetítés)

Valamely $R\langle \underline{A}\rangle$ reláció egy $B \subseteq A$ attribútumhalmazra vonatkozó *projekciója* (más néven *vetítése*) az $R\langle \underline{A}\rangle$ reláció $\underline{B}$-alrelációja, azaz

$$\pi_{\underline{B}}(R\langle \underline{A}\rangle) \triangleq R\langle \underline{A}\rangle\|_{\underline{B}} .$$

Megjegyzés

A projekció tehát az $R_{\underline{A}}$ táblából kiválasztja a B-beli oszlopokat, mégpedig a $\underline{B}$-beli sorrendnek megfelelően.

Példa

$R\langle X, Y, Z\rangle$:

X	Y	Z
1	2	5
3	4	6
1	2	7
1	2	8

$\pi_{\langle Y, X\rangle}(R\langle X, Y, Z\rangle)$:

Y	X
2	1
4	3
2	1
2	1

$Reduc(\pi_{\langle Y, X\rangle}(R\langle X, Y, Z\rangle))$:

Y	X
2	1
4	3

E példában a projekció is valódi multirelációt állított elő a monorelációból. Ezt azonban szükség esetén redukálhatjuk (például listakészítéskor) a *Reduc* függvény segítségével.

Dekompozíció

Legyen A egy attribútumhalmaz, továbbá $B = \{\underline{B}_1, \underline{B}_2, \ldots, \underline{B}_n\}$ egy attribútumvektor-halmaz, ahol minden $\underline{B}_i \in B$ esetén $B_i \subseteq A$ (tehát bármely $\underline{B}_i, \underline{B}_j \in B$ esetén $B_i \cap B_j \neq \emptyset$ megengedett), és

$$\bigcup_{\underline{B}_i \in B} B_i \;=\; A\,.$$

Ekkor valamely $R\langle \underline{A}\rangle$ relációnak a B attribútumvektor-halmazra vonatkozó *dekompozíciója*

$$R\langle \underline{A}\rangle / B \triangleq \{\, R_i \,\big|\, R_i = R\langle \underline{A}\rangle\|_{\underline{B}_i},\; \underline{B}_i \in B \,\},$$

ahol természetesen minden $\underline{B}_i \in B$ esetén B_i a $\underline{B}_i$ vektor alaphalmaza. (A dekompozíció tehát projekciók halmaza.)

Példa

$R\langle \underline{A}\rangle$:

P	R	S	T	U
1	5	9	4	3
3	6	8	4	3
2	5	7	5	4
3	6	8	4	3
1	5	9	5	3

Legyen $\underline{A} = \langle P, R, S, T, U\rangle$, $B = \{\underline{B}_1, \underline{B}_2, \underline{B}_3\}$, ahol $\underline{B}_1 = \langle U, P, R\rangle$, $\underline{B}_2 = \langle P, R, S\rangle$, $\underline{B}_1 = \langle S, T, U\rangle$, továbbá $R\langle \underline{A}\rangle$ a baloldali táblázattal adott reláció.
Határozzuk meg az $R\langle \underline{A}\rangle / \underline{B}$ dekompozíciót.

Ekkor $R\langle \underline{A}\rangle / B = \{\, R\langle \underline{A}\rangle\|_{\underline{B}_1}, R\langle \underline{A}\rangle\|_{\underline{B}_2}, R\langle \underline{A}\rangle\|_{\underline{B}_3} \}$, ahol

$R\langle \underline{A}\rangle\|_{\underline{B}_1}$:

U	P	R
3	1	5
3	3	6
4	2	5
3	3	6
3	1	5

$R\langle \underline{A}\rangle\|_{\underline{B}_2}$:

P	R	S
1	5	9
3	6	8
2	5	7
3	6	8
1	5	9

$R\langle \underline{A}\rangle\|_{\underline{B}_3}$:

S	T	U
9	4	3
8	4	3
7	5	4
8	4	3
9	5	3

Megjegyzés

A fenti példához hasonlóan a gyakorlatban a dekompozíciót „átfedéssel” valósítjuk meg, tehát a származtatott relációknak vannak közös attribútumaik. E közös attribútumok használatának célja általában az, hogy az egyes rekordok azonosíthatósága fennmaradjon.

Figyeljünk fel arra, hogy a származtatott relációk rekordjainak száma azonos, és megegyezik az eredeti reláció rekordjainak számával. (Ha a relációkat hipervektorokként definiáltuk volna, akkor ennél erősebb állítás is kimondható lenne.) Látható továbbá, hogy a dekompozíció szintén képes valódi multirelációt előállítani monorelációból.

Descartes-szorzás (szorzás)

Valamely $R\langle \underline{A} \rangle$ és $S\langle \underline{B} \rangle$ relációk *Descartes-szorzata* (vagy egyszerűen csak *szorzata*):

$$R\langle \underline{A} \rangle \times S\langle \underline{B} \rangle \triangleq \{\, \underline{r} \mid \underline{r} \in \mathbf{D}\langle \underline{C} \rangle,\ \underline{C} = \underline{A} \dot{+} \underline{B} \,\} ,$$

ahol $\dot{+}$ a pontozott láncolás műveletjele.

Megjegyzés

Nyilvánvalóan a relációk Descartes-szorzatának táblája annyi sorból fog állni, mint a paraméter-relációk tábláiban lévő sorok szorzata, azaz

$$\mu(R\langle \underline{A} \rangle \times S\langle \underline{B} \rangle) = \mu(R\langle \underline{A} \rangle) * \mu(S\langle \underline{B} \rangle) .$$

A pontozott láncolásnál bevezetett minősített névhasználattól a relációalgebrai gyakorlatban annyiban szoktunk eltérni, hogy minősített elemneveket csak a közös elemek esetén használunk.

Egy reláció önmagával való szorzásakor – szintén az attribútumok megkülönböztethetősége érdekében – az egyik operandust egy alias-névvel helyettesítjük.

Példa

$R\langle X, Y \rangle$:

X	Y
1	2
3	4

×

$S\langle Y, U, V \rangle$:

Y	U	V
5	6	7
8	9	10
11	12	13

=

X	$R.Y$	$S.Y$	U	V
1	2	5	6	7
1	2	8	9	10
1	2	11	12	13
3	4	5	6	7
3	4	8	9	10
3	4	11	12	13

Példa

$R\langle X, Y \rangle$:

X	Y
1	2
1	2

×

$S\langle Y, Z \rangle$:

Y	Z
2	3
4	5
4	5

=

X	$R.Y$	$S.Y$	Z
1	2	2	3
1	2	4	5
1	2	4	5
1	2	2	3
1	2	4	5
1	2	4	5

Láthatóan a Descartes-szorzás is csak akkor eredményez valódi multirelációt, ha az operandus-relációk legalább egyike is az.

Relációk osztása (osztás)

Valamely $R\langle \underline{A} \rangle$ reláció osztható egy $S\langle \underline{B} \rangle$ relációval, ha van olyan $T\langle \underline{C} \rangle$ reláció, melyre

1. $C = A \setminus B$, és
2. $S\langle \underline{B} \rangle \times T\langle \underline{C} \rangle = R\langle \underline{A} \rangle$.

Ekkor az $R\langle \underline{A} \rangle$ és az $S\langle \underline{B} \rangle$ *relációk hányadosa*:

$$R\langle \underline{A} \rangle \div S\langle \underline{B} \rangle \triangleq T\langle \underline{C} \rangle.$$

Könnyen belátható, hogy minden $\underline{r} \in R\langle \underline{A} \rangle$ esetén $\underline{r}\dagger_{(\underline{A \setminus B})} \in R\langle \underline{A} \rangle \div S\langle \underline{B} \rangle$.

Példa

$R\langle X, Y, U, V\rangle$:

X	Y	U	V
1	2	2	3
1	2	4	5
1	2	2	3
1	2	4	5

$\div$ $S\langle U, V\rangle$:

U	V
2	3
4	5

$=$

X	Y
1	2
1	2

Láthatóan a hányados csak akkor eredményezhet valódi multirelációt, ha az osztó is az. Az alábbi példából viszont az látható, hogy az osztó multireláció jellege önmagában még nem elég ahhoz, hogy a hányados is az legyen.

Példa

$R\langle X, Y, U, V\rangle$:

X	Y	U	V
1	2	2	3
1	2	4	5
1	2	2	3
1	2	4	5

$\div$ $S\langle Y, U, V\rangle$:

Y	U	V
2	2	3
2	4	5
2	2	3
2	4	5

$=$

X
1

Természetes összekapcsolás (Natural Join)

Legyen $R\langle \underline{A}\rangle$ és $S\langle \underline{B}\rangle$ relációk esetén $A \cap B = X$, és $X \neq \emptyset$ (vagyis legyenek az A és B halmazoknak közös attribútumaik, és ezeket ne különböztessük meg aszerint, hogy melyik halmazból vettük). Ekkor $R\langle \underline{A}\rangle$ és $S\langle \underline{B}\rangle$ relációk *természetes összekapcsolása*:

$$R\langle \underline{A}\rangle \bowtie S\langle \underline{B}\rangle \triangleq \{\, \underline{r} \mid \underline{r} \in \mathbf{D}\langle \underline{C}\rangle,\ \underline{C} = \underline{A} \overset{\cup}{+} \underline{B} \,\},$$

ahol $\overset{\cup}{+}$ az egyesítő láncolás műveletének jele.

Megjegyzés

A természetes összekapcsolás táblájában tehát az $R\langle \underline{A}\rangle$ és az $S\langle \underline{B}\rangle$ tábláinak azon sorai vesznek részt, melyek a közös (X-beli) attribútumokon azonos értékeket tartalmaznak.

Ebben az esetben tehát nem használunk minősített neveket, mivel a természetes összekapcsolás lényegét alkotják az azonos attribútumok.

Példa

$R\langle X, Y, Z\rangle$:

X	Y	Z
1	2	3
6	7	8
9	7	8

$\bowtie$ $S\langle Z, Y, U\rangle$:

Z	Y	U
3	2	4
3	2	5
8	7	9

$=$

X	Y	Z	U
1	2	3	4
1	2	3	5
6	7	8	9
9	7	8	9

Példa

$R\langle X, Y\rangle$:

X	Y
1	2
1	2

$\bowtie$ $S\langle Z, Y\rangle$:

Z	Y
3	2
5	4
5	4

$=$

X	Y	Z
1	2	3
1	2	3

A Descartes-szorzáshoz hasonlóan itt is csak akkor lesz az eredmény valódi multireláció, ha az operandus-relációk legalább egyike is az.

Általános összekapcsolás (Théta összekapcsolás, Join)

Legyen $R\langle \underline{A}\rangle$ és $S\langle \underline{B}\rangle$ reláció, $T_{A\dot\cup B}$ pedig egy logikai függvény az $A \dot\cup B$ attribútumhalmazon (ahol $\dot\cup$ a pontozott egyesítés műveletének jele). Ekkor az $R\langle \underline{A}\rangle$ és $S\langle \underline{B}\rangle$ relációknak a $T_{A\dot\cup B}$ logikai függvényre vonatkozó *általános összekapcsolása*:

$$R\langle \underline{A}\rangle \underset{T_{A\dot\cup B}}{\bowtie} S\langle \underline{B}\rangle \triangleq \{\, \underline{r} \mid \underline{r} \in \mathbf{D}\langle \underline{C}\rangle,\ \underline{C} = \underline{A} \dot{+} \underline{B},\ T_{A\dot\cup B} \,\} .$$

Megjegyzés

Az általános összekapcsolás a természetes összekapcsolásban előírt szigorú attribútum-egyezési feltétel helyett tetszőleges kapcsolati feltételt megenged a táblák sorainak egyesítéséhez. Mivel a különböző relációk esetlegesen egyező nevű attribútumai különböző feltételekben szerepelhetnek, azért természetesen minősített attribútumneveket (ennek érdekében pedig pontozott egyesítést) használunk.

Példa

Tekintsük az előző példánál bemutatott $R = R\langle X, Y, Z\ \rangle$ és az $S = S\langle Y, Z, U\rangle$ relációk tábláját, valamint a $T_1 = (X < U)$ és a $T_2 = (X < U) \wedge (R.Y \neq S.Y)$ logikai függvényeket.

Ekkor az $R \underset{T_1}{\bowtie} S$ reláció táblája:

X	$R.Y$	$R.Z$	$S.Y$	$S.Z$	U
1	2	3	2	3	4
1	2	3	2	3	5
1	2	3	7	8	10
6	7	8	7	8	10
9	7	8	7	8	10

és az $R \underset{T_2}{\bowtie} S$ reláció táblája:

X	$R.Y$	$R.Z$	$S.Y$	$S.Z$	U
1	2	3	7	8	10

Példa

$R\langle X, Y\rangle$:

X	Y
1	2
1	2

$\underset{R.Y<S.Y}{\bowtie}$

$S\langle Y, Z\rangle$:

Y	Z
2	3
4	5
4	5

=

X	$R.Y$	$S.Y$	Z
1	2	4	5
1	2	4	5
1	2	4	5
1	2	4	5

Az eredmény nyilván itt csak akkor lesz valódi multireláció, ha az operandus-relációk legalább egyike is az volt.

Relációműveletek kapcsolata

A természetes és az általános összekapcsolás olyan összetett relációs művelet, mely felépíthető Descartes-szorzásból és szelekcióból. A Descartes-szorzás hátránya, hogy az eredménytábla rettenetesen megnőhet, mivel mérete (sorainak száma) az operandus-táblák méreteinek szorzata. Célszerű ezért egyrészt a különböző szelekciós műveleteket (ha csak lehet) még az operandus-táblákon elvégezni, másrészt pedig Descartes-szorzás helyett természetes, vagy általános összekapcsolással a műveletből már eleve kizárni azokat az operandus-táblabeli sorokat, melyek egy kapcsolati feltételnek nem felelnek meg. (Descartes-szorzás használata esetén e feltételeknek nem megfelelő sorokat egy utólagos szelekcióval törölnénk, persze már egy jóval nagyobb táblából.)

A gyakorlatban még fontosabb a természetes összekapcsolás helyettesítése általánossal. Az előbbinek a veszélye abban rejlik, hogy a kapcsolat implicit (nem látható), egyszerűen a két tábla bizonyos oszlopainak azonos jelölésén (nevén) alapszik. Így utólag, pusztán „ránézésre” esetleg nem is vesszük észre, hogy a két tábla hogy is kapcsolódik. Súlyos problémát okoz, ha a kapcsolatban résztvevő valamelyik tábla egyik oszlopát átnevezzük, mivel ekkor a természetes összekapcsolás már nem is „működik”. Hasonlóan hibát okoz, ha nem akarunk az összes egyező nevű attribútum mentén kapcsolatot teremteni (például azért, mert az azonos nevű attribútumok egy része az egyik táblában mást jelent, mint a másikban). Ha két táblát össze kell kapcsolnunk, akkor tehát használjunk általános összekapcsolást. Ezt „támogatja” az SQL azáltal, hogy a természetes összekapcsolást nem is valósítja meg. A relációműveletek közötti legfontosabb kapcsolatok az alábbiak.

1. Az általános összekapcsolás és a Descartes-szorzat kapcsolata

Tetszőleges $R\langle \underline{A} \rangle$ és $S\langle \underline{B} \rangle$ relációk és $T_{A \dot\cup B}$ logikai függvény esetén:

$$R\langle \underline{A} \rangle \underset{T_{A\dot\cup B}}{\bowtie} S\langle \underline{B} \rangle = \sigma_{T_{A\dot\cup B}}(R\langle \underline{A} \rangle \times S\langle \underline{B} \rangle), \text{ és így nyilván}$$

$$(R\langle \underline{A} \rangle \underset{T_{A\dot\cup B}}{\bowtie} S\langle \underline{B} \rangle) \subseteq (R\langle \underline{A} \rangle \times S\langle \underline{B} \rangle).$$

2. A természetes összekapcsolás és a Descartes-szorzat kapcsolata

Legyen $R\langle \underline{A} \rangle$ és $S\langle \underline{B} \rangle$ relációk esetén $A \cap B \neq \emptyset$. Ekkor

$$R\langle \underline{A} \rangle \bowtie S\langle \underline{B} \rangle = \sigma_{T_{A\dot\cup B}}(R\langle \underline{A} \rangle \times S\langle \underline{B} \rangle).$$

3. Az általános és a természetes összekapcsolás kapcsolata

Legyen $R\langle \underline{A} \rangle$ és $S\langle \underline{B} \rangle$ relációk esetén $A \cap B \neq \emptyset$. Ekkor

$$R\langle \underline{A} \rangle \underset{T_{A\dot\cup B}}{\bowtie} S\langle \underline{B} \rangle = R\langle \underline{A} \rangle \bowtie S\langle \underline{B} \rangle.$$

A fentiekben a $T_{A\dot\cup B}$ logikai függvény például lehet a következő alakú:

$$T_{A\dot\cup B} = (R.C_1 = S.C_1) \wedge (R.C_2 = S.C_2) \wedge \ldots \wedge (R.C_n = S.C_n), \text{ melyben}$$

$$R.C_1, R.C_2, \ldots, R.C_n, S.C_1, S.C_2, \ldots, S.C_n \in A \dot\cup B.$$

Osztályozó műveletek

A szerkezetmódosító műveletekhez hasonlóan az osztályozó műveletek – melyek az algebrai osztályozáson, pontosabban az attribútum szerinti sémafelosztáson alapulnak - szintén módosítják a paraméter-reláció szerkezetét, ám ezt egészen másként teszik. Míg a korábban bemutatott szerkezetmódosító műveletek az eredmény-reláció szerkezetét a paraméter-reláció adataitól függetlenül, kizárólag a művelet jellegéből adódóan határozták meg, addig az osztályozó műveletek végrehajtásának módja alapvetően a paraméter-reláció adataitól függ. Ha ugyanis valamely kiválasztott attribútum mentén az egyes rekordok tartalmaznak különböző értékeket, akkor kialakulhat egy többosztályos felosztás, egyébként pedig nem.

Az osztályozó műveleteket, ha valamilyen aritmetikai művelet elvégzésére használjuk, nevezhetjük *relációaritmetikai műveleteknek* is. Nem tartoznak a hagyományos relációalgebrai műveletek közé, mégis célszerű őket bevezetni, mivel igen gyakran van szükségünk olyan kiválasztási, vagy lekérdezési tevékenységre, melynek során egy táblában bizonyos tulajdonságú sorok valamely attribútumait összegezzük, átlagoljuk, stb.

Az osztályozó műveletekben szereplő függvényeket az adatbázis-kezelő szoftverek tárgyalásánál *osztályfüggvényeknek*, *többsoros függvényeknek*, vagy *csoportfüggvényeknek* nevezzük, jelezve ezzel azt, hogy egyrészt a műveletek jellegéből adódóan tetszőleges számú operandusuk lehet, másrészt, hogy ezek az operandusok jellemzően egy adott tábla több sorának (rekordjának) azonos oszlopához (attribútumához) tartoznak.

A függvények másik csoportját alkotják az úgynevezett *egysoros függvények*, melyeket másnéven unáris, vagy egyoperandusú műveleteknek is hívunk. Ilyenek a különböző adatkonvertáló függvények, az előjel-, vagy az abszolútérték-függvények, stb. Ezek tetszőleges matematikai kifejezés tagjai lehetnek, a táblák szerkezetét nem befolyásolják, így – bár használni fogjuk őket – részletes tárgyalásukra nincs szükség.

Az osztályfüggvények alkalmazásához jelentek meg az SQL nyelv résztábla-kiválasztást lehetővé tevő SELECT utasításában a GROUP BY és a HAVING paraméterek. Ezek használatának megtanulását nehezíti, hogy egy összetett SELECT utasításban az osztályfüggvények és ezek a paraméterek általában igen távol kerülnek egymástól, és a kapcsolatuk, az alkalmazásuk módja sem nyilvánvaló.

Az osztályozó műveletek felírásánál a korábban bevezetett attribútum szerinti sémafelosztás jelölésétől egy kissé el fogunk térni éppen azért, hogy a fenti SQL utasításokkal minél jobban összhangban lehessünk. Mielőtt azonban megadnánk az osztályozó műveletek általános alakját, nézzünk egy példát egy gyakorlati osztályozási feladatra és megoldásának gondolatmenetére!

Bevezető példa

Legyen az *eladás*⟨SZALON, ÉV, BEVÉTEL, MÁRKA⟩ reláció táblája az alábbi:

SZALON	ÉV	BEVÉTEL	MÁRKA
Pipacs	1999	20000	Jaguar
Pipacs	1999	700	Mercedes
Pipacs	2000	14000	Jaguar
Pipacs	2000	2000	Mercedes
Pipacs	2001	10300	Jaguar
Pipacs	2001	5000	Mercedes
Szekér	1999	15000	Porsche
Szekér	1999	8700	Opel
Szekér	2000	12500	Porsche
Szekér	2000	22300	Opel
Szekér	2001	10500	Porsche
Szekér	2001	28000	Opel

Feladat

Készítsünk kimutatást, mely szalononkénti és évenkénti osztályozásban tartalmazza azon szalonok bevételeit (ezer euro-ban), melyeknek éves bevétele meghaladja a 20000-et!

Megoldás

A megoldást nyilván az alábbi táblájú *kimutatás*⟨SZALON, ÉV, ÖSSZBEVÉTEL⟩ eredmény-reláció adja:

SZALON	ÉV	ÖSSZBEVÉTEL
Pipacs	1999	20700
Szekér	1999	23700
Szekér	2000	34800
Szekér	2001	38500

Hogyan juthatunk el ehhez a megoldáshoz algebrai úton?

A megoldás módszere

Először kiválasztjuk a feladat relációjának (táblájának) a megoldás szempontjából fontos alrelációját (altábláját) annak érdekében, hogy kisebb legyen az elvégzendő feladat. A fentiekben például az eladott gépkocsik márkája érdektelen. Az első lépés tehát az *alreláció-kiválasztás* (altábla-kiválasztás). Második lépésként a SZALON és az ÉV attribútumok értékei szerint részrelációkra, vagyis rekord-osztályokra (résztáblákra, azaz sor-osztályokra) bontjuk a kapott alrelációt (*osztályozási* tevékenység), majd harmadik lépésként az így kapott részrelációk rekordjainak BEVÉTEL attribútumán részrelációnként kiszámítjuk az ÖSSZBEVÉTEL értéket előállító *Összegzés* osztályfüggvényt. Tehát az osztályozás után minden részrelációt egyetlen rekorddal (úgynevezett *reprezentáns-rekorddal*) kell helyettesíteni (ezt *faktorizációnak* nevezzük). Végül a negyedik lépés a reprezentáns-rekordok megszűrése megadott osztályszűrő feltétellel (ez az *osztályvizsgálat*). Az alábbiakban először elvileg tekintjük át ezeket a lépéseket, majd a gyakorlatban is elvégezzük őket.

1. lépés (Alreláció-kiválasztás)

Az *alreláció-kiválasztás* egy vetítési művelet (projekció), melynek során a vizsgált relációból (a paraméter-relációból) elhagyjuk mindazon attribútum-értékeket, melyekre a feladat megoldása során, vagy végeredménye szempontjából nincs szükségünk (tehát az altábla-kiválasztás oszlopelhagyást jelent a paraméter-táblából).

2. lépés (Osztályozás)

Az *osztályozás* célja egy halmaznak valamilyen szempontból azonos tulajdonságú elemei szerinti diszjunkt (egymást át nem fedő) részhalmazokra bontása. Eredményeként a halmaz olyan osztályozása jön létre, melyben az egy osztályon belül levő elemek (és csak azok) egyenlőek az osztályozási tulajdonság szerint. Speciális esetben előfordulhat, hogy minden osztály egyelemű (ezt triviális osztályozásnak nevezzük). A gyakorlatban ezzel ekvivalens, ha egyáltalán nem adunk meg osztályozási tulajdonságot. Esetünkben a halmaz elemei (a paraméter-reláció rekordjai) egy osztályba kerülnek, ha a SZALON és az ÉV attribútum-értékekben megegyeznek.

3. lépés (Faktorizáció)

A *faktorizáció* során az osztályokra (diszjunkt részhalmazokra) bontott halmazban az egyes osztályokat helyettesítjük egy úgynevezett *reprezentáns elemmel*. Ezeket azért nevezzük reprezentáns elemeknek, mert azt a tulajdonságot reprezentálják, melyben az általuk kiváltott osztály elemei megegyeznek. Feltehető a kérdés, hogy a relációalgebrában a reprezentáns elemek, mint rekordok milyen attribútumokkal rendelkeznek. A faktorizáció szempontjából a paraméter-reláció attribútumai az alábbi négy típusba sorolhatók:

- *Osztálytag-attribútumok*
  Azok az attribútumok, melyekre vonatkozóan az egy osztálybeli rekordok azonosak.
- *Osztályképző-attribútumok*
  Az osztálytag-attribútumoknak egy olyan részhalmaza, mely még alkalmas a szükséges osztályozás elvégzésére. Célszerű az ilyen tulajdonságú attribútumok minimális halmazát használni osztályozásra.
- *Osztályfüggvény-paraméter attribútum*
  Az alkalmazott osztályfüggvény ebből, mint paraméterből állítja elő az osztályfüggvény értéket, az úgynevezett *osztályfüggvény-attribútumot*.
- *Osztályidegen attribútumok*
  Ezek az eredeti reláció azon attribútumai, melyek nem osztálytagok (tehát egy osztályon belül is különböző értékeket vehetnek fel). Ezeket az attribútumokat (mivel nem reprezentálják az osztály elemeit) nem tartalmazzák a reprezentáns rekordok. Nyilván az osztályfüggvény-paraméter attribútum is osztályidegen.

A reprezentáns rekordokból álló eredmény-reláció tehát csak a paraméter-reláció *osztályképző-attribútumait* és az *osztályfüggvény-attribútumot* tartalmazhatja. Mást nem, hiszen a paraméter-relációból származtattuk (és legalább egy attribútuma biztos van, különben üres). De vajon kell-e tartalmaznia mind a kétféle attribútumot?

- *1. eset: Hiányoznak az eredményből az osztályképző-attribútumok.*
  Mivel az osztályozást általában éppen az azonos tulajdonságú osztályok valamilyen osztályfüggvény-értékének meghatározása miatt végezzük, az eredmény-relációból hiányozhatnak az osztályképző-attribútumok. Használatuk mégis célszerű, mivel ezek nélkül igen nehézkes az osztályfüggvény értékeinek értelmezése (hacsak nem egyetlen érték az osztályozó művelet eredménye).
- *2. eset: Hiányzik az eredményből az osztályfüggvény-attribútum.*
  Ha valamilyen (például az osztályfüggvényre vonatkozó) osztályszűrővel megszűrt reprezentáns rekordok osztálytag-attribútumai egy részének a listázására van szükségünk, akkor előfordul, hogy e listából kihagyjuk az osztályfüggvény értékét.
  Ha azonban nem használunk osztályszűrőt, akkor az eredmény-relációt osztályozó művelet nélkül is megkaphatjuk (`SELECT` utasítás esetén a megfelelő oszlopokat tartalmazó altábla kiválasztásával és a `DISTINCT` paraméter használatával).

Megjegyezzük, hogy triviális osztályozás esetén a faktorizáció azt eredményezi, hogy a paraméter-reláció minden rekordját önmaga reprezentálja. Tehát például, ha nem adunk meg osztályozási tulajdonságot, akkor a faktorizáció eredménye az eredeti paraméter-reláció lesz.

4. lépés (Osztályvizsgálat)

Az osztályvizsgálat során a faktorizáció eredményeként kapott reprezentáns-rekordok közül azokat tartjuk meg, melyek kielégítik a – valamilyen logikai függvénnyel megadott – osztályszűrő feltételt. Ezt röviden csak osztályszűrőnek nevezzük. Ez az osztályszűrő gyakran tartalmazza magát az osztályfüggvényt. Speciális esetként előfordulhat, hogy nem adunk meg osztályszűrőt. Ekkor teljesen elmarad az osztályvizsgálat, és nyilván a faktorizáció eredményeként kapott minden reprezentáns-rekord bekerül az eredményrelációba.

A megoldás gyakorlati lépései

Az alábbiakban bemutatjuk, hogy az osztályozó művelet végrehajtásának egyes lépéseit milyen módon alkalmazzuk a gyakorlatban. Ennek megfelelően az alábbi leírásban reláció helyett táblát, attribútum helyett oszlopot és rekord helyett sort fogunk mondani.

1. lépés (Altábla-kiválasztás)

Először létrehozzuk az *eladás* tábla *átmeneti* altábláját, mely csak a számunkra szükséges osztálytag-oszlopokat és az osztályfüggvény-paraméter oszlopot tartalmazza.

Az alábbiakban látható ennek az *átmeneti*⟨SZALON, ÉV, BEVÉTEL⟩ relációnak a táblája.

SZALON	ÉV	BEVÉTEL
Pipacs	1999	20000
Pipacs	1999	700
Pipacs	2000	14000
Pipacs	2000	2000
Pipacs	2001	10300
Pipacs	2001	5000
Szekér	1999	15000
Szekér	1999	8700
Szekér	2000	12500
Szekér	2000	22300
Szekér	2001	10500
Szekér	2001	28000

2. lépés (Osztályozás)

A ⟨SZALON, ÉV⟩ osztályozás eredménye:

osztályok	SZALON	ÉV	BEVÉTEL
1.	Pipacs	1999	20000
	Pipacs	1999	700
2.	Pipacs	2000	14000
	Pipacs	2000	2000
3.	Pipacs	2001	10300
	Pipacs	2001	5000
4.	Szekér	1999	15000
	Szekér	1999	8700
5.	Szekér	2000	12500
	Szekér	2000	22300
6.	Szekér	2001	10500
	Szekér	2001	28000

az ⟨ÉV, SZALON⟩ osztályozás eredménye:

osztályok	SZALON	ÉV	BEVÉTEL
1.	Pipacs	1999	20000
	Pipacs	1999	700
2.	Szekér	1999	15000
	Szekér	1999	8700
3.	Pipacs	2000	14000
	Pipacs	2000	2000
4.	Szekér	2000	12500
	Szekér	2000	22300
5.	Pipacs	2001	10300
	Pipacs	2001	5000
6.	Szekér	2001	10500
	Szekér	2001	28000

A feladat-kijelölés szerint az osztályképző-oszlopok vektora ⟨SZALON, ÉV⟩. Az osztályozás szempontjából nem lényeges, hogy azt az oszlopok milyen sorrendje szerint végezzük. Láthatóan a kétféle osztályozás eredménytáblájában ugyanazok az osztályok jönnek létre, csak a sorrendjük más. Mivel azonban a relációkat vektorhalmazoknak tekintettük, a rekordok (vagyis a táblabeli sorok) sorrendjének nincs elvi jelentősége.

3. lépés (Faktorizáció)

Az osztályhelyettesítő reprezentáns-sorok a kiválasztott osztályképző-oszlopok értékein kívül az egyes osztályokhoz tartozó osztályfüggvény értéket tartalmazzák.

Legyen az eredménytábla neve *kimutatás*, az eredményvektor pedig ⟨SZALON, ÉV, ÖSSZBEVÉTEL⟩. Ekkor a faktorizáció eredménye:

SZALON	ÉV	ÖSSZBEVÉTEL
Pipacs	1999	20700
Pipacs	2000	16000
Pipacs	2001	15300
Szekér	1999	23700
Szekér	2000	34800
Szekér	2001	38500

4. lépés (Osztályvizsgálat)

SZALON	ÉV	ÖSSZBEVÉTEL
Pipacs	1999	20700
Szekér	1999	23700
Szekér	2000	34800
Szekér	2001	38500

A feladat osztályszűrő feltétele szerint csak azon szalonok bevételeire van szükségünk, melyeknek éves bevétele meghaladja a 20000-et. Ez a fenti tábla nem megfelelő sorainak kiszűrését jelenti. Így már valóban a korábban felírt eredmény-táblát kaptuk.

Az osztályozó műveletek általános alakja

Az osztályozó művelet is egy reláció-transzformáció, mely egy paraméter-relációt leképez egy eredmény-relációba.

Az osztályozó műveleteket olymódon definiáljuk, hogy egyetlen összetartozó szerkezet tartalmazza

- magát az osztályfüggvényt (paramétereként egy attribútummal),
- a művelet paraméter-relációját (azt a relációt, amelyre alkalmazzuk az osztályozó műveletet, és amelynek egy attribútuma az osztályfüggvény paramétere),
- azoknak az attribútumoknak a vektorát, melyekre vonatkozóan az osztályképzést elvégezzük,
- valamint azokat a feltételeket, melyeknek eleget tevő osztályok a művelet eredményében szerepelhetnek.

Az osztályozó műveletek általános alakja:

$$OP(p)[\,R\langle \underline{A}\rangle \,|\, \underline{B}\,]\rfloor_T \Rightarrow \underline{C}\,,$$

ahol

OP: az osztályfüggvény neve
(megjegyezzük, hogy az Oracle-rendszerben az osztályfüggvényt többsoros-függvénynek nevezik),

$R\langle \underline{A}\rangle$: a *paraméter-reláció* (melyre a műveletet alkalmazzuk),

$\underline{B}$: az osztályképző-attribútumok vektora, ahol $B \subseteq B^*$, melyben
B^* az osztálytag-attribútumok halmaza (nyilván $B^* \subseteq A$),

p: a *paraméter-attribútum*, vagyis az OP művelet paramétere, ahol
$p \in P$, melyben $P = A \setminus B^*$, vagy
p egy$P$-beli attribútumkifejezés,

$\underline{C}$: az *eredmény-reláció attribútumvektora*, ahol $C \subseteq B \cup \{\,OP\,\}$,
melyben az OP az osztályfüggvény eredményeit tartalmazó *osztályfüggvény-attribútum* (ez az osztályfüggvény neve), végül

T: az *osztályszűrő*, mely egy logikai függvény (osztálykritérium) a C attribútumhalmaz felett (az előtte szereplő "$\rfloor$" szimbólum azt jelzi, hogy az osztályszűrő a teljes előtte álló kifejezés, vagyis az osztályfüggvény eredményére vonatkozik).

Megjegyzés

- Az eredmény-relációból hiányozhatnak az osztályképző attribútumok, vagy hiányozhat az osztályfüggvény-attribútum, de legalább egy attribútumnak szerepelnie kell.
- Ha az eredmény-reláció attribútumvektorában szerepel az osztályfüggvény-attribútum, akkor csak az osztályfüggvény neve kerülhet be, de a paramétere már nem.
- Ha az osztályszűrő hivatkozik az osztályfüggvényre, akkor ebben az esetben is csak annak neve kerülhet be, de a paramétere már nem.
- Az osztályképző-attribútumok vektora opcionális paraméter.
  Hiánya esetén az eredmény-reláció minden rekordja osztály-reprezentáns rekordként viselkedik. Ekkor az osztályozó műveletek alakja:

  $$OP(p)[\,R\langle \underline{A}\rangle\,]\rfloor_T \Rightarrow \underline{C}\,.$$

- A T osztályszűrő opcionális paraméter.
  Hiánya esetén nincs megszorítás az osztályképzés eredmény-relációjára.
  Ekkor az osztályozó műveletek alakja:

  $$OP(p)[\,R\langle \underline{A}\rangle \,\big|\, \underline{B}\,] \Rightarrow \underline{C}\,.$$

A bevezető példa folytatása

Írjuk fel az alfejezet bevezető példáját az osztályozó művelet fent bevezetett jelölésével!

Megoldás

A feladat paraméterei tehát:

$\underline{A} = \langle \text{szalon, év, bevétel, márka} \rangle$,
$R\langle \underline{A}\rangle = \textit{eladás}\langle \text{szalon, év, bevétel, márka} \rangle$,
$\underline{B} = \langle \text{szalon, év} \rangle$,
$p = \text{bevétel}$,
$OP = \textit{Sum}$,
$\underline{C} = \langle \text{szalon, év}, \textit{Sum} \rangle$,
$T = (\,\textit{Sum} > 20000\,)$

Igy a megoldás:

$$\begin{aligned}&\textit{kimutatás}\langle \text{SZALON, ÉV}, \textit{Sum}\rangle = \\ &\quad \textit{Sum}(\text{BEVÉTEL}) \\ &\qquad [\,\textit{eladás}\langle \text{SZALON, ÉV, BEVÉTEL, MÁRKA}\rangle \,\big|\, \langle \text{SZALON, ÉV}\rangle \\ &\qquad]\rfloor_{(\textit{Sum} > 20000)} \Rightarrow \langle \text{SZALON, ÉV}, \textit{Sum}\rangle\,.\end{aligned}$$

Ha figyelembe vesszük, hogy a bevezető példában az osztályfüggvény eredmény-attribútumát „ÖSSZBEVÉTEL” módon jelöltük, akkor egy átnevezőfüggvényt is alkalmaznunk kell.

Ekkor a megoldás:

$$kimutatás\langle \text{SZALON}, \text{ÉV}, \text{ÖSSZBEVÉTEL}\rangle =$$
$$Ren_{kimutatás\langle \text{SZALON, ÉV, ÖSSZBEVÉTEL}\rangle}$$
$$(Sum(\text{BEVÉTEL})$$
$$[\,eladás\langle \text{SZALON}, \text{ÉV}, \text{BEVÉTEL}, \text{MÁRKA}\rangle \,\big|\, \langle \text{SZALON}, \text{ÉV}\rangle$$
$$]\rfloor_{(Sum > 20000)} \Rightarrow \langle \text{SZALON}, \text{ÉV}, Sum\rangle$$
$$)\,.$$

Ha átnevezőfüggvény helyett az „**AS**” átnevező-operátort használjuk, akkor

$$kimutatás\langle \text{SZALON}, \text{ÉV}, \text{ÖSSZBEVÉTEL}\rangle =$$
$$Sum(\text{BEVÉTEL})$$
$$[\,eladás\langle \text{SZALON}, \text{ÉV}, \text{BEVÉTEL}, \text{MÁRKA}\rangle \,\big|\, \langle \text{SZALON}, \text{ÉV}\rangle$$
$$]\rfloor_{(Sum > 20000)} \Rightarrow \langle \text{SZALON}, \text{ÉV}, Sum \textbf{ AS } \text{ÖSSZBEVÉTEL}\rangle\,.$$

Miután ez az alak egyszerűbb és áttekinthetőbb, ezért általában ezt használjuk.

Osztályozó műveletek kompozíciója

Lehetőség van arra, hogy egy osztályképzés során több osztályfüggvényt is alkalmazzunk. Ennek természetesen az a feltétele, hogy az osztályképzés azonos legyen. Alkalmazására tipikus példa, amikor átlagot és szórást kell számítani. Az osztályozó műveletek kompozíciójának értelmezése:

$$(OP_1(p)[\,R\langle \underline{A}\rangle \,|\, \underline{B}\,]\rfloor_{T_1} \Rightarrow \underline{C}_1\,) \otimes (OP_2(p)[\,R\langle \underline{A}\rangle \,|\, \underline{B}\,]\rfloor_{T_2} \Rightarrow \underline{C}_2\,) =$$
$$(OP_1(p), OP_2(p)[\,R\langle \underline{A}\rangle \,|\, \underline{B}\,]\rfloor_{T} \Rightarrow \underline{C}\,)\,,$$

ahol

- $\otimes$ jelöli a műveletkompozíciót,
- $T = T_1 \wedge T_2$, melyben $\wedge$ a logikai ÉS művelet jele, és
- $\underline{C}$ pedig a $\underline{C}_1$ és $\underline{C}_2$ attribútumvektorokból összeállított olyan vektor, melyben mindkettő attribútumai szerepelnek, de az azonosak csak egyszer (tehát $C = C_1 \cup C_2$).

Példa
Írjuk fel az előző példát oly módon, hogy az egyes osztályokra az összegen kívül az átlagot is meghatározzuk!

Megoldás

$$kimutatás\langle \text{SZALON}, \text{ÉV}, \text{ÖSSZBEVÉTEL}, \text{ÁTLAG}\rangle =$$
$$Sum(\text{BEVÉTEL}), Avg(\text{BEVÉTEL})$$
$$[\,eladás\langle \text{SZALON}, \text{ÉV}, \text{BEVÉTEL}, \text{MÁRKA}\rangle \,\big|\, \langle \text{SZALON}, \text{ÉV}\rangle$$
$$]\rfloor_{(Sum > 20000)} \Rightarrow \langle \text{SZALON}, \text{ÉV}, Sum \textbf{ AS } \text{ÖSSZBEVÉTEL}, Avg \textbf{ AS } \text{ÁTLAG}\rangle.$$

Eredményként ekkor az alábbi táblát kapjuk:

SZALON	ÉV	ÖSSZBEVÉTEL	ÁTLAG
Pipacs	1999	20700	10350
Szekér	1999	23700	11850
Szekér	2000	34800	17400
Szekér	2001	38500	19250

Egyszerű lekérdezések

Az alábbiakban néhány olyan mintapéldát mutatunk be, melyek ugyan egyszerű alapfeladatokat oldanak meg, mégis jól szemléltetik a relációalgebrai műveletek használatát.

1. Példa
Legyen $\mathit{film}\langle \text{CÍM, ÉV, HOSSZ, STÚDIÓ} \rangle$ egy reláció.

Feladat
Melyek a Fox stúdióban készült, legalább 100 perc hosszúságú filmek, és ezek, mikor készültek?

a.) Megoldás

$$\pi_{\text{CÍM, ÉV}}\big(\sigma_{\text{HOSSZ} \geq 100}(\mathit{film}\langle \text{CÍM, ÉV, HOSSZ, STÚDIÓ} \rangle) \bigcap \sigma_{\text{STUDIÓ} = \text{‘Fox’}}(\mathit{film}\langle \text{CÍM, ÉV, HOSSZ, STÚDIÓ} \rangle)\big)\ .$$

b.) Megoldás

$$\pi_{\text{CÍM, ÉV}}\big(\sigma_{(\text{ HOSSZ} \geq 100) \wedge (\text{STUDIÓ} = \text{‘Fox’ })}(\mathit{film}\langle \text{CÍM, ÉV, HOSSZ, STÚDIÓ} \rangle)\big)\ .$$

Megjegyzés
A két egyenértékű megoldás közül nyilván a második a hatékonyabb, hiszen gyorsabb egy táblán egy összetett feltételt vizsgálni, mint metszet művelettel két tábla közös sorait keresni (amely két táblán még külön is kell egy-egy kiválasztást végezni).

2. Példa
Legyen $\mathit{film1}\langle \text{CÍM, ÉV, HOSSZ, STÚDIÓ} \rangle$, és
$\mathit{film2}\langle \text{CÍM, ÉV, SZÍNÉSZ} \rangle$ két reláció.

Feladat
Kik azok a színészek, akik legalább 100 perc hosszú filmekben szerepelnek

a.) Megoldás

$$\pi_{\text{SZÍNÉSZ}}\big(\sigma_{\text{HOSSZ} \geq 100}(\mathit{film1}\langle \text{CÍM, ÉV, HOSSZ, STÚDIÓ} \rangle \bowtie \mathit{film2}\langle \text{CÍM, ÉV, SZÍNÉSZ} \rangle)\big)\ .$$

b.) Megoldás

$$\pi_{\text{SZÍNÉSZ}}\big(\sigma_{\text{HOSSZ} \geq 100}(\mathit{film1}\langle \text{CÍM, ÉV, HOSSZ, STÚDIÓ} \rangle) \bowtie \mathit{film2}\langle \text{CÍM, ÉV, SZÍNÉSZ} \rangle\big)\ .$$

Megjegyzés
Ezúttal is a második megoldás a hatékonyabb, mivel a természetes összekapcsolást így kisebb (a HOSSZ $\geq$ 100 feltételű kiválasztással csökkentett méretű) táblával kell elvégezni.

3. Példa

Legyen $szállító\langle$ CÉG, CÍM, TELEFON, ÜGYINTÉZŐ, TERMÉK, ÁR $\rangle$, és
$rendelés\langle$ CÉG, TERMÉK, MENNYISÉG $\rangle$ két reláció.

Feladat

Kik azok az ügyintézők, és mi a telefonszámuk, akiknek a cégétől 100 eFt-nál nagyobb értékben vásároltak valamilyen árut?

Megoldás

$$\begin{aligned}&Sum(\text{MENNYISÉG} * \text{ÁR})\\&\quad[\,rendelés\langle \text{CÉG, TERMÉK, MENNYISÉG}\rangle \bowtie\\&\quad\; szállító\langle \text{CÉG, CÍM, TELEFON, ÜGYINTÉZŐ, TERMÉK, ÁR}\rangle \,\big|\, \langle \text{CÉG, TERMÉK}\rangle\\&\quad]\,\rfloor_{(Sum\, >\, 100000)} \Rightarrow \langle \text{ÜGYINTÉZŐ, TELEFON}\rangle .\end{aligned}$$

Megjegyzés

Az eredmény olyan osztálytag-attribútumokból áll, melyek egyike sem osztályképző.

4. Példa

Legyen egy adatbázis modellje:
$film1\langle$ CÍM, ÉV, HOSSZ, STÚDIÓ $\rangle$,
$film2\langle$ CÍM, ÉV, SZÍNÉSZ $\rangle$, és
$színészek\langle$ SZÍNÉSZ, ÉV, KOR, SZEREPSZÁM $\rangle$.

Feladat

Listázza ki a legalább 100 perces filmek színészeinek nevét és korát a film készítésekor!

a.) Megoldás

$$\begin{aligned}&\pi_{\text{SZÍNÉSZ},\ színészek.\text{KOR} - (színészek.\text{ÉV} - film1.\text{ÉV})\ \textbf{AS}\ \text{KORA}}\\&\quad(\sigma_{\text{HOSSZ} \geq 100}(film1\langle \text{CÍM, ÉV, HOSSZ, STÚDIÓ}\rangle) \bowtie\\&\qquad\qquad film2\langle \text{CÍM, ÉV, SZÍNÉSZ}\rangle) \underset{T}{\bowtie}\\&\qquad\qquad színészek\langle \text{SZÍNÉSZ, ÉV, KOR, SZEREPSZÁM}\rangle) ,\end{aligned}$$

ahol $T = (film2.\text{SZÍNÉSZ} = színészek.\text{SZÍNÉSZ})$.

Megjegyzés

A *film2*.SZÍNÉSZ, valamint a *színészek*.SZÍNÉSZ kifejezésekben a korábban már bevezetett úgynevezett *pontozott jelölést*, azaz *minősített nevet* alkalmaztuk.

A fenti példa jól mutatja azt az esetet, amikor nem szabad természetes összekapcsolást használni. Ha a második összekapcsolásban természetes összekapcsolást használnánk, akkor automatikusan létrejönne a *film2*.ÉV = *színészek*.ÉV kapcsolat is, holott ez helytelen lenne. A *film2* relációban ugyanis az ÉV nyilván a film megjelenésének éve, míg a *színészek* reláció-ban az ÉV a színészek aktuális évbeli adatait tartalmazzák. Az alábbiakban megmutatjuk a megoldást arra az esetre, ha teljesen elhagyjuk a természetes összekapcsolást. (Ez már csak azért is indokolt, mivel a gyakorlati adatbázis-kezelő rendszerek nem is valósítják meg a természetes összekapcsolást, csak az általános összekapcsolást.)

b.) Megoldás

$$\pi_{\underline{B}}(\sigma_{T_1}(\mathit{film1}\langle \text{CÍM, ÉV, HOSSZ, STÚDIÓ}\rangle) \bowtie_{T_2}$$
$$\mathit{film2}\langle \text{CÍM, ÉV, SZÍNÉSZ}\rangle) \bowtie_{T_3}$$
$$\mathit{színészek}\langle \text{SZÍNÉSZ, ÉV, KOR, SZEREPSZÁM}\rangle) ,$$

ahol

$\underline{B} = \langle \mathit{film2}.\text{SZÍNÉSZ},\ \mathit{színészek}.\text{KOR} - (\mathit{színészek}.\text{ÉV} - \mathit{film1}.\text{ÉV})\ \mathbf{AS}\ \text{KORA}\rangle$,

$T_1 = (\mathit{film1}.\text{HOSSZ} \geq 100)$,

$T_2 = (\mathit{film1}.\text{CÍM} = \mathit{film2}.\text{CÍM}) \wedge (\mathit{film1}.\text{ÉV} = \mathit{film2}.\text{ÉV})$,

$T_3 = (\mathit{film2}.\text{SZÍNÉSZ} = \mathit{színészek}.\text{SZÍNÉSZ})$.

5. Példa

Tekintsük az előző példa adatbázis modelljét.

Feladat

Adja meg a legalább 100 perces filmek színészeinek átlagéletkorát a film készítésekor!

Megoldás

$$\mathit{Avg}(\text{KORA})\ [\pi_{\text{SZÍNÉSZ},\ \mathit{színészek}.\text{KOR} - (\mathit{színészek}.\text{ÉV} - \mathit{film1}.\text{ÉV})\ \mathbf{AS}\ \text{KORA}}$$
$$(\sigma_{\text{HOSSZ} \geq 100}(\mathit{film1}\langle \text{CÍM, ÉV, HOSSZ, STÚDIÓ}\rangle) \bowtie$$
$$\mathit{film2}\langle \text{CÍM, ÉV, SZÍNÉSZ}\rangle) \bowtie_{T}$$
$$\mathit{színészek}\langle \text{SZÍNÉSZ, ÉV, KOR, SZEREPSZÁM}\rangle)$$
$$] \Rightarrow \langle \mathit{Avg}\ \mathbf{AS}\ \text{ÁTLAGÉLETKOR}\rangle,$$

ahol $T = (\mathit{film2}.\text{SZÍNÉSZ} = \mathit{színészek}.\text{SZÍNÉSZ})$.

Megjegyzés

Ez egy olyan speciális esete az osztályozó műveleteknek, melyben

- hiányzik az osztályképző-attribútum vektor
  (nincs rá szükség, mivel a teljes paraméter-reláción kiszámítjuk az osztályfüggvényt),
- hiányzik az osztályszűrő
  (nincs rá szükség, mivel a paraméter-relációban szereplő kiválasztás feltétele – a T logikai függvény – már elvégezte a megfelelő szűrést),
- az eredményben csak az osztályfüggvény-attribútum szerepel
  (nincs szükség értelmező osztályképző-attribútumokra, mivel a feladat csupán egyetlen szám meghatározása volt).

6. Példa

Legyen egy adatbázis modellje:

$\mathit{kézirat}\langle \text{CÍM, SZERZŐ, TÉMA}\rangle$,

$\mathit{könyv}\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR}\rangle$,

$\mathit{könyvkiadó}\langle \text{KIADÓ, ORSZÁG, POSTA\_CÍM, TELEFON}\rangle$.

Feladat

Listázza ki a krimi-kiadók nevét és postai címét az ország megjelölésével. Az elkészült lista a $\langle$ KIADÓ, CÍME, ORSZÁG $\rangle$ fejlécet tartalmazza. Megjegyezzük, hogy egy kiadót "krimi-kiadónak" nevezünk, ha legalább egy krimit kiadott.

Megoldás

$$\pi_{\underline{B}}(\sigma_T(\textit{kézirat}\langle \text{CÍM, SZERZŐ, TÉMA}\rangle) \bowtie \textit{könyv}\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR}\rangle \bowtie \textit{könyvkiadó}\langle \text{KIADÓ, ORSZÁG, POSTA\_CÍM, TELEFON}\rangle\ ,$$

ahol

$\underline{B} = \langle \textit{könyvkiadó}.\text{KIADÓ}, \textit{könyvkiadó}.\text{POSTA\_CÍM } \textbf{AS} \text{ CÍME}, \textit{könyvkiadó}.\text{ORSZÁG}\rangle$,

$T = (\textit{kézirat}.\text{TÉMA} = „\text{krimi}”)$.

7. Példa

Tekintsük az előző példa adatbázis modelljét.

Feladat

Listázza ki azon szerzőket, akiknek az 1990 és 2000 között kiadott könyveinek átlagára legalább 1000,-Ft. A lista tartalmazza a szerzők nevein kívül a könyveik átlagárát is a fenti időszakban. A lista fejléce legyen ⟨SZERZŐ, ÁTLAGÁR⟩ alakú.

Megoldás

$$Avg(\text{ÁR})\ [\sigma_{T_1}(\textit{könyv}\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR}\rangle) \mid \langle \text{SZERZŐ}\rangle\ ]\rfloor_{T_2} \Rightarrow \langle \text{SZERZŐ}, Avg \ \textbf{AS}\ \text{ÁTLAGÁR}\rangle,$$

ahol $T_1 = (\textit{könyv}.\text{ÉV} \in \{1990,\ldots,2000\})$,

$T_2 = (Avg \geq 1000)$.

Egy összetett mintapélda

A relációalgebra alfejezetének lezárásaként tekintsünk egy összetett mintafeladatot, mely bemutatja az eddig tárgyalt szerkezetmódosító és osztályozó műveleteket.

Példa

Legyen egy adatbázis modellje a következő:

kézirat⟨CÍM, SZERZŐ, TÉMA⟩,
könyv⟨CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR⟩,
könyvkiadó⟨KIADÓ, ORSZÁG, POSTA_CÍM, TELEFON⟩ .

Listázza ki az 1 Millió $ éves bevételt meghaladó krimi-kiadók nevét, az ország nevét, amelyben működnek és éves bevételüket 1990-től 2000-ig éves bontásban! Az elkészült lista fejléce ⟨KIADÓ, ORSZÁG, ÉV, ÉVES BEVÉTEL⟩ legyen, és a listában ne legyenek azonos sorok!

Megjegyzés

- Egy kiadó akkor krimi-kiadó, ha a vizsgált időszakban legalább egy krimit kiadott.
- Az éves árbevételt egy adott év kiadott könyveinek darabszáma és ára határozza meg.

Megoldás

A megoldáshoz a feladatot részfeladatokra bontjuk.

1. lépés (A lista reláció specifikálása)

A *lista* reláció specifikálásához két követelményt kell első lépésként figyelembe venni. Egyrészt a fejléc felépítésére vonatkozót, másrészt azt, hogy a *lista* ne tartalmazzon azonos sorokat. Ezek rögzítése már csak azért is célszerű, mivel a továbbiak ezt már nem fogják befolyásolni. Persze azért a megoldás végén célszerű meggondolni, hogy a *lista* előállítása során vajon valóban keletkezhetnek-e azonos sorok, mert ha nem, akkor a redukáló függvény alkalmazására természetesen nincs szükség (mint tudjuk ez jelentősen lassíthatja a *lista* előállítását).

$$redukált\_lista\langle \text{KIADÓ, ORSZÁG, ÉV, ÉVES BEVÉTEL} \rangle = Reduc(lista\langle \text{KIADÓ, ORSZÁG, ÉV, ÉVES BEVÉTEL} \rangle)\ . \quad (1)$$

2. lépés (A lista reláció elemzése és esetleges felbontása)

A *lista* relációval kapcsolatban az első kérdés, hogy vajon mely paraméter-relációkra van szükség a megvalósításához. Az attribútumvektor alapján megállapíthatjuk, hogy

- a KIADÓ attribútumot a *könyv* és a *könyvkiadó* relációkból nyerhetjük,
- az ORSZÁG attribútumot a *könyvkiadó* relációból,
- az ÉV attribútumot a *könyv* relációból, és
- az ÉVES BEVÉTEL származtatott attribútumot pedig szintén a *könyv* relációból.

Ezek után célszerű a *Lista* relációt az alábbi módon felbontani két rész-relációra:

$$lista\langle \text{KIADÓ, ORSZÁG, ÉV, ÉVES BEVÉTEL} \rangle = R_1\langle \text{KIADÓ, ÉV, ÉVES BEVÉTEL} \rangle \underset{T_1}{\bowtie} R_2\langle \text{KIADÓ, ORSZÁG} \rangle\ , \quad (2)$$

ahol

az $R_1\langle \text{KIADÓ, ÉV, ÉVES BEVÉTEL} \rangle$ reláció a *könyv* relációból van származtatva,
az $R_2\langle \text{KIADÓ, ORSZÁG} \rangle$ reláció pedig a *könyvkiadó* relációból,
a T_1 logikai függvény összekapcsolja az R_1 és az R_2 relációkat a következőképpen:

$$T_1 = (R_1.\text{KIADÓ} = R_2.\text{KIADÓ})\ . \quad (3)$$

3. lépés (A rész-relációk elemzése és esetleges felbontása)

3.1. lépés (Az R_1 reláció elemzése)

Tulajdonképpen az R_1 reláció valósítja meg a feladat szerinti osztályképzés eredményét, mégpedig oly módon, hogy az általános

$$OP(p)[\,R\langle \underline{A} \rangle \,|\, \underline{B}\,]\rfloor_T \Rightarrow \underline{C}$$

alakban

$OP = Sum$ az osztályfüggvény (ugyanis összegezni kell),
$R\langle \underline{A} \rangle = R_3\langle \text{KIADÓ, ÉV, DARAB, ÁR} \rangle$ paraméter-reláció a *könyv* relációból származtatva,
$\underline{B} = \langle \text{KIADÓ, ÉV} \rangle$ az osztályképző-attribútumok vektora,
$p = \text{DARAB} * \text{ÁR}$ az osztályozó művelet paramétere, amely egy attribútumkifejezés,
$\underline{C} = \langle \text{KIADÓ, ÉV}, Sum\ \textbf{AS}\ \text{ÉVES BEVÉTEL} \rangle$ az eredmény-reláció attribútumvektora, és
$T = (Sum > 1000000)$ az osztályszűrő.

A fentiek alapján tehát az R_1 reláció származtatása:

$$R_1\langle \text{KIADÓ, ÉV, ÉVES BEVÉTEL} \rangle =$$
$$\mathit{Sum}(\text{DARAB} * \text{ÁR})$$
$$[\,R_3\langle \text{KIADÓ, ÉV, DARAB, ÁR} \rangle \,\big|\, \langle \text{KIADÓ, ÉV} \rangle$$
$$]\,\big\rfloor_{(\mathit{Sum} > 1000000)} \Rightarrow \langle \text{KIADÓ, ÉV, } \mathit{Sum}\ \textbf{AS}\ \text{ÉVES BEVÉTEL} \rangle\ . \tag{4}$$

3.1.a. lépés (Az R_3 reláció elemzése)

Az R_3 relációnak a *könyv* relációból való származtatása egy kiválasztás (melyben a kiválasztó T_2 logikai függvény meghatározása lesz a következő részfeladat), másrészt a *Sum* osztályozó művelethez szükséges attribútumok vetítése. Tehát

$$R_3\langle \text{KIADÓ, ÉV, DARAB, ÁR} \rangle =$$
$$\pi_{\text{KIADÓ, ÉV, DARAB, ÁR}}(\sigma_{T_2}(\mathit{könyv}\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR} \rangle))\ . \tag{5}$$

3.1.b. lépés (A T_2 logikai függvény elemzése)

A T_2 logikai függvény feladata azon kiadók kiválasztása, melyek bizonyos összetett feltételeknek megfelelnek. Ezek egyrészt vonatkoznak a *kézirat* relációra (a TÉMA attribútum értéke legyen „krimi”), másrészt a *könyv* relációra (az ÉV attribútum értéke legyen 1999 és 2000 közötti). További részfeladat e feltételeknek megfelelő R_4 reláció meghatározása, melynek attribútumhalmaza mind a *kézirat*, mind a *könyv* attribútumait fogja tartalmazni. Így

$$T_2 = (\,\mathit{könyv}.\text{KIADÓ} \in$$
$$\pi_{\text{KIADÓ}}(R_4\langle \text{CÍM, SZERZŐ, TÉMA, ÉV, DARAB, KIADÓ, ÁR} \rangle))\ . \tag{6}$$

3.1.c. lépés (Az R_4 reláció elemzése)

Az előbb beláttuk, hogy R_4 egy összetett reláció, a *kézirat* relációra és a *könyv* relációra vonatkozó feltételekkel, és természetesen egy olyan kapcsolattal ezek között, melyet a CÍM, SZERZŐ párosok által kijelölt könyvek biztosítanak. (Az nyilvánvaló, hogy egy könyvet csak e páros tagjainak egyidejű használatával tudunk egyértelműen kijelölni.) Bontsuk fel tehát ennek megfelelően az R_4 relációt:

$$R_4\langle \text{CÍM, SZERZŐ, TÉMA, ÉV, DARAB, KIADÓ, ÁR} \rangle =$$
$$R_5\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR} \rangle \underset{T_3}{\bowtie} R_6\langle \text{CÍM, SZERZŐ, TÉMA} \rangle\ , \tag{7}$$

ahol

az $R_5\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR} \rangle$ relációt a *könyv* relációból származtattuk,
az $R_6\langle \text{CÍM, SZERZŐ, TÉMA} \rangle$ relációt pedig a *kézirat* relációból,
a T_3 logikai függvény összekapcsolja az R_5 és az R_6 relációkat a következőképpen:

$$T_3 = (\,R_5.\text{CÍM} = R_6.\text{CÍM}\,) \wedge (\,R_5.\text{SZERZŐ} = R_6.\text{SZERZŐ}\,)\ . \tag{8}$$

3.1.d.1. lépés (Az R_5 reláció elemzése)

Az R_5 relációnak a *könyv* relációból való származtatása szintén egy kiválasztás, melyben a kiválasztó T_4 logikai függvény a kiválasztott könyv megjelenési évére vonatkozó feltétel.

Tehát

$R_5\langle$ CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR $\rangle$ =

$$\sigma_{T_4}(könyv\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR}\rangle)\,, \quad (9)$$

$$T_4 = (\,könyv.\text{ÉV} \in \{1990,\ldots,2000\}\,)\,. \quad (10)$$

3.1.d.2. lépés (Az R_6 reláció elemzése)

Az R_6 relációnak a *kézirat* relációból való származtatása az a kiválasztás, melyben a kiválasztó T_5 logikai függvény a kiválasztott könyv témájára vonatkozó feltétel. Tehát

$R_6\langle$ CÍM, SZERZŐ, TÉMA $\rangle$ =

$$\sigma_{T_5}(kézirat\langle \text{CÍM, SZERZŐ, TÉMA}\rangle)\,, \quad (11)$$

$$T_5 = (\,kézirat.\text{TÉMA} = \text{„krimi”}\,)\,. \quad (12)$$

3.2. lépés (Az R_2 reláció elemzése)

Az R_2 relációra az ORSZÁG attribútum megszerzése érdekében van szükségünk, azonban ahhoz, hogy összeköthető legyen a T_1 logikai függvény segítségével az R_1 relációval, szükség van még a KIADÓ attribútumra is. Ennek megfelelően egy olyan kiválasztást (projekciót) használunk, mely *könyvkiadó* relációból ezt a kettőt tartja meg, azaz

$R_2\langle$ KIADÓ, ORSZÁG $\rangle$ =

$$\pi_{\text{KIADÓ, ORSZÁG}}(könyvkiadó\langle \text{KIADÓ, ORSZÁG, POSTA CÍM, TELEFON}\rangle)\,. \quad (13)$$

4. lépés (A megoldás)

A feladat teljes algebrai megoldása az előzőekben felírt (1) ,…, (13) összefüggések együttese. Az alábbiakban ezeket foglaljuk össze. A későbbiekben megmutatjuk a megoldást SQL nyelven is annak érdekében, hogy összehasonlíthassuk a két eszközt.

$redukált\_lista\langle$ KIADÓ, ORSZÁG, ÉV, ÉVES BEVÉTEL $\rangle$ =

$$Reduc(lista\langle \text{KIADÓ, ORSZÁG, ÉV, ÉVES BEVÉTEL}\rangle) \quad (1)$$

$lista\langle$ KIADÓ, ORSZÁG, ÉV, ÉVES BEVÉTEL $\rangle$ =

$$R_1\langle \text{KIADÓ, ÉV, ÉVES BEVÉTEL}\rangle \bowtie_{T_1} R_2\langle \text{KIADÓ, ORSZÁG}\rangle \quad (2)$$

$$T_1 = (\,R_1.\text{KIADÓ} = R_2.\text{KIADÓ}\,) \quad (3)$$

$R_1\langle$ KIADÓ, ÉV, ÉVES BEVÉTEL $\rangle$ =

$$Sum(\,\text{DARAB} * \text{ÁR}\,)$$
$$[\,R_3\langle \text{KIADÓ, ÉV, DARAB, ÁR}\rangle \,|\, \langle \text{KIADÓ, ÉV}\rangle$$
$$]\rfloor_{(Sum > 1000000)} \Rightarrow \langle \text{KIADÓ, ÉV, } Sum \textbf{ AS } \text{ÉVES BEVÉTEL}\rangle \quad (4)$$

$R_3\langle$ KIADÓ, ÉV, DARAB, ÁR $\rangle$ =

$$\pi_{\text{KIADÓ, ÉV, DARAB, ÁR}}(\sigma_{T_2}(könyv\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR}\rangle)) \quad (5)$$

$$T_2 = (\,könyv.\text{KIADÓ} \in \pi_{\text{KIADÓ}}(R_4\langle \text{CÍM, SZERZŐ, TÉMA, ÉV, DARAB, KIADÓ, ÁR}\rangle)) \quad (6)$$

$R_4\langle$ CÍM, SZERZŐ, TÉMA, ÉV, DARAB, KIADÓ, ÁR $\rangle$ =

$$R_5\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR}\rangle \bowtie_{T_3} R_6\langle \text{CÍM, SZERZŐ, TÉMA}\rangle \quad (7)$$

$$T_3 = (R_5.\text{CÍM} = R_6.\text{CÍM}) \wedge (R_5.\text{SZERZŐ} = R_6.\text{SZERZŐ}) \quad (8)$$

$$R_5\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR} \rangle = \sigma_{T_4}(\mathit{könyv}\langle \text{CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR} \rangle) \quad (9)$$

$$T_4 = (\mathit{könyv}.\text{ÉV} \in \{1990, \ldots, 2000\}) \quad (10)$$

$$R_6\langle \text{CÍM, SZERZŐ, TÉMA} \rangle = \sigma_{T_5}(\mathit{kézirat}\langle \text{CÍM, SZERZŐ, TÉMA} \rangle) \quad (11)$$

$$T_5 = (\mathit{kézirat}.\text{TÉMA} = \text{„krimi”}) \quad (12)$$

$$R_2\langle \text{KIADÓ, ORSZÁG} \rangle = \pi_{\text{KIADÓ, ORSZÁG}}(\mathit{könyvkiadó}\langle \text{KIADÓ, ORSZÁG, POSTA\_CÍM, TELEFON} \rangle) \quad (13)$$

Feladatok

A most következő feladatok célja, hogy az Olvasó kipróbálhassa, mennyire sikerült elsajátítania a relációalgebrai műveletek használatát. Persze vannak köztük összetettebbek is, a nagyobb ambíciók kielégítése érdekében.

1.
Legyen egy adatbázis modellje a következő:

$\mathit{dolgozó}\langle$AZONOSÍTÓ, NÉV, MUNKAKÖR, FŐNÖK_AZONOSÍTÓ, BELÉPÉSI_DÁTUM, FIZETÉS, PÓTLÉK, OSZTÁLY_AZONOSÍTÓ$\rangle$,

$\mathit{osztály}\langle$OSZTÁLY_AZONOSÍTÓ, OSZTÁLY_NÉV, VÁROS$\rangle$.

a.) Listázza ki a vállalat összes dolgozójának és főnökének nevét.
b.) Listázza ki a debreceni dolgozók és főnökeik nevét 2000-ben.
c.) Listázza ki „Zelei László” osztályvezető osztályán dolgozó könyvelők nevét és azok összjövedelmét 2000-ben. Megjegyezzük, hogy a FIZETÉS és a PÓTLÉK adatok havi értékeket jelentenek.
d.) Listázza ki az egyes osztályok dolgozóinak átlagjövedelmét 2000-ben.

2.
Legyen egy adatbázis modellje a következő:

$\mathit{kézirat}\langle$CÍM, SZERZŐ, TÉMA$\rangle$,

$\mathit{könyv}\langle$CÍM, SZERZŐ, ÉV, DARAB, KIADÓ, ÁR$\rangle$,

$\mathit{könyvkiadó}\langle$KIADÓ, ORSZÁG, POSTA_CÍM, TELEFON$\rangle$.

a.) Gyűjtse ki a krimi témájú könyvek kiadóinak nevét és postai címét.
b.) Listázza ki azon krimi-írókat, akiknek az 1990 és 2000 között kiadott könyveinek átlagára legalább 1000,-Ft. A lista tartalmazza a szerzők nevein kívül a könyveik átlagárát is a fenti időszakban. A lista fejléce legyen $\langle$SZERZŐ, ÁTLAGÁR$\rangle$ alakú. Megjegyezzük, hogy egy szerzőt "krimi-írónak" nevezünk, ha a vizsgált időszakban legalább egy krimit írt.

3.

Legyen egy adatbázis modellje a következő:

film1⟨CÍM, ÉV, HOSSZ, RENDEZŐ, STÚDIÓ⟩ ,
film2⟨CÍM, ÉV, SZÍNÉSZ⟩ ,
szinészek⟨SZÍNÉSZ, ÉV, KORA⟩ .

a.) Listázza ki a ’Dracula’ című filmek szereplői közül azokat, akik a film készítésekor 80 évnél idősebbek voltak.

b.) Listázza ki a Paramount-színészek nevét és azok éves szerepszámát 1980-tól 1990-ig éves bontásban. Az elkészült lista a ⟨SZÍNÉSZ, ÉV, SZEREPEK SZÁMA⟩ fejlécet tartalmazza. Megjegyezzük, hogy egy színészt "Paramount-színésznek" nevezünk, ha a vizsgált időszakban legalább egy Paramount stúdió által készített filmben szerepelt.

c.) Listázza ki azokat a stúdiókat, amelyek által 1990 és 2000 között készített filmek átlagos hossza legalább 100 perc. A lista tartalmazza a stúdiók nevein kívül a fenti időszakban általuk készített filmek átlagos hosszát is. Az elkészült lista a ⟨STÚDIÓ, ÁTLAGOS FILMHOSSZ⟩ fejlécet tartalmazza.

d.) Listázza ki a Paramount-színészek és rendezők nevét 1980-tól 1990-ig éves bontásban. Az elkészült lista a ⟨RENDEZŐ, ÉV, SZÍNÉSZ⟩ fejlécet tartalmazza. Megjegyezzük, hogy egy színész, vagy egy rendező Paramount-résztvevő, ha a vizsgált időszakban legalább egy Paramount stúdió által készített film készítésében részt vett.

4.

Legyen egy adatbázis modellje a következő:

telephely⟨OSZTÁLY, FŐNÖK, CÍM⟩ ,
munkakör⟨BEOSZTÁS, FIZETÉS, ÉV⟩ ,
dolgozó⟨NÉV, BEOSZTÁS, OSZTÁLY, KEZDŐ_ÉV, ZÁRÓ_ÉV⟩ .

a.) Adja meg „Tóth Béla” osztályvezető osztályának bérösszegét 2000-ben. Megjegyezzük, hogy KEZDŐ_ÉV a munkába lépés évét, a ZÁRÓ_ÉV munkahely elhagyásának évét jelenti, és az aktív dolgozók ZÁRÓ_ÉV adata: 9999.

b.) Listázza ki az osztályok és főnökeik nevét, valamint az évszámokat, továbbá az egyes osztályok bérösszegét 1990-től 2000-ig éves bontásban.