5. fejezet

Algebrai és logikai lekérdező nyelvek

A jelen fejezet során a relációs adatbázisok modellezése helyett a programozásra fektetjük a hangsúlyt. A tárgyalást két absztrakt programozási nyelv leírásával kezdjük: egy algebraival, illetve egy logikán alapulóval. Az algebrai programozási nyelvet (vagy relációs algebrát) már a 2.4. alfejezetben bemutattuk, amely során láthattuk, milyen műveletek szerepelnek a relációs modellben. Most azonban többről van szó, mint az ott bemutatott algebráról, ezért a 2.4. alfejezetben tárgyalt halmaz alapú algebrát kiterjesztjük multihalmazokra. Ez a megfogalmazás már jobban fogja tükrözni a relációs modell gyakorlatban használt megvalósítási módjait. A kiterjesztés miatt a korábbiakon felül lesz még néhány újabb művelet is, mint például a reláció oszlopainak összegzése (például átlag).

A fejezetet egy másik, logikán alapuló lekérdező nyelvnek a formalizmusával zárjuk le. Ezt *Datalog*nak fogjuk nevezni. Ebben a nyelvben már megfogalmazhatunk a kívánt eredménynek megfelelő lekérdezéseket is anélkül, hogy az eredményt kiszámító olyan algoritmust adnánk, amely a relációs algebrai megközelítésnél elhanyagolhatatlan volt.

5.1. Relációs műveletek multihalmazokon

Ebben az alfejezetben a relációkat inkább multihalmaznak tekintjük, mint halmaznak. Éppen ezért ugyanaz a sor többször is megjelenhet egy adott relációban. Néhány relációs műveletet azonban át kell fogalmaznunk, amennyiben a relációkra multihalmazokként tekintünk. Első lépésben tekintsük az alábbi egyszerű példát, amelyben a reláció egy valódi multihalmaz, azaz nem egy halmaz.

5.1. példa. Az 5.1. ábrán látható reláció tulajdonképpen sorokból álló multihalmaz, amelyben az $(1, 2)$ sor háromszor és a $(3, 4)$ sor egyszer szerepel. Ha az 5.1. ábra relációját halmaznak tekintenénk, akkor ki kellene küszöbölni az $(1, 2)$ sor két megjelenését. A multihalmazként kezelt relációban megengedjük ugyan, hogy ugyanazon sor többször szerepeljen, viszont a sorok sorrendje itt sem számít, éppúgy, mint a halmazként kezelt relációk esetén. □

A	B
1	2
3	4
1	2
1	2

5.1. ábra. Egy multihalmaz

5.1.1. Mire jók a multihalmazok?

A forgalomban lévő ABKR-ek a relációkat multihalmazként valósítják meg (nem pedig halmazként), mint ahogy azt már korábban is említettük. A relációk hatékony megvalósítása szempontjából a relációk multihalmazokként való kezelése több módon is gyorsíthatja a relációs műveleteket. Például

1. Ha két multihalmazként értelmezett reláció unióját vesszük, akkor az egyik reláció sorait egyszerűen lemásoljuk, és ehhez a másolathoz hozzáadjuk a másik reláció összes sorát. Ebben az esetben nem kell megszüntetnünk azon ismétlődő sorokat sem, amelyek mind a két relációban szerepelnek.

2. Ha a reláció vetítését halmazként fogjuk fel, akkor minden vetített sort össze kell hasonlítanunk az összes többi vetített sorral azért, hogy meggyőződjünk arról, hogy a vetítésben minden sor csak egyszer szerepel. Ezzel szemben, hogyha az eredményt multihalmaznak tekintjük, akkor minden sort levetítünk, és egyszerűen hozzáadjuk ezeket az eredményhez. Így egyetlen összehasonlítást sem kell végeznünk a vetítésből származó sorok között.

A	B	C
1	2	5
3	4	6
1	2	7
1	2	8

5.2. ábra. Az 5.2. példában szereplő multihalmaz

5.2. példa. Az 5.1. ábrán látható multihalmaz, akár az 5.2. ábrán látható reláció A és B attribútumokra történő vetítésének eredménye is lehetne, feltéve, ha megengednénk, hogy az $(1, 2)$ sor többször is szerepeljen az eredményben, azaz az eredményt multihalmazként kezelnénk. Ha a hagyományos relációs algebrai vetítési operátort használnánk, és ily módon kiküszöbölnénk az azonos sorokat az eredményből, akkor a következő relációt kapnánk:

A	B
1	2
3	4

Figyeljük meg, annak ellenére, hogy a multihalmazként kezelt eredmény nagyobb méretű, mégis gyorsabban ki lehet számolni, hiszen nem kell összehasonlítsuk az összes $(1, 2)$, illetve $(3, 4)$ sort a megelőző lépésben megkapott sorokkal. □

Egy másik motiváció a relációk multihalmazként történő kezelésére: előfordulhat olyan helyzet, mikor a kívánt választ csak úgy kaphatjuk meg, ha legalább is átmenetileg multihalmazos megközelítést használunk. A következő példa ezt mutatja be.

5.3. példa. Ha például az 5.2. ábrán lévő halmazként kezelt reláció A oszlopára vonatkozó átlagot akarjuk venni, akkor nem használhatjuk a halmaz modell A-ra vonatkozó vetítését, mivel ez esetben az átlagra 2-t kapunk. Ugyanis az 5.2. ábrán az A-nak csak két különböző értéke van (a 3 és az 1), ezeknek az átlaga pedig 2. Ha viszont az A oszlopra mint multihalmazra $\{1, 3, 1, 1\}$ tekintünk, akkor a valódi átlagot kapjuk, amelynek értéke – az 5.2. ábra 4 sorának felhasználásával számolva – 1.5 lesz. □

5.1.2. Multihalmazok egyesítése, metszete, különbsége

Mind a három műveletnek új értelmezése lesz multihalmazok esetén. Tegyük fel, hogy R és S multihalmazok, és legyen t az R n-szer, illetve az S m-szer előforduló sora. Megengedve azt is, hogy akár n, akár m, akár mindkettő 0 legyen. Ekkor:

- Az $R \cup S$ multihalmazon értelmezett egyesítésben a t sor $n + m$-szer fog előfordulni.
- Az $R \cap S$, tuple t multihalmazon értelmezett metszetben a t sor $\min(n, m)$-szer fog előfordulni.
- Az $R - S$ multihalmazon értelmezett különbségben a t sor $\max(0, n - m)$-szer fog előfordulni. Azaz ha a t többször fordul elő R-ben, mint S-ben, akkor t az $R - S$-ben annyiszor fordul elő, mint ahányszor előfordult R-ben mínusz ahányszor előfordul S-ben. Amennyiben viszont S-ben legalább annyiszor előfordul a t, mint R-ben, akkor t nem lesz benne $R - S$

különbségben. Informálisan tehát t minden egyes előfordulása S-ben „semlegesíti” t egy előfordulását R-ben.

5.4. példa. Legyen az R az 5.1. ábrán látható reláció, amely valójában egy multihalmaz, amelyben az $(1, 2)$ sor kétszer és a $(3, 4)$ sor egyszer szerepel. Legyen az S a következő multihalmaz:

A	B
1	2
3	4
3	4
5	6

Ebben az esetben az $R \cup S$ egyesítés eredménye az a multihalmaz, amelyben az $(1, 2)$ négyszer (háromszor az r miatt és egyszer az S miatt), a $(3, 4)$ háromszor és az $(5, 6)$ egyszer szerepel.

Az $R \cap S$ metszet eredménye a következő multihalmaz:

A	B
1	2
3	4

Mind az $(1, 2)$, mind a $(3, 4)$ sor csak egyszer szerepel az $R \cap S$ eredményben. Az $(1, 2)$ háromszor szerepel az R-ben és egyszer az S-ben és $\min(3, 1) = 1$. Hasonló módon a $(3, 4)$ sorra $\min(1, 2) = 1$. Az $(5, 6)$ sor egyszer szerepel az S-ben és nem szerepel az R-ben, $\min(0, 1) = 0$, tehát az $R \cap S$ eredményben ez a sor nem szerepel.

Az $R - S$ különbség eredménye a következő multihalmaz:

A	B
1	2
1	2

Az $(1, 2)$ háromszor szerepel az R-ben és egyszer az S-ben, $\max(0, 3 - 1) = 2$, ezért az $R - S$ különbségben ez a sor kétszer jelenik meg. A $(3, 4)$ sor egyszer szerepel az R-ben és kétszer szerepel az S-ben, $\max(0, 1 - 2) = 0$, tehát ez a sor nem szerepel az $R - S$ különbségben. Mivel az R-nek nincs más sora, ezért az $R - S$ eredmény sem tartalmazhat más sort.

Az $S - R$ különbség a következő multihalmaz:

A	B
3	4
5	6

A $(3, 4)$ sor egyszer jelenik meg, hiszen ki kell vonni az S-beli előfordulások számából az R-beli előfordulások számát. Az $(5, 6)$ sor ugyanilyen okból szerepel egyszer az eredményben. Ebben az esetben a multihalmazként kezelt eredmény történetesen halmaz. □

Multihalmaz-műveletek halmazokon

Képzeljük el, hogy van két halmazunk, az R és az S. Bármely halmaz tekinthető multihalmaznak, olyan multihalmaznak, amelynek történetesen mindegyik sora különböző. Tegyük fel, hogy az $R \cap S$ metszetet akarjuk kiszámolni, de R-et és S-t multihalmaznak tekintjük és az ennek megfelelő szabályt alkalmazzuk. Ebben az esetben ugyanazt az eredményt kapjuk, mintha az R-et és S-t halmaznak tekintettük volna. Ha R-et és S-t multihalmaznak tekintjük, akkor egy t sor annyiszor szerepel majd az $R \cap S$ eredményben, amennyi a két multihalmazban található előfordulások minimuma. Mivel az R és S halmazok, ezért a t előfordulási száma 0 vagy 1. Azaz mindegy, hogy halmaz- vagy multihalmazszabállyal számoljuk ki a metszetet, a t sor legfeljebb egyszer szerepelhet az $R \cap S$ eredményben, és pontosan egyszer akkor fog megjelenni, ha mind az R-ben, mind az S-ben szerepel a t sor. Hasonló módon, ha az $R - S$ vagy az $S - R$ kiszámításához multihalmaz-különbséget használunk, akkor ugyanazt az eredményt kapjuk, mintha halmazkülönbséget használtunk volna.

Az egyesítés azonban különböző eredményt ad attól függően, hogy az R-et és az S-t halmaznak vagy multihalmaznak tekintjük. Ha multihalmazszabályt alkalmazunk az $R \cup S$ kiszámításához, akkor előfordulhat, hogy az eredmény nem halmaz lesz annak ellenére, hogy mind az R, mind az S halmaz. Azaz, ha a t sor az R-ben és az S-ben is szerepel, akkor multihalmazszabályt alkalmazva a t sor kétszer jelenik meg az $R \cup S$ eredményben, ha viszont halmazszabályt alkalmazunk, akkor a t sor csak egyszerre jelenik meg az $R \cup S$ eredményben.

5.1.3. Multihalmazok vetítése

A multihalmazok vetítéséről már volt szó. Amint az 5.2. példában láthattuk, a vetítés során mindegyik sort külön kezeljük. Ha az R reláció az 5.2. ábrán látható multihalmaz, akkor a $\pi_{A,B}(R)$ eredménye az 5.1. ábrán látható multihalmaz.

Ha vetítés során különböző sorokból egy vagy több attribútum elhagyásával egyforma sorokat kapunk, akkor ezeket az egyforma sorokat nem kell kiküszöbölni a multihalmaz vetítésének eredményéből. Ha az 5.2. ábra R relációjának $(1, 2, 5)$, $(1, 2, 7)$ és $(1, 2, 8)$ sorait levetítjük az A és B attribútumokra, akkor mindhárom sor ugyanazt az $(1, 2)$ sort eredményezi. Ez azt jelenti, hogy a multihalmaz-eredményben az $(1, 2)$ sor háromszor szerepel, míg a hagyományos vetítéssel csak egyszer szerepelne.

Algebrai azonosságok multihalmazok esetén

Az algebrai azonosság két relációs algebrai kifejezés közötti ekvivalencia. A kifejezések argumentumai relációk. Az ekvivalencia szerint mindegy, hogy milyen relációkat helyettesítünk be, a két kifejezés ugyanazt az eredményt adja. Egy jól ismert példa az egyesítésre vonatkozó azonosság: $R \cup S = S \cup R$. Ez az azonosság fennáll, ha az R és S relációkat halmazként kezeljük, de akkor is fennáll, ha a relációkat multihalmazként kezeljük. Létezik azonban számos olyan azonosság, amely fennáll abban az esetben, ha a relációs algebrai műveleteket a hagyományos halmazelméleti módon értelmezzük, viszont nem érvényesek, ha a relációkat multihalmazként kezeljük. Ilyen azonosság például a különbség és az egyesítés disztributív volta: $(R \cup S) - T = (R - T) \cup (S - T)$. Ez az azonosság érvényes halmazokra, viszont nem érvényes multihalmazokra. Hogy lássuk, miért is nem igaz multihalmazok esetén, tegyük fel, hogy a t sor egyszer szerepel mindhárom relációban, az R-ben, az S-ben és a T-ben. Ebben az esetben a bal oldali kifejezés eredményében egyszer szerepel a t sor, a jobb oldali kifejezés eredményében viszont nem szerepel a t sor. Ha halmazszabályt alkalmazunk, akkor egyik oldal eredményében sem szerepel a t sor. Az 5.1.4. és 5.1.5. feladatokban megvizsgáljuk további algebrai azonosságok érvényességét multihalmazok esetén.

5.1.4. Multihalmazokon értelmezett kiválasztás

A multihalmazon végzett kiválasztás azt jelenti, hogy a kiválasztás feltételét minden egyes sorra alkalmazzuk. A multihalmazoknál megszokott módon itt sem küszöböljük ki az azonos sorokat.

5.5. példa. Legyen R a következő multihalmaz:

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

a $\sigma_{C \geq 6}(R)$ multihalmazos kiválasztás eredménye:

A	B	C
3	4	6
1	2	7
1	2	7

Azaz, az első sor kivételével mindegyik sor teljesíti a feltételt. A két utolsó sor az R-ben is megegyezik, így mindkettő szerepel az eredményben is. □

5.1.5. Multihalmazok szorzata

A multihalmazok Descartes-szorzatára vonatkozó szabály pontosan az, amit vártunk: az első reláció minden egyes sorát össze kell párosítanunk a második reláció mindegyik sorával függetlenül attól, hogy az adott sor hányszor szerepel az adott relációban. Következésképpen, ha az r sor m-szer szerepel az R relációban és az s sor n-szer szerepel az S relációban, akkor az rs sor mn-szer szerepel az $R \times S$ szorzat eredményében.

5.6. példa. Legyen az R és S reláció az 5.3. ábrán látható két reláció. Ekkor az $R \times S$ szorzat hat sorból áll, amint az az 5.3. (c) ábrán is látható. Megjegyezzük, hogy a hagyományos szorzatnál bevezetett, attribútumnevekre vonatkozó jelölés multihalmazokra is kitűnően alkalmazható. Ekképpen az R és S relációkban egyaránt megtalálható B attribútum kétszer jelenik meg a szorzatban, és előtagként a megfelelő reláció neve szolgál. □

A	B
1	2
1	2

(a) Az R reláció

B	C
2	3
4	5
4	5

(b) Az S reláció

A	$R.B$	$S.B$	C
1	2	2	3
1	2	2	3
1	2	4	5
1	2	4	5
1	2	4	5
1	2	4	5

(c) Az $R \times S$ szorzat

5.3. ábra. Multihalmazok szorzatának kiszámítása

5.1.6. Multihalmazok összekapcsolása

A multihalmazok összekapcsolása szintén nem szolgál különösebb meglepetéssel. Az első reláció minden egyes sorát összehasonlítjuk a második reláció valamennyi sorával, és eldöntjük, hogy az így kapott sorpárok sikeresen összekapcsolhatók-e vagy sem. Ha az összekapcsolás elvégezhető, akkor az eredményből nem kell kiküszöbölni a megegyező sorokat.

5.7. példa. Az 5.3. ábrán látható R és S reláció természetes összekapcsolásának eredménye, $R \bowtie S$:

A	B	C
1	2	3
1	2	3

Az R reláció $(1, 2)$ sora összekapcsolható az S $(2, 3)$ sorával. Mivel az R relációban az $(1, 2)$ sor kétszer szerepel és az S relációban a $(2, 3)$ sor egyszer szerepel, ezért az eredményben az $(1, 2, 3)$ kétszer jelenik meg. Az R és S egyéb sorai nem kapcsolhatók össze.

Ugyanezeken a relációkon végezzünk most egy théta-összekapcsolást:

$$R \bowtie_{R.B<S.B} S$$

Az eredményül kapott multihalmaz:

A	$R.B$	$S.B$	C
1	2	4	5
1	2	4	5
1	2	4	5
1	2	4	5

Az összekapcsolást a következő módon végezzük: az R $(1, 2)$ sora és az S $(4, 5)$ sora megfelel a feltételnek. Mivel mindkét sor kétszer szerepel a megfelelő relációkban, ezért az összekapcsolt sor $2 \times 2 = 4$ alkalommal jelenik meg az eredményben. Ezenkívül még összekapcsolhatnánk az R $(1, 2)$ sorát az S $(2, 3)$ sorával, de ez a sorpár nem teljesíti a feltételt, ily módon az eredményben sem jelenik meg. □

5.1.7. Feladatok

5.1.1. feladat. Tekintsük a 2.21. (a) ábrán szereplő `PC` relációt. Számítsuk ki a $\pi_{\texttt{sebesség}}(\texttt{PC})$ kifejezést. Mi lesz az értéke, ha a relációkat halmazokként kezeljük? És ha multihalmazokként? Mennyi lesz az értékek átlaga ebben a projekcióban, ha a relációkat halmazokként kezeljük? Mennyi, ha multihalmazokként?

5.1.2. feladat. Végezzük el az 5.1.1. feladatot a $\pi_{\texttt{merevlemez}}(\texttt{PC})$ kifejezéssel.