Alkalmazott Matematikai Lapok **15** (1990–91) 181–196

ELÁGAZÓ FÜGGŐSÉG RELÁCIÓS ADATBÁZISOKBAN*

DEMETROVICS JÁNOS, KATONA GYULA ÉS SALI ATTILA

Budapest

Egy relációs adatmodell úgy tekinthető, mint egy mátrix, melynek sorai az individuumok, az oszlopai az attributumok. Ezen elméletnek egy jól ismert fogalma a funkcionális függőségek fogalma. Ha A oszlopoknak egy halmaza, b pedig egy oszlop, b–ről azt mondják, hogy függ A–tól, ha nincs két olyan sor, amely b–ben különböző, de A–ban azonos. Ezt a fogalmat általánosítjuk itt. Legyenek $p \leq q$ pozitív egészek. Azt mondjuk, hogy b (p,q)–függ A–tól, ha nincs $q+1$ olyan sor, amelyek b–ben teljesen különböznek, de A minden oszlopában legfeljebb p különféle érték áll. A hagyományos függőségekre ismert több tétel van a cikkben általánosítva.

1. Bevezetés

Egy adatbázis felfogható, mint egy mátrix, amelynek oszlopai az *adatfajták, attributumok* (pl. név, születési hely, idő, stb.), míg egy-egy sor tartalmazza egy-egy *individuum* adatait. Jelölje X az attributumok (azaz a mátrix oszlopainak) halmazát. Legyen $A \subseteq X$, $b \in X$. Azt mondjuk (ARMSTRONG [1], CODD [2]), hogy b *funkcionálisan függ* A–tól (jelölésben: $A \rightarrow b$), ha az A–ban álló adatok már meghatározzák a b–ben álló adatot is, azaz nincs két olyan sor, amely A–ban egyezik és b–ben különbözik.

A funkcionális függőségek nagyon hasznosnak bizonyultak a gyakorlatban. A létező adatbáziskezelő rendszerek mind ezen a fogalmon alapszanak. Tekintsünk egy példát. Tegyük fel, hogy $X = \{x_1, x_2, x_3, x_4\}$, $x_1 \rightarrow x_2$ és $x_3 \rightarrow x_4$. Ha az egész mátrixot tároljuk a memóriában, az a legrosszabb esetben $4N_1N_3$ helyet jelent, ahol N_1 ill. N_3 jelöli az x_1 ill. x_3 által felvehető különböző értékek számát. Hiszen x_1 és x_3 egymástól függetlenül vehetik fel értékeiket (viszont x_2–t és x_4–et már meghatározzák), így a sorok száma maximálisan N_1N_3. Azonban — kihasználva az adott két funkcionális függőséget — rengeteg helyet megspórolhatunk. Ugyanis elegendő az x_1 és x_3 oszlopából álló mátrixon kívül (ami $2N_1N_3$ értéket tartalmaz), két kis mátrixot tárolnunk. Az egyik mátrix két oszlopa x_1 és x_2 értékeit tartalmazza. Az első oszlopban állnak x_1 lehetséges értékei, a másodikban az $x_1 \rightarrow x_2$ funkcionális függőség által meghatározott értékek. A másik mátrixot is hasonlóan építjük fel x_3–ból és x_4–ből. A tárolt értékek száma legfeljebb $2N_1N_3 + 2(N_1 + N_3)$, ami általában sokkal kisebb $4N_1N_3$–nál.

Jelen cikkben a funkcionális függőségnél egy általánosabb (gyengébb) fogalmat fogunk bevezetni. Először csak egy nagyon speciális esetben, majd illusztráljuk a

*Az 2575–ös számú Országos Tudományos Kutatási Alap által támogatva.

fogalom hasznosságát. Ha $A \subseteq X$, $b \in X$, akkor azt mondjuk, hogy b (1,2)-függ A-tól, ha az A-ban álló értékek „kétértelműen" meghatározzák a b-ben álló értékeket, azaz nincs 3 olyan sor amelyek mind megegyeznek A-ban, de 3 különböző értéket tartalmaznak b-ben. Jelölése $A \text{—}(1,2) \to b$. Hasonlóan, $A \text{—}(1,q) \to b$, ha nincs $q+1$ olyan sor, amelyek A-ban megegyeznek, de b-ben $q+1$ különböző értéket tartalmaznak.

Tegyük fel, hogy az adatbázis egy kamion útjait tartalmazza, pontosabban csak az általa érintett országokat. Az egyszerűség kedvéért tegyük fel, hogy egy útja során egy kamion pontosan 4 országot érint (beleértve az indulási és érkezési országot is) és nem megy vissza olyan országba, ahol már volt. Tegyük fel továbbá, hogy 30 ország jön szóba, egy országnak legfeljebb 5 szomszédja van. Jelölje x_1, x_2, x_3, x_4 az első, második, harmadik ill. negyedik országot mint attributumot. Könnyen látható, hogy $x_1 \text{—}(1,5) \to x_2$, $\{x_1, x_2\} \text{—}(1,4) \to x_3$ és $\{x_2, x_3\} \text{—}(1,4) \to x_4$. Most nem tudjuk a tárolt mátrix nagyságát csökkenteni, mint a funkcionális függőség ((1,1)-függőség) esetén, de tudjuk a mátrix elemeinek értékkészletét csökkenteni. Az eredeti mátrix minden eleme 30 féle értéket vehet fel, az országnevet, vagy azok valamilyen kódjait (5 bit). Tároljunk egy kis (legfeljebb $30 \times 5 \times 5 = 750$-ös) táblázatot, amely minden ország szomszédainak egy számozását tartalmazza, azaz hozzájuk rendeli a 0,1,2,3,4 számokat. Ekkor az x_2 attributumot helyettesíthetjük ezen számokkal, hiszen x_1 már megadja az induló országot, az $x_2{}^*$ attributum értéke a kis táblázatból megállapíthatóan meghatározza a második országot. Ugyanez érvényes x_3-ra is, de itt még csökkenthetjük a lehetséges értékek számát, ha a lehetséges x_1, x_2 párokhoz harmadikként csatlakozó országokat megszámozó táblázatot is megadunk. Ekkor az így kapott $x_3{}^*$ már csak 4 különböző értéket vehet fel. Ugyanez érvényes x_4-re is. Tehát, míg az eredeti mátrixban minden elem 5 bittel volt leírható, most 2 kis póttáblázat árán a második oszlopban állókat 3 bitre, a harmadik és negyedik oszlopban állókat 2 bitre csökkentettük.

Könnyen látható, hogy ez a gondolat teljesen hasonlóan alkalmazható minden olyan esetben, amikor valamilyen gráf útjait kell tárolni, ahol a gráfban a maximális fok sokkal kisebb a szögpontok számánál. Vagy másképpen fogalmazva, ha valamilyen folyamat állapotsorozatait kell tárolni, ahol az állapotok száma sokkal nagyobb az állapotok rákövetkező állapotainak maximális számánál. De más esetekben is, ha sok $(1,q)$-függés van, ahol q kicsi.

Az általános fogalom, amit tárgyalni fogunk, a (p,q)-függés ($1 \leq p \leq q$, egészek). Akkor mondjuk, hogy b az A-tól (p,q)-függ, ha nincs $q+1$ olyan sor, amely A minden oszlopában legfeljebb p különböző értéket tartalmaz, b-ben pedig $q+1$ csupa különbözőt. Ez nyilván általánosítása a funkcionális függőség fogalmának. A cikk fő célja bizonyos, a funkcionális függőségre vonatkozó tételek általánosítása a (p,q)-függőségekre.

2. A (p,q)-függőségek jellemzése

Ha adott az M adatbázis (vagy mátrix), definiáljuk a $\mathcal{J}_{Mpq}(A)$, $A \subseteq X$

függvényt a következőképpen:

$$\mathcal{J}_{M_{pq}}(A) = \{b : A \!-\!\!\!\!\!(p,q) \to b\}.$$

Ha egyébként nem okoz félreértést, az M, p, q indexeket elhagyjuk.

Könnyen látható, hogy ha $b \in A$, akkor $A \!-\!\!\!\!\!(p,q) \to b$, tehát

$$A \subseteq \mathcal{J}_{M_{pq}}(A). \tag{2.1}$$

Másrészt, ha $A \subseteq B$ és $A \!-\!\!\!\!\!(p,q) \to b$, akkor $B \!-\!\!\!\!\!(p,q) \to b$. Tehát

$$\mathcal{J}_{M_{pq}}(A) \subseteq \mathcal{J}_{M_{pq}}(B), \text{ ha } A \subseteq B. \tag{2.2}$$

A (2.1) és (2.2) feltételeket kielégítő halmaz-halmaz függvényeket ***növelő-monoton függvényeknek*** nevezzük. Ha egy $\mathcal{N}$ növelő-monoton függvényhez található egy olyan M mátrix, hogy $\mathcal{J}_{M_{pq}} = \mathcal{N}$, akkor azt mondjuk, hogy (p,q)–reprezentálható.

2.1. TÉTEL. *Ha X-en adott egy $\mathcal{N}$ növelő-monoton függvény, melyre $\mathcal{N}(\emptyset) = \emptyset$ teljesül, és $1 < q$, akkor $\mathcal{N}$ $(1,q)$–reprezentálható.*

Bizonyítás. Nevezzük *láncnak* X részhalmazainak egy $\emptyset \neq A_1 \subset \cdots \subset A_k$ sorozatát, amelyekre teljesülnek a következő feltételek:

$$\mathcal{N}(A_i) = A_{i+1} \quad (1 \leq i \leq k), \tag{2.3 }$$

$$\mathcal{N}(A_k) = A_k. \tag{2.4}$$

Egy ilyen adott L lánchoz készítsük el a következő $M(L)$ mátrixot. (Lásd a táblázatot.)

A_1	$A_2 - A_1$	$A_3 - A_2$	$\cdots$	$A_k - A_{k-1}$	$X - A_k$
1	1	1		1	1
1	1	1		1	2
1	.	.		.	3
.	.	.		.	4
.	.	.		.	.
.	1	$r+1$		$(k-2)r+1$	$kr+1$
1	2	$r+2$		$(k-2)r+2$	$kr+2$
1	3	$r+3$		$(k-2)r+3$	$kr+3$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
1	$r+1$	$2r+1$		$(k-1)r+1$	$(k+1)r+1,$

ahol $r = \lceil \frac{q}{2} \rceil$.

A mátrix minden oszlopa néhány egyessel kezdődik, majd egy helytől kezdve a természetes számok következnek növekvő sorrendben: $1, \ldots, 1, 1, 2, 3, 4, \ldots$. Az

$A_i - A_{i-1}$-be ($1 \leq i \leq k$, $A_0 = \emptyset$) tartozó oszlopok mind egyformák, ugyanez érvényes $X - A_k$-ra is. $X - A_k$ oszlopai legyenek ilyenek: $1, 2, 3, 4, \ldots$. Másrészt A_1 oszlopai álljanak csupa 1-ből. $A_2 - A_1$ oszlopai A_1 oszlopaihoz képest r-rel vannak eltolva, tehát az elejükön r-rel kevesebb az egyes, az utolsó elemük viszont $r + 1$. Általában, $A_i - A_{i-1}$ oszlopai az $A_{i-1} - A_{i-2}$ ($1 < i \leq k$) oszlopaihoz képest r-rel vannak eltolva. Ennyivel kevesebb az elejükön az egyes, és ennyivel nagyobb az utolsó elemük. $X - A_k$ oszlopai viszont már $2r$-rel vannak eltolva $A_k - A_{k-1}$ oszlopaihoz képest. $A_i - A_{i-1}$ ($1 \leq i \leq k$) a lánc definíciója miatt nem lehet üres. $X - A_k$ viszont lehet üres. Ekkor a mátrix ilyen oszlopokat nem tartalmaz.

A könnyebb érthetőség kedvéért bemutatjuk a $k = 4$, $q = 4$, $r = 2$ esetet:

A_1	$A_2 - A_1$	$A_3 - A_2$	$A_4 - A_3$	$X - A_4$
1	1	1	1	1
1	1	1	1	2
1	1	1	1	3
1	1	1	1	4
1	1	1	1	5
1	1	1	2	6
1	1	1	3	7
1	1	2	4	8
1	1	3	5	9
1	2	4	6	10
1	3	5	7	11

E mátrixnak a következő, könnyen ellenőrízhető tulajdonságait fogjuk felhasználni:

(i) A sorok száma $(k + 1)\lceil q/2 \rceil + 1$.

(ii) Ha $A_i - A_{i-1}$ $(1 < i \leq k)$ egy oszlopában két pozíció azonos számokat tartalmaz (ami csak egy lehet), és $j < i$ akkor $A_j - A_{j-1}$ megfelelő pozícióiban is azonos számok állnak (amik szintén egyesek).

(iii) Kiválasztva egy 1-est $A_i - A_{i-1}$ egy oszlopában, $A_{i+1} - A_i$ egy oszlopának megfelelő pozíciójában nem állhat más, csak 1 vagy 2 vagy 3 vagy $\ldots r + 1$. Ha viszont egy 1-től különböző s számot választunk ki $A_i - A_{i-1}$ egy oszlopában, akkor $A_{i+1} - A_i$ egy oszlopának megfelelő pozíciójában csak $s + r$ állhat ($1 \leq i < k$).

(iv) Ha $k \geq j > i + 1 \geq 2$, akkor $A_j - A_{j-1}$ egy oszlopában kiválasztható $2r + 1$ különböző szám (mégpedig az $1, 2, \ldots, 2r + 1$) úgy, hogy $A_i - A_{i-1}$ minden oszlopában a megfelelő helyeken csupa 1-es álljon. Jegyezzük meg, hogy $2r + 1 \geq q + 1$.

(v) $X - A_k$ egy oszlopában kiválasztható $2r + 1$ ($\geq q + 1$) különböző szám (mégpedig az $1, 2, \ldots, 2r + 1$) úgy, hogy $A_i - A_{i-1}$ ($1 \leq i \leq k$) minden oszlopában a megfelelő helyeken csupa 1-es álljon.

Vegyünk láncoknak egy olyan $L_1, \ldots, L_m$ halmazát, melyben minden olyan A, b párra, amelyre $A \neq \emptyset$, $b \notin \mathcal{N}(A)$ teljesül, található egy olyan lánc és benne egy olyan A_i halmaz, melyre

$$A \subseteq A_i \text{ és } b \notin \mathcal{N}(A_i) \tag{2.5}$$

fennálnak. Láncoknak ilyen halmazát kapjuk például, ha A_1-ként minden lehetséges nem üres A-t kiválasztunk. Ekkor $m = 2^{|X|} - 1$, de általában kevesebb lánc is elegendő.

Az $M(L_1), \ldots, M(L_m)$ mátrixokat írjuk egyszerűen egymás alá, de a bennük szereplő számokat úgy módosítsuk, hogy két mátrixban ne szerepeljen ugyanaz a szám. Ha történetesen valamely oszlopban kevesebb, mint $q+1$ különböző érték szerepel, akkor $M(L_1)$-ből vegyünk annyi különböző példányt, teljesen különböző számjegyekkel, hogy már minden oszlopban legyen $q+1$ különböző érték. Az így kapott M mátrix által definiált $\mathcal{J} = \mathcal{J}_{M1q}$ — mint látni fogjuk — éppen $\mathcal{N}$ lesz.

Ehhez két dolgot kell igazolnunk; 1) Ha $b \notin \mathcal{N}(A)$, akkor $b \notin \mathcal{J}(A)$ és 2) ha $b \in \mathcal{N}(A)$, akkor $b \in \mathcal{J}(A)$.

1) Tegyük fel tehát, hogy $b \notin \mathcal{N}(A)$. Ha $A = \emptyset$, akkor a $b \notin \mathcal{J}(\emptyset)$ állítás abból következik, hogy M minden oszlopában legalább $q+1$ különböző érték van. Ha viszont $A \neq \emptyset$, válasszunk egy L_v $(1 \leq v \leq m)$ láncot és benne egy A_i-t (2.5) szerint. Mivel $b \notin \mathcal{N}(A_i) = A_{i+1}$ (lásd (2.3)), így $b \in A_j - A_{j-1}$, $k \geq j > i+1$, vagy $b \in X - A_k$ teljesül. Az első esetben használjuk (iv)-et. Eszerint $M(L_v)$-ben kiválasztható $q+1$ sor, ami $A_i - A_{i-1}$ oszlopaiban csupa egyest tartalmaz, míg b-ben mind a $q+1$ érték különböző. (2.5) és (ii) miatt A minden oszlopában is csupa egyes áll. Tehát valóban $b \notin \mathcal{J}(A)$. Ha $b \in X - A_k$, akkor (v)-öt használva ugyanúgy járunk el.

2) Most tegyük fel, hogy $b \in \mathcal{N}(A)$. Feltehető, hogy $A \neq \emptyset$, mert $\mathcal{N}(\emptyset)$ üres. Tekintsünk egy tetszőleges L_v láncot: $A_1, \ldots, A_k$. Legyen $i = i(L_v) = k+1$, ha $A \cap (X - A_k)$ nem üres, ellenkező esetben viszont i legyen a legnagyobb index, amire $A \cap (A_i - A_{i-1})$ nem üres. $A \subseteq A_i$-ből (2.2) miatt $b \in \mathcal{N}(A) \subseteq \mathcal{N}(A_i) = A_{i+1}$ következik, ha $i < k$. Ellenkező esetben $b \in \mathcal{N}(A_i) = A_i$ (ahol $A_{k+1} = X$).

Vegyünk ki $q+1$ sort M-ből úgy, hogy ezekben a sorokban A minden oszlopa csupa azonos számból álljon. M konstrukciójából következik, hogy a soroknak ugyanazon $M(L_v)$-ben (vagy $M(L_1)$-nek ugyanabban a példányában) kell lenniük. Tehát $A \cap (A_i - A_{i-1})$ az $M(L_v)$ mátrixnak e $q+1$ sorában csupa egyes. Ha $b \in A_{i+1}$, akkor (iii)-ből következik, hogy b-ben e sorokban legfeljebb $r + 1 = \lceil q/2 \rceil + 1 \leq q$ különböző érték állhat. Ha $b \in A_i$, akkor b-ben e sorokban ráadásul csupa egyes áll. Tehát $b \in \mathcal{J}(A)$, amint azt bizonyítani kívántuk. □

2.1. *Probléma.* Igaz-e a 2.1 tétel állítása tetszőleges (p, q) $(p < q)$ függőségekre? Elhagyható-e az $\mathcal{N}(\emptyset) = \emptyset$ feltétel?

A választ jelenleg még a (2,3) esetre sem tudjuk.

Ha $p = q$, a helyzet jelentősen megváltozik. Mint azt [6]-ban már megmutattuk, $\mathcal{J} = \mathcal{J}_{Mpp}$ (2.1) és (2.2) mellett még egy fontos feltételt köteles kielégíteni:

2.2. ÁLLÍTÁS.

$$\mathcal{J}_{Mpp}(\mathcal{J}_{Mpp}(A)) = \mathcal{J}_{Mpp}(A). \tag{2.6}$$

Bizonyítás. Az egyik irányú tartalmazás (2.1)-ből és (2.2)-ből következik. Tehát csak azt kell igazolni, hogy $b \in \mathcal{J}(\mathcal{J}(A))$-ból következik $b \in \mathcal{J}(A)$. Vegyük a mátrix sorainak egy olyan halmazát, melyekben A összes oszlopa legfeljebb p különböző értéket tartalmaz. $\mathcal{J}$ definíciója szerint ekkor $\mathcal{J}(A)$ minden oszlopára is teljesül ez. A $b \in \mathcal{J}(\mathcal{J}(A))$ feltevés szerint b-ben is legfeljebb p különböző érték állhat e sorokban, s ez bizonyítja $b \in \mathcal{J}(A)$-t. □

A (2.1), (2.2) és (2.6) feltételeket kielégítő halmaz-halmaz függvényeket *lezárásoknak* nevezi az irodalom. Jól ismert (más formában lásd ARMSTRONG [1], ilyen alakban lásd pl. [5]), hogy $p = 1$-re az állítás megfordítható, azaz minden lezáráshoz található olyan M mátrix, melyre $\mathcal{J}_{M11}$ éppen ez az adott lezárás, vagyis, hogy minden lezárás (1,1)-reprezentálható. A következőkben belátjuk, hogy ugyanez igaz $p = 2$-re is, de $p > 3$-ra már általában nem.

A bizonyításhoz szükségünk van néhány, a lezárásokra vonatkozó, egyszerű fogalomra és állításra, amelyeket itt bizonyítás nélkül közlünk, megtalálhatók pl. [5]-ben. *Zártnak* nevezünk egy A halmazt az $\mathcal{L}$ lezárás szerint, ha $\mathcal{L}(A) = (A)$. A zárt halmazok összességét jelölje $\mathcal{Z} = \mathcal{Z}(\mathcal{L})$. Két zárt halmaz metszete zárt. $\mathcal{L}(A)$ egyenlő az A-t tartalmazó zárt halmazok metszetével. Jelölje továbbá $\mathcal{M} = \mathcal{M}(\mathcal{Z}(\mathcal{L}))$ azon zárt halmazok összességét, amelyek nem kaphatók meg tőlük különböző két zárt halmaz metszeteként. Bármely zárt halmaz előállítható $\mathcal{M}$-beliek metszeteként, következésképpen $\mathcal{L}(A)$ egyenlő az A-t tartalmazó $\mathcal{M}$-beli halmazok metszetével.

2.3. TÉTEL. *Minden lezárás (2,2)-reprezentálható.*

Bizonyítás. Legyen a lezárás $\mathcal{L}$, és $\mathcal{M} = \mathcal{M}(\mathcal{Z}(\mathcal{L})) = \{X, G_1, \ldots, G_m\}$. Egyelőre tegyük fel, hogy $m > 1$. Könnyen látható, hogy $G = \mathcal{L}(\emptyset)$ minden zárt halmaznak, s így minden G_i-nek részhalmaza.

Minden G_i-hez készítsünk az M mátrixban két sort, a $2i-1$-ediket és a $2i$-ediket. A G-nek megfelelő helyeken álljon 0. A $G_i - G$-nek megfelelő helyeken mindkét sorban álljon i. A többi helyeken a j-edik sorban $j + 2m$ álljon.

Be kell látnunk, hogy erre az M mátrixra $\mathcal{J}_{M22}(A) = \mathcal{L}(A)$ teljesül minden $A \subseteq X$-re. Azaz $b \in \mathcal{J}_{M22}(A)$ akkor és csak akkor, ha $b \in \mathcal{L}(A)$.

Tegyük fel, hogy $b \notin \mathcal{L}(A)$. Ha $A = \emptyset$, akkor felhasználva azt az állítást, hogy $\mathcal{L}(\emptyset)$ egyenlő az összes $\mathcal{M}$-beli metszetével, található egy G_i, amelyre $b \notin G_i$ teljesül. Ez ad két különböző értéket b oszlopában. X, ami mindig egy G, ad egy újabbat. Egy tetszőleges harmadik sorban vagy 0 áll, vagy újabb érték. Tehát b oszlopában van legalább 3 különböző érték, $b \notin \mathcal{J}(\emptyset)$. Ha $A \neq \emptyset$, akkor $\mathcal{L}(A)$ ismét egyenlő az A-t tartalmazó $\mathcal{M}$-beliek metszetével, tehát van egy olyan G_i, amelyre $A \subseteq G_i$ és $b \notin G_i$ fennállnak. De ekkor a G_i-nek megfelelő két sor A-ban megegyezik, b-ben viszont különbözik. E két sort kiegészítve bármely további sorral, találtunk 3 sort, ami bizonyítja, hogy $b \notin \mathcal{J}_{M22}(A)$ valóban igaz.

Tegyük most fel, megfordítva, hogy $b \in \mathcal{L}(A)$. Két esetet különböztetünk meg. Ha $A \subseteq \mathcal{L}(\emptyset)$, akkor (2.2) miatt $\mathcal{L}(A) \subseteq \mathcal{L}(\emptyset)$. Ekkor $b \in G$, tehát b oszlopában csupa 0 áll. $b \in \mathcal{J}_{M22}(\emptyset)$ valóban következik. Ha viszont $A \not\subseteq \mathcal{L}(\emptyset)$, akkor A–nak van egy $\mathcal{L}(\emptyset)$–n kivüli a eleme. Vegyünk ki 3 sort a mátrixból, amelyek A oszlopaiban legfeljebb két különböző értéket tartalmaznak. Az a oszlopot tekintve, azt nyerjük, hogy a három sor közül kettőnek ugyanahhoz a G_i–hez kell tartoznia. Továbbá fenn kell állnia $A \subseteq G_i$–nek is. G_i zártsága és (2.2) miatt $\mathcal{J}(A) \subseteq G_i$, tehát ez a két sor b–ben is egyforma. $b \in \mathcal{J}_{M22}(A)$ következik.

Az $m = 1$ esetben egy csupa új elemből álló harmadik sor hozzáadása segít. □

Vezessük be a következő jelölést:

$$\mathcal{L}^k{}_n(A) = \begin{cases} A, & \text{ha } |A| < k, \\ X, & \text{ha } |A| \geq k, \end{cases}$$

ahol k egy egész, $1 \leq k \leq |X|$, és $A \subseteq X$. $\mathcal{L}^k{}_n$ nyilván lezárás.

2.4. TÉTEL. *Ha $p > 2$ és $n > 6$, akkor $\mathcal{L}^2{}_n$ nem (p,p)–reprezentálható.*

Bizonyítás. Tegyük fel indirekte, hogy van egy olyan n oszlopú M mátrix, amelyik $\mathcal{L}^2{}_n$–t (p,p)–reprezentálja. Legyen M sorainak száma minimális. Mivel $\mathcal{L}^2{}_n$–ben az üres halmaz lezárása önmaga, ezért M minden oszlopában van legalább $p+1$ különböző érték. Ha az a oszlopban minden érték különböző lenne, akkor $b \in \mathcal{L}^2{}_n(\{a\})$ teljesülne egy tetszőleges $b \in X$–re, ellentmondásban $\mathcal{L}^2{}_n$ definíciójával.

Tegyük fel most, hogy az r és s sorok egyenlőek az a és a b oszlopokban. Definíció szerint, $c \in \mathcal{L}^2{}_n(\{a,b\})$ fennáll tetszőleges $c \in X$–re. Válasszunk ki r–hez és s–hez még $p-1$ egy sort úgy, hogy értékeik c–ben különbözzenek mind r–től, mind s–től, és egymástól is. (Ez megtehető, mert van a c oszlopban is $p+1$ különböző érték.) E $p+1$ sorban a és b egyaránt legfeljebb p különböző értéket vesz fel, tehát c is. Ez csak úgy lehet, ha r és s megegyeznek c–ben is. Vagyis az r és s sorok végig megegyeznek. Ez ellentmondásban van a minimalitással, tehát nem egyezhet meg két sor két különböző helyen.

Tegyük fel, hogy az első oszlopban a t–edik és az u–adik sorok egyeznek meg, míg a második oszlopban az r–edik és s–edik sorok ($t \neq u$, $r \neq s$). Az előző bekezdés szerint $\{t,u\} \neq \{r,s\}$, így csak a következő két esetet különböztetjük meg: (i) $t = r, u \neq s$, (ii) mind a négy sor különböző.

(i) Az első és második oszlop legfeljebb 2 különböző értéket vesz fel e három sorban. $\mathcal{L}^2{}_n$ definíciója miatt minden oszlop is csak legfeljebb két értéket vehet fel ezekben a sorokban. Ha az oszlopok száma több mint 3, akkor kell lenni olyan két oszlopnak, amelyek a fenti három sor valamelyik kettőjében megegyeznek, ellentmondásban a korábban igazolt állítással.

(ii) Az előző esethez hasonlóan belátható, hogy minden oszlop legfeljebb 3 értéket tartalmazhat a t, u, r, s sorokban. Mivel az egyezésekre csak 6–féle lehetőség van, $n > 6$ miatt ismét lesz két olyan oszlop, amely két adott sorban megegyezik. Az ellentmondás bizonyítja állításunkat. □

A most következőkben egy karakterizációt adunk arra, hogy mikor lehet egy lezárást (p,p)–reprezentálni. Egy ehhez szükséges definíció:

Legyen $\mathcal{B} = \{A_{i,j}\}$ az n–elemű X halmaz egy két indexes halmazrendszere, ahol $1 \leq i < j \leq m$. Akkor mondjuk, hogy teljesíti a háromszögfeltételt, ha minden $i < j < k$–ra $A_{i,j}$, $A_{j,k}$ és $A_{i,k}$ közül bármely kettő metszetét tartalmazza a harmadik.

2.5. LEMMA. *Akkor és csak akkor található olyan n oszlopú és m sorú mátrix melynek i–edik és j–edik sorai éppen az $A_{i,j}$–nek megfelelő helyeken egyeznek meg, ha $\{A_{i,j}\}$ kielégíti a háromszögfeltételt.*

Bizonyítás. A csak akkor rész triviális. Tegyük tehát fel, hogy a háromszögfeltétel teljesül és konstruáljunk egy megfelelő mátrixot. Méghozzá mohó módon, soronként. Tegyük fel, hogy az első k sor már megvan, és kielégíti a feltételt. A $k+1$–edik sor konstruálásánál csak az lehet a probléma, hogy egy helyen két különböző értékkel kell megegyezzen, mondjuk a g–edik és h–adik $(g < h)$ sorokban felvett értékekkel. Ez viszont ellentmond annak, hogy $A_{g,h}$, $A_{g,k+1}$ és $A_{h,k+1}$ kielégítik a háromszögfeltételt. □

Most már meg tudjuk adni a karakterizációt.

2.6. TÉTEL. *Az $\mathcal{L}$ lezárás akkor és csak akkor (p,p)–reprezentálható, ha létezik egy $\mathcal{B} = \{A_{i,j}\}$ $(1 \leq i < j \leq m)$ halmazrendszer úgy, hogy kielégíti a háromszögfeltételt, a következő alakú halmazok mind zártak $\mathcal{L}$ szerint:*

$$\bigcup_{0 \leq r < s \leq p} A_{j_r, j_s} \tag{2.7}$$

$(1 \leq j_0, j_1, \ldots, j_p \leq m$ tetszőlegesen rögzített különböző egészek), és minden zárt halmaz előállítható (2.7) alakú halmazok metszeteként.

Bizonyítás. Bizonyítsuk először a következő lemmát.

2.7. LEMMA. *Tegyük fel, hogy az m sorú M mátrix (p,p)–reprezentálja az $\mathcal{L}$ lezárást és M i–edik és j–edik sorai az $A_{i,j}$ halmazban egyeznek meg. $A \subseteq X$ akkor és csak akkor zárt, ha (2.7) alakú halmazok metszete.*

Bizonyítás. A (p,p)–függés definíciója szerint $A \not\to (p,p) \to b$ akkor és csak akkor teljesül egy M mátrixban, ha található $p+1$ sor, amely A oszlopaiban legfeljebb p értéket vesz fel, viszont a b oszlopban mind a $p+1$ érték különböző. Ezt úgy is lehet fogalmazni, hogy valamilyen $1 \leq j_0, j_1, \ldots, j_p \leq m$ különböző egészekre

$$A \subseteq \bigcup_{0 \leq r < s \leq p} A_{j_r, j_s} \not\ni b$$

teljesül. A tehát zárt akkor és csak akkor, ha ilyen található minden $b \notin A$–hoz. Ez utóbbi viszont akkor és csak akkor teljesül, ha A (2.7) alakú halmazok metszete.

□

Térjünk vissza a tétel bizonyításához. Ha (p,p)–reprezentálható, akkor a reprezentáló M mátrix definiálja az $A_{i,j}$ soregyenlőséghalmazokat. Ezek a 2.5 lemma miatt kielégítik a háromszögfeltételt. Egy (2.7) alakú halmaz ilyen alakú halmazoknak (egynek) a metszete, tehát a 2.7 lemma miatt zárt. Ugyancsak a 2.7 lemmából következik, hogy minden zárt halmaz (2.7) alakúak metszete.

Megfordítva, ha találhatók a tétel feltételeit kielégítő $A_{i,j}$ halmazok, a háromszögfeltétel miatt létezik egy M mátrix, amelyben a soregyenlőséghalmazok éppen $A_{i,j}$–k. Az $\mathcal{L}$ szerint zárt halmazok a tétel feltételei miatt előállnak (2.7) alakú halmazok metszeteként. Megfordítva, a nem zártak nem állíthatók elő, mert a (2.7) alakúak szerint zártak, s tudjuk, hogy zárt halmazok metszetei is zártak. Így a 2.7 lemma befejezi a bizonyítást. □

Sajnos a tétel feltétele algoritmikusan nem hatékony. A következő következmény mutatja azonban, hogy mégse teljesen haszontalan.

2.8. ÁLLÍTÁS. *Tegyük fel, hogy az $\mathcal{L}$ lezárás olyan, hogy $\mathcal{M}(\mathcal{Z}(\mathcal{L})) = \{G_1,\ldots,G_t\}$ zárt az unió képzésére. Ekkor (p,p)–reprezentálható minden p–re.*

Bizonyítás. Tegyük fel először, hogy $t \geq p$. Legyen a 2.6 tételbeli $m = 2t$, $A_{2i-1,2i} = G_i$ $(1 \leq i \leq t)$, míg a többi A legyen az üres halmaz. A rendszer kielégíti a háromszögfeltételt. A (2.7) alakú halmazok a G–k uniói lesznek, a $t \geq p$ feltétel miatt a $p+1$ darab index választható úgy, hogy (2.7) éppen egy (bármelyik) G legyen. A feltétel szerint a G–k uniói mind zártak, másrészt minden zárt halmaz előállítható $\mathcal{M}$–beliek (azaz G–k) metszeteként. Így a 2.6 tételből következik az állítás.

Ha $t < p$, akkor G–k valamelyikét ismételjük meg annyiszor, hogy m elérje a szükséges méretet. □

A következő egyszerű állítást bizonyítás nélkül közöljük.

2.9. ÁLLÍTÁS. *Legyen az $\mathcal{L}$ lezárás egy n–elemű halmazon megadva. Ha $\mathcal{L}$ $(2n,2n)$–reprezentálható, akkor (k,k)–reprezentálható is minden $k > 2n$ esetén.*

Összefoglalva, a kérdés nagyrészt még nyitott maradt:

2.2. *Probléma.* Keressünk egy algoritmikusan hatékony jellemzését azon lezárásoknak, amelyek (p,p)–reprezentálhatóak.

A következő probléma azonban könnyebbnek látszik. Legyen $\mathcal{N}$ egy növelő-monoton függvény az X halmazon. A K halmazt kulcsnak mondjuk, ha $\mathcal{N}(K) = X$. K *minimális kulcs,* ha kulcs, és nincs olyan valódi része, ami kulcs. Könnyen látható, hogy a minimális kulcsok nem tartalmazhatják egymást, így a minimális kulcsok $\mathcal{K}$ rendszere kielégíti a *Sperner-feltételt:* Ha $K_1, K_2 \in \mathcal{K}$, $K_1 \neq K_2$, akkor $K_1 \not\subset K_2$. Ezt úgy is mondhatjuk, hogy $\mathcal{K}$ *Sperner-rendszer.* Az X–en adott *Sperner-rendszerről* azt mondjuk, hogy (p,q)–reprezentálható $(p < q)$, ha X–en van egy növelő-monoton függvény, ami (p,q)–reprezentálható és amiben a minimális kulcsok rendszere éppen $\mathcal{K}$.

A (p,p)–reprezentálhatóság definíciója analóg, csak egy lezárás keresendő.

2.3. *Probléma.* Igaz-e hogy minden nem-üres *Sperner-rendszer* (p,q)–reprezentálható bármely $p < q$ mellett?

3. A (p,q) függőség egymáshoz való viszonyai

Jelölje $(p,q) \Rightarrow (p',q')$ azt, hogy ha $b \in \mathcal{J}_{Mpq}(A)$–ból minden M mátrixra következik $b \in \mathcal{J}_{Mp'q'}(A)$. A következő állítások triviálisak.

3.1. LEMMA.

$$(p,q) \Rightarrow (p,q+1)$$
$$(p,q) \Rightarrow (p-1,q).$$

□

Több állítható, ha feltesszük, hogy az M mátrix minden oszlopában legalább m különböző érték áll. Erre az esetre használjuk a $(p,q) \underset{m}{\Rightarrow} (p',q')$ jelölést.

3.2. LEMMA.

$$(p,q) \underset{q+1}{\Rightarrow} (p-1,q-1), \quad de \quad (p,q) \underset{q}{\not\Rightarrow} (p-1,q-1).$$

Bizonyítás. Az első állítás bizonyításához tegyük fel, hogy valamely M mátrixban $b \in \mathcal{J}_{Mpq}(A)$. Bizonyítani kivánjuk, hogy $b \in \mathcal{J}_{M,p-1,q-1}(A)$. Ha ez utóbbi nem teljesülne, akkor a mátrixnak volna q olyan sora, amely A–ban legfeljebb $p-1$ különböző értéket vesz fel, míg b–ben q különböző értéket. A feltevés szerint a b oszlopban van legalább $q+1$ különböző érték. Az előző q sort egészítsük ki egy újabb sorral úgy, hogy b–ben $q+1$ különböző érték legyen. A semelyik oszlopában se lehet több, mint p érték, s ez ellentmondásban van $b \in \mathcal{J}_{Mpq}(A)$–val.

A második állítást egy olyan mátrix bizonyítja, amelyben b oszlopa pontosan q különböző értéket tartalmaz, A minden oszlopa pedig legfeljebb $p-1$–et. □

3.3. LEMMA. $(p,q) \underset{q}{\not\Rightarrow} (1,q-1)$.

Bizonyítás. A következő ellenpélda adja a bizonyítást, ahol az első oszlop A oszlopait jelképezi, a második pedig b–ét.

$$\begin{array}{cc}
1 & 1 \\
1 & 2 \\
\vdots & \vdots \\
1 & q \\
2 & 1 \\
2 & 2 \\
\vdots & \vdots \\
2 & q \\
\vdots & \vdots \\
q & 1 \\
q & 2 \\
\vdots & \vdots \\
q & q
\end{array}$$

□

3.4. Lemma. $(p,q) \underset{m}{\not\twoheadrightarrow} (1, q-p), \quad (p < q)$.

Bizonyítás. A következő ellenpélda adja a bizonyítást, ahol az első oszlop A oszlopait jelképezi, a második pedig b-ét.

$$\begin{array}{cc}
1 & 1 \\
1 & 2 \\
\vdots & \vdots \\
1 & q-p+1 \\
2 & q-p+2 \\
\vdots & \vdots \\
m & q-p+m
\end{array}$$

□

3.5. Lemma. $(p,q) \underset{m}{\not\twoheadrightarrow} (p+1, N), \quad (p+1 \leq N)$.

Bizonyítás. Először egy olyan konstrukciót adunk, ami $(p,q) \not\twoheadrightarrow (p+1, N)$-et bizonyítja. A mátrixnak $N+1$ sora van, b-ben sorban az $1, 2, \ldots, N+1$ számok állnak. A oszlopaiban az $1, 2, \ldots, p+1$ számok állhatnak. Az oszlopokat úgy válasszuk meg, hogy bármely $p+1$ sorhoz legyen egy olyan oszlop, amelyikben $p+1$ különböző érték áll. Ha A oszlopainak száma elég nagy, ez elérhető. Könnyen látható, hogy ezen M mátrixban $b \in \mathcal{J}_{Mpp}(A)$ azaz a 3.1 lemma miatt $b \in \mathcal{J}_{Mpq}(A)$. Másrészt azonban $b \notin \mathcal{J}_{M,p+1,N}(A)$.

Már csak úgy kell módosítanunk M-et, hogy minden oszlopban legalább m érték legyen. Írjunk az $N+1+i$-edik sorba csupa $N+1+i$-t $(1 \leq i \leq m-p-1)$. Ez a módosítás nem rontja el a fenti tulajdonságot. □

3.6. TÉTEL. *Legyen $m > q$.*

$$(p, q) \underset{m}{\Rightarrow} (p', q'), \tag{3.1}$$

akkor és csak akkor teljesül, ha $1 \leq p' \leq p$, és $q - p \leq q' - p'$. Ha viszont $m \leq q$, akkor (3.1) szükséges és elégséges feltétele $1 \leq p' \leq p$ és $q \leq q'$.

Bizonyítás. A 3.1–3.5 lemmákból az állítás könnyen következik. □

Vegyük észre, hogy a 3.5 lemma bizonyításában A-nak elég nagynak kell lennie. Ez azt jelenti, hogy kis méretű A-ra a $b \in \mathcal{J}_{Mpq}(A)$ feltevésből még következik $b \in \mathcal{J}_{M,p+1,N}(A)$. Ha például $|X| = n$ is kisebb, mint ez a határ, akkor itt minden A-ra is teljesül ez a következtetés. Ez veti fel a következő problémát.

3.1. *Probléma.* Milyen méretű A-kra teljesül, hogy $b \in \mathcal{J}_{Mpq}(A)$-ból következik $b \in \mathcal{J}_{M,p+1,N}(A)$ minden M-re (p, q és N rögzített)?

Két speciális esetet bizonyítás nélkül közlünk:

3.7. ÁLLÍTÁS. *Ha $|A| < \lceil \log(N+1) \rceil$, akkor $b \in \mathcal{J}_{M11}(A)$-ból következik $b \in \mathcal{J}_{M2N}(A)$ minden M mátrixra, de ha $|A| \geq \lceil \log(N+1) \rceil$, akkor nem.* □

3.8. ÁLLÍTÁS. *Ha*

$$|A| < \left\lceil \frac{q+2}{2(q-p+1)} \right\rceil,$$

akkor $b \in \mathcal{J}_{Mpp}(A)$-ból következik $b \in \mathcal{J}_{M,p+1,q+1}(A)$ minden M mátrixra, de ha A ennél nagyobb, akkor nem. □

4. Minimális reprezentáció

E fejezetben azzal a kérdéssel foglalkozunk, hogy ha létezik reprezentáció, akkor mi a reprezentáló M mátrix minimális sorszáma. Ha $\mathcal{N}$ egy növelő-monoton függvény, jelölje $s_{pq}(\mathcal{N})$ ezt a minimális sorszámot. A 2.1 tétel biztosítja, hogy $s_{1q}(\mathcal{N})$ létezik. Az ott szereplő konstrukcióban a táblázatban szereplő mátrix sorainak száma (lásd (i)) $= (k+1)\lceil q/2 \rceil + 1 \leq 2nq$. Másrészt annyi ilyen mátrixot veszünk, amennyi a szükséges láncok száma. Ez legfeljebb 2^n. Tehát a következő igaz.

4.1. TÉTEL. *Ha $q > 1$ és $\mathcal{N}$ egy tetszőleges növelő-monoton függvény, akkor*

$$s_{1q}(\mathcal{N}) < 2qn2^n.$$

□

Ez a becslés lényegesen javítható volna, ha be tudnánk látni, hogy viszonylag kevés lánc elég a (2.5) feltétel kielégítéséhez. A másik javítási lehetőség a láncok tényleges hosszának figyelembevétele. Általában a növelő-monoton függvények szerkezetének tanulmányozása szükséges. Mindenesetre a 4.1 tétel becslése sokkal rosszabb, mint amit $s_{11}(\mathcal{L})$-re ismerünk (n alatt $\lfloor n/2 \rfloor$, lásd [3]).

A továbbiakban csak a $p = q = 2$ esettel foglakozunk. $s_{pp}(\mathcal{K}) = s_p(\mathcal{K})$ ill. $s_{pp}(\mathcal{L}) = s_p(\mathcal{L})$ jelölje a $\mathcal{K}$ *Sperner-rendszer* ill. az $\mathcal{L}$ lezárás (p,p)-reprezentálásához szükséges minimális sorszámot.

4.2. TÉTEL.

$$s_2(\mathcal{K}) < 2\binom{n}{\lfloor \frac{n}{2} \rfloor}.$$

Bizonyítás. A 2.3 tétel bizonyításából kiolvasható. □

Ez az eredmény már csak egy kettes faktorral rosszabb az (1,1) eset becslésénél (lásd [3]).

4.3. TÉTEL. *Van olyan Sperner-rendszer, melyre*

$$s_2(\mathcal{K}) \geq \frac{1}{n^2}\binom{n}{\lfloor \frac{n}{2} \rfloor}.$$

Bizonyítás. A bizonyítás szó szerint ugyanaz, mint [4] bizonyítása. □

4.4. LEMMA.

$$\binom{s_2(\mathcal{L}^k{}_n)}{3} \geq \binom{n}{k-1}.$$

Bizonyítás. Legyen M egy olyan $s_2(\mathcal{L}^k{}_n)$ sorú mátrix, amely $(2,2)$–reprezentálja $\mathcal{L}^k{}_n$–t. Ha B egy $k-1$–elemű részhalmaza X–nek, akkor van három olyan sora M–nek melyek legfeljebb két különböző értéket tartalmaznak B–beli oszlopokban, és van olyan oszlop, ahol három különbözőt. Ha B–hez a (b,c,d) sorok, C–hez pedig az (e,f,g) sorok tartoznak, akkor $(b,c,d) \neq (e,f,g)$, mert különben $B \cup C$ nem lenne kulcs. Azaz minden $k-1$ elemű halmazhoz különböző sorhármas tartozik, vagyis a lehetséges sorhármasok száma legalább akkora mint a $k-1$–elemű oszlophalmazok száma. □

Bizonyos speciális k értékekre $s_2(\mathcal{L}^k{}_n)$–t pontosan is meg tudjuk határozni, másokra pedig jó felső becslést tudunk adni:

4.5. TÉTEL.
(i) $s_2(\mathcal{L}^1{}_n) = 3$,
(ii) $s_2(\mathcal{L}^2{}_n) = 2n$, *ha* $n > 3$,
(iii) $s_2(\mathcal{L}^n{}_n) \leq n+2$.

Bizonyítás. (i) Az $s_2(\mathcal{L}^1{}_n) \geq 3$ egyenlőtlenség a 4.4 lemmából következik. A másik irányt a következő konstrukció adja:

$$\begin{matrix} 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \\ 2 & 2 & \dots & 2 \end{matrix}$$

(ii) Az $s_2(\mathcal{L}^2{}_n) \leq 2n$ egyenlőtlenség a 2.3 tétel bizonyításából kiolvasható. A másik irányú egyenlőtlenséget direkt bizonyítjuk. A 4.4 lemma szerint minden oszlophoz van három sor, hogy azok legfeljebb két különböző értéket tartalmaznak ezen oszlopban. Azaz minden oszlophoz van két sor, melyek ott megegyeznek. Belátjuk, hogy a különböző oszlopokhoz tartozó sorpárok diszjunktak.

a) Tegyük fel, hogy az i és j oszlopokban az r és s sorok megegyeznek. Mivel $\{i,j\} \multimap (2,2) \to X$ szükséges, ezért r és s az összes többi oszlopban is meg kell egyezzen. Ekkor r–et elhagyva, a kapott mátrix még mindig reprezentálja $\mathcal{L}^2{}_n$–t.

b) Ha van 3 olyan sor, hogy az i és j oszlopokban is legfeljebb két értéket vesznek fel, akkor az összes többi oszlopban is legfeljebb két különböző értéket vesznek fel. Ekkor viszont, ha $n > 3$, akkor az a) esetre jutunk vissza.

Azt kaptuk tehát, hogy az egyes oszlopokhoz tartozó párok diszjunktak. Jegyezzük meg, hogy $s_2(\mathcal{L}^2{}_2) = 4$, $s_2(\mathcal{L}^2{}_3) = 5$.

(iii) Az $s_2(\mathcal{L}^n{}_n) \leq n + 2$ felső korlátot a következő konstrukció bizonyítja:

$$\begin{array}{cccc} 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \\ 2 & 0 & \dots & 0 \\ 0 & 2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2. \end{array}$$

□

4.1. *Probléma.* Határozzuk meg $s_2(\mathcal{L}^n{}_n)$–et, $s_2(\mathcal{L}^3{}_n)$–at és $s_2(\mathcal{L}^4{}_n)$–et.

4.2. *Probléma.* Határozzuk meg $s_2(\mathcal{L}^k{}_n)$ nagyságrendjét rögzített k és nagy n esetére.

Az adatbázisok elméletében fontos szerepe van a lezárások direkt szorzatának. Ha $\mathcal{L}$ egy lezárás az U alaphalmazon, $\mathcal{N}$ pedig a V halmazon ($U \cap V = \emptyset$), akkor $\mathcal{L}$ és $\mathcal{N}$ direkt szorzata az $U \cup V$ alaphalmazon a következőképpen definiált függvény:

$$(\mathcal{L} \times \mathcal{N})(A) = \mathcal{L}(A \cap U) \cup \mathcal{N}(A \cap V).$$

Könnyen látható, hogy ez is egy lezárás.

4.6. TÉTEL. *Legyenek $\mathcal{L}$ és $\mathcal{N}$ lezárások az U ill. V alaphalmazokon. Ekkor*

$$s_p(\mathcal{L} \times \mathcal{N}) \leq s_p(\mathcal{L}) + s_p(\mathcal{N}) - p.$$

Bizonyítás. Ha $\mathcal{L}$ vagy $\mathcal{N}$ valamelyike nem reprezentálható, akkor értelmezzük az s_p–t végtelennek. Ekkor az egyenlőtlenség triviális. Legyen tehát M_1 egy minimális reprezentáló mátrixa $\mathcal{L}$–nek, M_2 pedig $\mathcal{N}$–nek. Képezzük a következő mátrixot:

$$M = \begin{pmatrix} Q & W \\ R & T \\ Y & P \end{pmatrix},$$

ahol Q az M_1-ből az utolsó p sor levágásával keletkezik. W az M_2 első sora annyiszor véve, ahány sora van Q-nak, R az M_1, T pedig az M_2 első p sora, P úgy keletkezik M_2-ből, hogy első p sorát levágjuk, végül Y pedig M_1 utolsó sora annyiszor véve, ahány sora van P-nek.

Tegyük fel, hogy $y \in U$ és $y \notin (\mathcal{L} \times \mathcal{N})(A)$ valamely $A \subseteq U \cap V$-re. Ez pontosan akkor van, ha $y \notin \mathcal{L}(A \cap U)$, azaz M_1-nek van $p+1$ sora, melyek A-n legfeljebb p értéket vesznek fel, de y-n mind különböznek. Ez a $p+1$ sor fellép az M mátrixban is, mégpedig úgy, hogy a V-re való kiterjesztésükben legfeljebb p különböző érték lép fel, azaz $y \notin \mathcal{J}_{Mpp}(A)$.

Legyen most $y \in (\mathcal{L} \times \mathcal{N})(A)$, és legyenek $r_0, \dots, r_p$ olyan különböző sorok, melyek A-ban legfeljebb p értéket vesznek fel. Ha az $r_0, \dots, r_p$ közül legalább kettő esik az Y részre, illetve még R utolsó sora valamelyik, akkor y-ban az a két sor megegyezik, tehát $r_0, \dots, r_p$ legfeljebb p értéket vesz fel y-on. Ha viszont csak legfeljebb egy sor esik az r_j-k közül a mátrix „alsó" felébe, akkor az A-n és y-on felvett értékeik ugyanazok, mint az M_1 megfelelő soráé, tehát $y \in \mathcal{J}_{Mpp}$.

4.3. *Probléma.* Milyen feltételekkel áll a 4.6 tételben egyenlőség?

IRODALOM

[1] Armstrong, W. W., "Dependency structures of data base relationships ", *Information Processing* **74** (North Holland, Amsterdam, 1974), 580–583.

[2] Codd, E. F., "A relational model of data for large shared data banks ", *Comm. ACM* **13** (1970), 377–387.

[3] Demetrovics, J., Füredi, Z. and Katona, G. O. H., "Minimum matrix representation of closure operations ", *Discrete Appl. Math.* **11** (1985), 115–128.

[4] Demetrovics, J. and Gyepesi, Gy., "A note on minimal matrix representation of closure operations ", *Combinatorica* **3** (1983), 177–180.

[5] Demetrovics, J. and Katona , G. O. H., "Combinatorial problems of database models ", *Coll. Math. Soc. János Bolyai* **42**, Algebra, Combinatorics and Logic in Computer Science, Győr (Hungary), 1983, 331–353.

[6] Demetrovics, J. and Katona, G. O. H., "Extremal combinatorial problems of database models ", MFDBS' 87, 1st Symposium on Mathematical Fundamentals of Database Systems, Dresden, GDR, January, 1987, Springer Verlag, Lecture Notes in Computer Science, 1987.

(Beérkezett: 1988. november 28.)
(Végleges változat: 1991. szeptember 5.)

DEMETROVICS JÁNOS
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1132 BUDAPEST, VICTOR HUGO U. 18

KATONA GYULA ÉS SALI ATTILA
MTA MATEMATIKAI KUTATÓ INTÉZET
1053 BUDAPEST, REÁLTANODA U. 13–15

BRANCHING DEPENDENCY IN RELATIONAL DATABASES

J. Demetrovics, G. O. H. Katona and A. Sali

A relational database model can be considered as a matrix, where the individuums are the rows and the attributes are the columns. A well-known notion in this theory is the notion of functional dependencies. If A is a set of columns and b is a column then b is said to depend on A if there are no two rows different in b but identical in A. This concept is generalized here. Let $p \leq q$ be positive integers. We say that b (p,q)–depend on A if there are no $q+1$ rows having all different entries in b but having at most p different entries in the columns of A. Several results known for the traditional dependencies are generalized.