Alkalmazott Matematikai Lapok **8** (1982) 183—193

A RELÁCIÓS ADATBÁZIS EXTREMÁLIS PROBLÉMÁI

DEMETROVICS JÁNOS és KATONA GYULA

Budapest

A dolgozatban azzal a kérdéssel foglalkozunk, hogy mit lehet mondani egy mátrix funkcionális függéseinek kombinatorikus bonyolultságáról.

Bebizonyítjuk, hogy a minimális kulcsok rendszere *Sperner-rendszer* és hogy minden *Sperner-rendszer* reprezentálható egy mátrixszal, mint minimális kulcsok rendszere. Becsléseket adunk adott *Sperner-rendszert* reprezentáló reláció sorainak a számára, valamint aszimptotikusan pontosan meghatározzuk a „lényeges" funkcionális függések maximális számát.

1. Bevezetés

Egy adatbázis struktúrájának a felhasználó számára legfontosabb jellemzője az, hogy a tárolt adatok közti összefüggések milyen formában vannak jelen az adatbázisban. Mivel egy adatbázis struktúráját a rekordtípusok és a rekordtípusok közti kapcsolatok alkotják, azért az adatmodelleket aszerint osztályozzuk, hogy a rekordtípusok között milyen kapcsolat definiálása megengedett. Első látásra talán meglepőnek hangzik, de ennek alapján két adatmodellt különböztethetünk meg:

— az egyik az ún. *hálózatos adatmodell,* melyben rekordtípusok között tetszőleges, pointerezéssel megadható kapcsolat létesíthető;
— a másik az ún. *relációs adatmodell,* melyben a rekordtípusok között semmilyen kapcsolat nem létesíthető.

A relációs adatbázisokban tehát a rekordtípus az egyetlen kapcsolatteremtő eszköz az adatbázis entitásai között.

Rekordtípus megadása matematikai szempontból nem más, mint egy „értelmezési tartomány" és egy „értékkészlet" definiálása; az így adott rekordtípus előfordulásai az adott értelmezési tartományon értelmezett, értékeit az adott értékkészletből felvevő függvények. Az értelmezési tartomány elemeit attributumoknak nevezik. Egy adatbázis egy rekordtípusának egy adott pillanatban létező előfordulásai tehát egy mátrixot alkotnak. Ennek sorai az egyes előfordulások, oszlopai pedig az egyes attributumokon az előfordulások értékei.

Az adatbázistervezés során az adatbázis struktúrájának megadása utáni lépés azon feltételek előírása, melyeket a leendő adatbázisnak minden pillanatban teljesítenie kell. Például, ha egy prímszámokat kereső program eredményeit kívánjuk egy adatbázisban tárolni, akkor a „minden érték prímszám" egy ilyen feltétel.

Az előírható feltételek formája természetesen függ az adatbázis struktúrájától. Ezért az adatmodell kutatásának egyik fontos területe megtalálni az adott adatmodellben azokat a „feltételfféléket", amelyek egyrészt kielégítik a gyakorlatban felmerülő igényeket, másrészt bizonyos tulajdonságaik matematikailag elég jól vizsgálhatók ahhoz, hogy hatékony adatbáziskezelő rendszerek legyenek kiépíthetők.

A relációs adatbázisokban az egyik legfontosabb „feltételféle" az E. F. CODD [1] által definiált funkcionális függés. A fentiekben megállapítottuk, hogy egy relációs adatbázis pillanatnyi tartalmát mint mátrixokat képzelhetjük, ezért a funkcionális függést mátrixokra definiáljuk. Ha A és B a mátrix értelmezési tartományának (azaz az attributumok halmazának) részhalmazai és $\mathbf{M}$ a mátrix, akkor azt mondjuk, hogy B funkcionálisan függ A-tól az $\mathbf{M}$ mátrixban, ha $\mathbf{M}$ bármely két olyan sora, amelyek A minden elemén egyenlő értékeket vesznek fel, B minden elemén is megegyeznek.

A funkcionális függéssel kapcsolatos problémák két csoportra oszthatók. Az egyik problémakör gyakorlati motivációja a következő: relációs adatbázis tervezése során a felhasználó által megadott feltételek a leendő adatbázisra általában nem teljesek; azaz bizonyos, meg nem fogalmazott újabb feltételek következnek belőlük. Ezek ismerete gyakran fontos a felhasználó számára, tehát szükséges olyan algoritmusok kidolgozása, melyekkel adott feltételrendszerből következő feltételek megtalálhatók. Ennek a problémának matematikai logikai és algoritmuselméleti vonatkozásai vannak.

Ebben a cikkben a másik, nem kevésbé fontos problémakörrel foglalkozunk, amelynek alapkérdése röviden így fogalmazható:

mit mondhatunk egy mátrix funkcionális függéseinek kombinatorikus bonyolultságáról.

A 2. pontban az ún. minimális kulcsok rendszerének struktúráját vizsgáljuk. Bebizonyítjuk, hogy a minimális kulcsok rendszere *Sperner-rendszer* és hogy minden *Sperner-rendszer* reprezentálható egy mátrixszal, mint minimális kulcsok rendszere.

A 3. pontban adott *Sperner-rendszert* reprezentáló reláció sorainak számára adunk becsléseket.

A 4. pontban aszimptotikusan pontosan meghatározzuk a „lényeges" funkcionális függések maximális számát.

2. A minimális kulcsok maximális száma

Legyen $\mathbf{M}$ egy nem negatív egész értékeket tartalmazó, különböző sorokból álló $m\times n$-es mátrix. $\mathbf{M}$ oszlopainak egy K részhalmazát *kulcsnak* nevezzük, ha az ezen oszlopok által meghatározott almátrix sorai különbözők. Ez a feltétel azt jelenti, hogy egy sort már az ezen oszlopokban álló értékei meghatározzák. Vagyis, ha egy sort akarunk megtalálni, elég megadni a K-hoz tartozó oszlopokban álló értékeket.

Ha K_1 egy kulcs és $K_1\subset K_2$, akkor K_2 nyilvánvalóan szintén kulcs. Azok tehát az érdekes kulcsok, melyekhez nem található „kisebb" kulcs. Az olyan K kulcsokat, melyekben nem található $K'\subsetneq K$ kulcs, *minimális kulcsoknak* nevezzük. A minimális kulcsok halmazát $\mathscr{K}$-val jelöljük.

A modellben eddig 4 lényeges paraméter van: 1. az oszlopok száma, n; 2. a sorok száma, m; 3. a mátrixban álló legnagyobb egész szám, N; 4. a minimális kulcsok száma, $|\mathscr{K}|$. A legáltalánosabb probléma az lehetne, hogy hármat ismerve a fenti paraméterek közül, milyen határok közé lehet szorítani a negyediket. E fejezetben egy ennél szerényebb cél irányába törekszünk. Két, illetve 3 paraméter közötti egyenlőtlenségeket vizsgálunk.

Az első probléma az n és $|\mathcal{K}|$ közötti kapcsolat felderítése. A következő tétel ezt a problémát egy halmazrendszerekre vonatkozó ismert tételre redukálja.

2.1. TÉTEL [3]: *Egy n-elemű halmaz részhalmazainak egy $\mathcal{K}$ rendszeréhez akkor és csak akkor található egy $n\times m$-es mátrix, melyben a minimális kulcsok rendszere éppen $\mathcal{K}$, ha $\mathcal{K}$ elemeire a*

$$K_1, K_2 \in \mathcal{K}, \quad K_1 \neq K_2 \Rightarrow K_1 \not\subset K_2 \tag{2.1}$$

feltétel teljesül.

Bizonyítás. Vezessük be a következő jelölést: $\mathcal{K}^{-1}=\{B: B$ nem tartalmaz $\mathcal{K}$-beli részhalmazt és maximális az ilyen tulajdonságú halmazok között$\}$. Legyen $\mathcal{K}^{-1}=\{B_1, B_2, \ldots, B_m\}$. A megkonstruálandó $\mathbf{M}$ mátrixnak $m+1$ sora lesz. A 0-adik sor legyen a csupa 0-ból álló sor; az i-edik sor $(1\leqq i\leqq m)$ j-edik eleme legyen 0 vagy i, aszerint, hogy $j\in B_i$ vagy sem. Világos, hogy ha $A\subseteqq\{1, 2, \ldots, n\}$ olyan halmaz, amely nem tartalmaz $\mathcal{K}$-beli részhalmazt, akkor van olyan i, hogy $A\subseteqq B_i$ és így A-n megegyezik a 0-adik és i-edik sor, tehát A nem kulcs. Ha viszont $K\in\mathcal{K}$, akkor minden i-re $(1\leqq i\leqq m)$ $K\setminus B_i\neq\emptyset$, és így az i-edik sor egy K-beli oszlopában i áll, emiatt K kulcs. Tehát $\mathbf{M}$ minimális kulcsainak a rendszere éppen $\mathcal{K}$. A bizonyítást befejeztük.

Nézzünk egy illusztratív példát. Legyen $n=5$, $\mathcal{K}=\{\{1, 2, 3\}, \{1, 2, 4\}, \{3, 4\}, \{4, 5\}\}$. Ekkor $\mathcal{K}^{-1}=\{\{1, 4\}, \{2, 4\}, \{1, 3, 5\}, \{2, 3, 5\}, \{1, 2, 5\}\}$ és a megfelelő $\mathbf{M}$ mátrix:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 2 & 0 & 2 & 0 & 2 \\ 0 & 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 & 0 \\ 0 & 0 & 5 & 5 & 0 \end{bmatrix}.$$

Eredeti célunk az n oszlopú mátrix minimális kulcsai száma felső határának megállapítása volt (az alsó határ érdektelen). Az első tétel alapján ez arra a problémára redukálódott, hogy hány egymást nem tartalmazó részhalmaza van egy n elemű halmaznak. E kérdést SPERNER [2] tétele válaszolja meg, mely szerint ez a maximum $\binom{n}{\left\lfloor \frac{n}{2} \right\rfloor}$ ($\lfloor x\rfloor$ az x egész részét jelöli):

2.2. TÉTEL [3]: *Egy n oszloppal rendelkező $\mathbf{M}$ mátrixban legfeljebb $\binom{n}{\left\lfloor \frac{n}{2} \right\rfloor}$ minimális kulcs van.*

3. Adott kulcsokhoz individuumok minimális száma

A 2.1. tétel szerint minden $\mathcal{K}$ *Sperner-rendszerhez* található olyan **M** mátrix, amely minimális kulcsainak rendszere $\mathcal{K}$. Azonban ott nem foglalkoztunk azzal, hogy mennyi a sorok minimális száma. Jelölje tehát $s(\mathcal{K})$ ezt a minimális sorszámot. E fejezetben $s(\mathcal{K})$ értékeivel foglalkozunk különféle $\mathcal{K}$ *Sperner-rendszerek* esetére. A 2.1. tétel bizonyítása ad becslést $s(\mathcal{K})$-ra.

Látható, hogy $\mathcal{K}^{-1}$ *Sperner-rendszer*, azaz a *Sperner-tétel* alapján kimondható a

3.1. TÉTEL.

$$s(\mathcal{K}) \leqq 1+|\mathcal{K}^{-1}| \leqq 1+\binom{n}{\left[\frac{n}{2}\right]}.$$

A következő tétel azt mutatja meg, hogy van olyan $\mathcal{K}$, amelyre $s(\mathcal{K})$ „közel" jut e felső korláthoz.

3.2. TÉTEL [5]: *Minden n természetes szám esetén található olyan $\mathcal{K}$ Sperner-rendszer n elemen, melyre*

$$\frac{1}{n^2}\binom{n}{\left[\frac{n}{2}\right]} \leqq s(\mathcal{K}). \tag{3.1}$$

Bizonyítás. Legyen $s(n)=\max\limits_{\mathcal{K}}\{s(\mathcal{K})\}$, ahol a maximumot az n elemen készíthető *Sperner-rendszerekre* vesszük. Azt mondjuk, hogy az **M** mátrix realizálja $\mathcal{K}$-t, ha **M** minimális kulcsainak rendszere $\mathcal{K}$. A fenti definíciók szerint bármely n elemen felvett $\mathcal{K}$ realizálható egy $m\times n$-es mátrixszal, ahol $m\leqq s(n)$. Ha az $(m+1)$-edik, $(m+2)$-edik, ..., $s(n)$-edik sorba csupa $(m+1)$-est, $(m+2)$-est, ..., $s(n)$-est írunk, a kibővített mátrix is $\mathcal{K}$-t realizálja. Azaz bármely $\mathcal{K}$ realizálható egy $s(n)\times n$-es mátrixszal. (Feltehető, hogy az eredeti mátrixban csak $1, 2, \dots m$ van, lásd alább.)

Tekintsük egy ilyen **M** első oszlopában álló elemeket nagyság szerint rendezve: $i_1<i_2<\dots<i_r$ $(1\leqq r\leqq s(n))$, és cseréljük ki i_j-t j-re. Könnyen látható, hogy ez a változtatás nem változtatja meg a minimális kulcsok rendszerét. Ugyanezt a változtatást elvégezve az összes oszlopban, egy olyan **M**-hez jutunk, amelyben minden elem $1, \dots, s(n)$. Az ilyen mátrixok száma $s(n)^{ns(n)}$. Így

$$s(n)^{ns(n)} \geqq \mathcal{K}\text{—k száma.} \tag{3.2}$$

Mivel valahány $\left[\frac{n}{2}\right]$-elemű halmazból álló rendszer mindig *Sperner-rendszer*, (3.2) jobb oldalát csökkentve a

$$s(n)^{ns(n)} > 2^{\binom{n}{\left[\frac{n}{2}\right]}} \tag{3.3}$$

egyenlőtlenséghez jutunk. (3.3)-at átalakítva a

(3.4) $$s(n)\log_2 s(n) > \frac{1}{n}\binom{n}{\left[\frac{n}{2}\right]}$$

egyenlőtlenséghez jutunk. Ha (3.4)-ben $s(n)$ helyére $\frac{1}{n^2}\binom{n}{\left[\frac{n}{2}\right]}$-őt írunk, akkor (3.4) jobb oldalánál kisebbet kapunk, azaz (3.4)-ből következik $s(n) > \frac{1}{n^2}\binom{n}{\left[\frac{n}{2}\right]}$, ami éppen a bizonyítandó állítás.

A következőkben néhány egyszerű $\mathscr{K}$ *Sperner-rendszerre* próbáljuk $s(\mathscr{K})$-t meghatározni. Jelölje $\mathscr{F}_k^n$ az n elemű halmaz összes k-elemű halmazainak rendszerét. $\mathscr{F}_k^n$ nyilvánvalóan *Sperner-rendszer*. Legelőször következzék egy $s(\mathscr{F}_k^n)$-ra vonatkozó egyszerű lemma.

3.1. LEMMA [6].
$$\binom{s(\mathscr{F}_k^n)}{2} \geqq \binom{n}{k-1}, \quad 0 < k \leqq n.$$

Bizonyítás. Tegyük fel, hogy egy $m\times n$-es **M** mátrix realizálja $\mathscr{F}_k^n$-et. Bármely $k-1$ különböző oszlophoz található egy sor-pár, amelyek ezen oszlopokban megegyeznek. Másrészt $k-1$ oszlop egy másik rendszeréhez másik pár kell, hogy tartozzék. Így $\binom{m}{2} \geqq \binom{n}{k-1}$, a bizonyítást befejeztük.

E lemmából $s(\mathscr{F}_1^n) \geqq 2$ következik. A következő konstrukció bizonyítja, hogy $s(\mathscr{F}_1^n) = 2$:

$$\begin{bmatrix} 0 & 0 \ldots 0 \\ 1 & 1 \ldots 1 \end{bmatrix}.$$

Határozzuk meg most $s(\mathscr{F}_2^n)$-et. Ha $m\times n$-es mátrix realizálja $\mathscr{F}_2^n$-et, akkor a 3.1. lemma szerint

(3.5) $$n \leqq \binom{m}{2}.$$

Másrészt, ha (3.5) fennáll, az alábbiakban megmutatjuk, hogy lehet egy $m\times n$-es mátrixot konstruálni, mely $\mathscr{F}_2^n$-et realizálja. **M** minden oszlopa két nullát tartalmaz, különböző oszlopokban a nulla-pár különböző helyeken áll. Az i-edik sor többi eleme legyen i. A konstrukció megvalósíthatósága (3.5)-ön múlik. Könnyen látható, hogy 1 oszlopban van két egyforma sor, míg 2 oszlopban már nincs. Így $s(\mathscr{F}_2^n)$ a (3.5)-öt m-ben kielégítő legkisebb egész szám.

Alkalmazzuk most a lemmát $k=n-1$-re:

$$s(\mathscr{F}_{n-1}^n) \geqq n.$$

A következő konstrukció biztosítja az egyenlőséget:

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & \vdots & & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$s(\mathcal{F}_n{}^n)$ meghatározása egy kicsit bonyolultabb. Egy újabb lemma szükséges. Ha $\mathbf{M}$ egy $m\times n$-es mátrix, $G(\mathbf{M})$ legyen az a gráf, melynek szögpontjai $\mathbf{M}$ sorai, két szögpont össze van kötve, ha a két sor legalább egy helyen megegyezik. Az élre ráírjuk a közös helyek A halmazát.

3.2. Lemma [6]: *Legyen* $\mathbf{M}$ *egy mátrix és legyenek* $G(\mathbf{M})$ *egy körére az* $A_1, \dots$ $\dots, A_r$ *halmazok írva. Ekkor*

$$\text{(3.6)} \qquad \textit{minden } 1 \leqq j \leqq r\textit{-re} \bigcap_{\substack{i=1 \\ i \neq j}}^{r} A_i - A_j = \emptyset.$$

Bizonyítás. Tegyük fel indirekt, hogy (3.6) nem üres, azaz létezik egy oszlop, mondjuk az u-adik, amely minden A_i-ben benne van, kivéve A_j-t. Legyenek a kör mentén levő szögpontok $k_1, \dots, k_r$ úgy, hogy (k_i, k_{i+1})-re van A_i írva $(1 \leqq i < r)$, és (k_r, k_1)-re A_r. $u \in A_{j+1}$-ből következik, hogy a k_{j+1}-edik és k_{j+2}-edik értékek az u-adik sorban egyenlők. Ugyanez mondható el a k_{j+2}-edik és k_{j+3}-adik értékről is stb. Tehát a k_{j+1}-edik, k_{j+2}-edik, ..., k_r-edik, k_1-edik, ..., k_j-edik értékek az u-adik oszlopban megegyeznek. Ebből $u \in A_j$ következik, ellentmondásban a feltevéssel. A bizonyítás teljes.

Most már meg tudjuk határozni $s(\mathcal{F}_n{}^n)$-et. Tegyük fel, hogy egy $m\times n$-es $\mathbf{M}$ adatbázis realizálja $\mathcal{F}_n{}^n$-et. Mivel egy $n-1$ oszlopból álló halmaz nem kulcs, szerepelnie kell $G(\mathbf{M})$ valamelyik élére írva. Tehát $G(\mathbf{M})$-nek van n különböző éle, melyekre a különböző $(n-1)$-elemű halmazok vannak írva. Ezek az élek a 3.2. lemma miatt nem alkothatnak kört, mert (3.6)-ot nem elégíti ki $(n-1)$-elemű halmazok együttese. Ezek az élek tehát fát alkotnak, következésképpen $G(\mathbf{M})$-nek legalább $n+1$ szögpontja van: $m \geqq n+1$. A következő $(n+1)\times n$-es mátrix realizálja $s(\mathcal{F}_n{}^n)$-et:

$$\begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & \vdots & & \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Az eddigi eredményeket a következő tételben foglaljuk össze:

3.3. Tétel [6]:

$$s(\mathcal{F}_1^n) = 2, \quad s(\mathcal{F}_2^n) = \left\lceil \frac{1+\sqrt{1+8n}}{2} \right\rceil,$$

$$s(\mathcal{F}_n^n) = n+1, \quad s(\mathcal{F}_{n-1}^n) = n.$$

A 3.1. lemmából $s(\mathcal{F}_3{}^n) \geqq n$ is következik.

Sejtés. $s(\mathcal{F}_3^n)=n$, ha $n\geqq 7$.
Bizonyos esetekben ez a sejtés bizonyítást nyert.

3.4. TÉTEL (FÜREDI):

$$s(\mathcal{F}_3^n)=n, \quad ha \quad n=4^r.$$

Bizonyítás. Konstruálni fogunk egy $4^r\times 4^r$-es mátrixot, mely $\mathcal{F}_3^n$-et realizálja. A módszer a *Sperner-rendszerek* irodalmának egy bevált módszere.

Tekintsük a $2r$ dimenziós 0-kból és 1-ekből álló 4^r vektor V halmazát. Be fogjuk látni — teljes indukcióval —, hogy ez a halmaz a csupa-nulla vektort leszámítva, beosztható 3-elemű $A_1, \dots, A_{(4^r-1)/3}$ részhalmazokba úgy, hogy az egy részhalmazba eső három vektorra, v_1, v_2, v_3-ra $v_1+v_2=v_3$ teljesül a koordinátánkénti mod 2 összeadásra nézve (ekkor $v_1+v_3=v_2$, $v_2+v_3=v_1$ már következik). $r=1$-re az állítás igaz: $(0,1)+(1,0)=(1,1)$. Tegyük fel, hogy r-re már igaz, és bizonyítsuk $r+1$-re. Az r-re készített minden $A=\{v_1, v_2, v_3\}$ halmazból 4 újat fogunk csinálni:

1. v_1, v_2, v_3 mindegyike végére 00-t írunk;
2. v_i, v_{i+1}, v_{i+2} után 01-et, 10-át és 11-et írunk $i=1, 2, 3$-ra. $(v_4=v_1, v_5=v_2)$.

Ezenkívül a $2r$ hosszúságú csupa-nulla vektorból is csinálunk egy 3-elemű halmazt 01, 10 és 11 utánaírásával. Könnyen látható, hogy így minden $2(r+1)$-dimenziós 0, 1 vektort pontosan egyszer kaptuk meg, azaz ez a kívánt felbontást adja.

Igazoltuk tehát, hogy létezik a következő felbontás:

$$(3.7) \quad \begin{aligned} &V = A_0\cup A_1\cup\dots\cup A_{(4^r-1)/3}, \quad A_i\cap A_j=\emptyset \quad (0\leqq i<j\leqq (4^r-1)/3),\\ &A_0=\{(0,\dots,0)\},\\ &|A_i|=3 \quad (1\leqq i\leqq (4^r-1)/3);\end{aligned}$$

A_i bármely két elemének összege a harmadikat adja. $(1\leqq i\leqq (4^r-1)/3)$.

Legyen $v\in V$. $A+v$ jelentse az $a+v$ (koordinátánként mod 2) vektorok halmazát, ahol $a\in A$. Az $\mathbf{M}$ mátrix sorait és oszlopait egyaránt V elemeivel számozzuk meg (mondhatjuk ezt úgy is, hogy azokkal az egészekkel számozzuk meg, amelyeknek a fenti vektorok a kettes számrendszerbeli alakjaik). $\mathbf{M}$ v_1-edik sorának v_2-edik eleme legyen a

$$(3.8) \qquad v_1+v_2\in A_i$$

reláció által meghatározott i szám. (3.7) alapján ilyen i létezik és egyértelműen meg van határozva. Így például a v_1-edik sor v_1-edik eleme 0. (v_1 a $\mathbf{0}$-vektor.)

$\mathbf{M}$-ről először azt kell bizonyítani, hogy benne semmilyen 2 oszlop sem kulcs, azaz a tetszőlegesen választott v_2 és $v_2'\neq v_2$ oszlopokhoz találhatók v_1 és $v_1'\neq v_1$ sorok, melyek a v_2 és v_2' oszlopokban megegyeznek. (3.7) miatt valamilyen $i\neq 0$ egészre $v_2+v_2'\in A_i$ teljesül. Legyen A_i egy további eleme a, továbbá legyen

$$v_1=a+v_2 \quad \text{és} \quad v_1'=a+v_2'.$$

Innen

$$v_1+v_2=a, \quad v_1'+v_2'=a,$$

$$v_1+v_2'=a+v_2+v_2', \quad v_1'+v_2=a+v_2+v_2'.$$

Mivel a és $a+v_2+v_2' \in A_i$, így a (3.8) definíció alapján a v_1, v_1' sorok és v_2, v_2' oszlopok metszéspontjaiban csupa i áll **M**-ben. A v_2 és v_2' oszlopok valóban nem alkotnak kulcsot.

Vegyünk most 3 különböző oszlopot, v_2-t, v_2'-t és v_2''-t. Tegyük fel indirekte, hogy valamilyen v_1 és $v_1' \neq v_1$ sorok ezekben az oszlopokban megegyeznek, azaz a (3.8) definíció alapján

$$v_1+v_2 \in A_i, \quad v_1'+v_2 \in A_i, \tag{3.9}$$

$$v_1+v_2' \in A_j, \quad v_1'+v_2' \in A_j, \tag{3.10}$$

$$v_1+v_2'' \in A_k, \quad v_1'+v_2'' \in A_k \tag{3.11}$$

teljesül valamilyen $i, j, k\ (\neq 0)$-ra. (3.9) miatt $v_1+v_2+v_1'+v_2=v_1+v_1' \in A_i$. (3.10)-ből és (3.11)-ből hasonlóan következik $v_1+v_1' \in A_j$ és $v_1+v_1' \in A_k$. Azaz $A_i=A_j=A_k$, vagyis $i=j=k$. Mivel v_2, v_2' és v_2'' különbözők, így $a=v_1+v_2$, $b=v_1+v_2'$ és $c=v_1+v_2''$ is különböző elemei A_i-nek. Hasonlóan következik, hogy $v_1'+v_2$, $v_1'+v_2'$ és $v_1'+v_2''$ is különbözők. Mivel $v_1'+v_2 \neq v_1+v_2$, így vagy $v_1'+v_2=b$ vagy $v_1'+v_2=c$. Az első esetben $v_1'+v_2'$ nem lehet a, mert akkor $v_1'+v_2''$ csak c lehetne, ami a $v_1+v_2''=v_1'+v_2''$ ellentmondásra vezetne. Így $v_1'+v_2'=c$. Tehát ez esetben $v_1'+v_2=b$, $v_1'+v_2'=c$, $v_1'+v_2''=a$. A végső ellentmondás a

$$c = a+b = (v_1+v_2)+(v_1'+v_2) = v_1+v_1',$$

$$a = b+c = (v_1+v_2')+(v_1'+v_2') = v_1+v_1'$$

egyenlőségből következik. A második esetben a fentihez hasonlóan nyerhető az ellentmondás. Tehát nem lehetséges, hogy eredeti feltevésünk teljesüljön, hogy 3 oszlop nem alkot kulcsot. A bizonyítást befejeztük.

FÜREDInek sikerült a sejtést az $n=12k+1$ és $n=12k+4$ alakú számokra is bizonyítani, HANANI [7] eredményeit felhasználva.

4. A függőségek számáról

Egy újabb fogalmat kell még bevezetnünk. Az oszlopok egyik halmazának egy másik halmazától való függősége fogalmát. Heurisztikus megfogalmazásban akkor mondjuk, hogy A-tól függ B (A és B egy **M** mátrix oszlopainak halmazai), ha **M** bármely sorának A-ba eső elemei egyértelműen meghatározzák a B-be eső elemeket. Pontosabban tehát **M** oszlophalmazainak egy (A, B) párját *függőség*nek nevezzük, ha **M** bármely két sora, amelyek megegyeznek az A-nak megfelelő helyeken, megegyeznek a B-nek megfelelő helyeken is. Az $A \to B$ jelölést használjuk annak jelölésére, hogy (A, B) egy függőség. A definíciókból látható, hogy az $A \to$ (**M** összes oszlopa) feltétel pontosan azt jelenti, hogy A kulcs.

Mint a kulcsok esetében, itt is csak bizonyos „meghatározó" vagy „lényeges" függőségeket fogunk figyelembe venni. Könnyen látható, hogy

$$A \to A. \tag{4.1}$$

$$\text{Ha } A \to B \text{ és } A \subset A', \text{ akkor } A' \to B. \tag{4.2}$$

$$\text{Ha } A \to B \text{ és } B' \subset B, \text{ akkor } A \to B'. \tag{4.3}$$

Az ezek alapján „felesleges" függőségektől eltekintve *lényeges függőségnek* nevezzük az $A \to B$ függőséget, ha

(4.4) $$A \neq B,$$

(4.5) nincs $A' \subset A$, $A' \neq A$ úgy, hogy $A' \to B$,

(4.6) nincs $B' \supset B$, $B' \neq B$ úgy, hogy $A \to B'$.

Mivel a lényeges függőségekből levezethetők az összes további függőségek (4.1)—(4.3) alapján, a lényeges függőségek számát tekinthetjük a mátrix egy bonyolultságának. A továbbiakban az n oszloppal rendelkező mátrixok ilyen értelemben vett bonyolultságának maximumával foglalkozunk. Legyen $N(n)$ tehát az n oszlopú mátrixok lényeges függőségeinek maximális száma. Könnyű konstruálni egy olyan mátrixot, amelyben a lényeges függőségek mind $A \to A \cup \{x\}$ alakúak, ahol x az utolsó oszlop, és $x \notin A$. Ebben a lényeges függőségek száma 2^{n-1}. Azonban ennél több is lehet, mint látni fogjuk.

4.1. TÉTEL [4]:

(4.7) $$2^n\left(1-\frac{4}{\log e}\cdot\frac{\log\log n}{\log n}(1+o(1))\right) \leqq N(n) \leqq 2^n\left(1-\frac{1}{n+1}\right),$$

ahol a $\log$ *2-alapú.*

Bizonyítás. Alsó becslés. Legyenek $q_1, \ldots, q_k$ pozitív egészek, $\sum_{i=1}^{k} q_i = n$. A $\left(\binom{q_i}{2}+1\right) \times q_i$ méretű $\mathbf{Q}_i$ mátrixot a következőképpen definiáljuk. $\mathbf{Q}_i$ első sora 1-esekből áll. A többi sor q_i-2 1-est tartalmaz az összes lehetséges módon. Az i-edik sor maradék két helyére i-t írunk $\left(2 \leqq i \leqq \binom{q_i}{2}+1\right)$. Az $\mathbf{M} = \mathbf{Q}_1 \times \ldots \times \mathbf{Q}_k$ mátrix sorait a $\mathbf{Q}_i$ mátrixok sorainak összes lehetséges egymásután való írásával képezzük. Azaz $\mathbf{M}$ egy sorát úgy nyerjük, hogy $\mathbf{Q}_1$ egy tetszőleges sora után $\mathbf{Q}_2$ egy tetszőleges sorát, majd $\mathbf{Q}_3$ egy tetszőleges sorát írjuk stb. Világos, hogy $\mathbf{M}$ oszlopainak száma $\sum_{i=1}^{k} q_i = n$. Az $n=5$, $k=2$, $q_1=3$, $q_2=2$ esetre lássuk példaként a konstrukciót:

$$\left[\begin{array}{ccc:cc}
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 2 & 2 \\
1 & 2 & 2 & 1 & 1 \\
1 & 2 & 2 & 2 & 2 \\
3 & 1 & 3 & 1 & 1 \\
3 & 1 & 3 & 2 & 2 \\
4 & 4 & 1 & 1 & 1 \\
4 & 4 & 1 & 2 & 2
\end{array}\right].$$

Jelölje Q_i^* azon oszlopok indexeit $\mathbf{M}$-ben, amelyek $\mathbf{Q}_i$-nak felelnek meg, azaz $Q_i^* = \{q_1+\ldots+q_{i-1}+1, \ldots, q_1+\ldots+q_i\}$. Könnyen látható, hogy a lényeges $A \to B$

függőségek azok, melyekre

$$\min(q_i - |A \cap Q_i^*|) = 1,$$

$$B = A \cup \bigcup Q_i^*,$$

ahol az $\cup$ azon i-ken fut végig, melyekre $|A \cap Q_i^*| = q_i - 1$ teljesül. Az ilyen $A \to B$ párok száma

$$\prod_{i=1}^{k} (2^{q_i} - 1) - \prod_{i=1}^{k} (2^{q_i} - q_i - 1). \tag{4.8}$$

Ez $N(n)$-re egy alsó becslést ad. Például a fenti példa esetében $17 \leqq N(5)$, ami jobb mint a triviális $2^4 = 16$ becslés. Nagy n-ekre a

$$q = q(n) = \left[\log n - \log\left(\frac{1}{\log e}(\log\log n - \log\log\log n - \log\log e - 1)\right)\right]$$

értéket válasszuk. $k(n)$-et, $r(n)$-et az $n = q(n)k(n) + r(n)$, $0 \leqq r(n) < q(n)$ maradékos osztással definiáljuk. Végül legyen

$$q_1 = q_2 = \ldots = q_r = q+1,$$

$$q_{r+1} = \ldots = q_k = q.$$

Ezekkel az értékekkel (4.8) éppen a (4.7) bal oldalát adja némi elemi, de bonyolult számolás után (lásd [4]).

Felső becslés. Legyen $\mathscr{F}$ azon A halmazok családja, melyekhez tartozik egy B halmaz, melyre $A \to B$ lényeges függőség. Ha $A \to B$ egy lényeges függőség, akkor

$$A \subset B, \quad \text{de} \quad A \neq B \quad ((4.4) \text{ és } (4.6) \text{ alapján}).$$

Vegyünk egy tetszőleges, de az $A \subset C \subset B$, $|C| = |A| + 1$ feltételeket kielégítő halmazt. Könnyen látható, hogy $C \notin \mathscr{F}$. Ellenkező esetben ugyanis $C \to D$ lényeges függőség volna. Ez minden esetben ellentmondásra vezet. Mivel egy ilyen C $\mathscr{F}$-nek legfeljebb n tagjából keletkezhet, van legalább $|\mathscr{F}|/n$ halmaz, ami $\mathscr{F}$-nek nem eleme. Innen az $|\mathscr{F}| + \frac{|\mathscr{F}|}{n} \leqq 2^n$ egyenlőtlenség következik. Ez ekvivalens (4.7) jobb oldalával. A bizonyítást befejeztük.

(4.7) jobb oldalán az $\frac{1}{n+1}$ tag nagyságrendileg $2/n$-re könnyen javítható, hisz $\mathscr{F}$ elemeinek túlnyomó többsége $n/2$ körüli elemmel bír, így a megfelelő C halmazok elemszáma is közelítőleg ennyi. Azonban a felső becslés valószínűleg ennél sokkal jobban is javítható. (4.7)-ben a második tagok között még igen nagy a különbség.

IRODALOM

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

[2] SPERNER, E., „Ein Satz über Untermengen einer endlichen Menge", *Math. Z.* **27** (1928) 544—548.

[3] DEMETROVICS, J., "On the equivalence of candidate keys with Sperner systems", *Acta Cybernetica* **4** (1979) 247—252.
[4] BÉKÉSSY, J., DEMETROVICS, J., HANNÁK, L., FRANKL, P. and KATONA, G. O. H., "On the number of maximal dependencies in a data base relation of fixed order", *Disc. Math.* **30** (1980) 83—88.
[5] DEMETROVICS, J. and GYEPESI, GY., "On the functional dependency and some generalization of it", *Acta Cybernetica* 5 (1981) 296—305.
[6] DEMETROVICS, J. and KATONA, G. O. H., "Extremal combinatorial problems in relational data base", *Fundamentals of Computation Theory, Lecture Notes in Computer Science* 117, (1981) 110—119.
[7] HANANI, H., "The existence and construction of balanced inclomplete block designs", *Ann. Math. Statist.* **32** (1961) 361—386.

(Beérkezett: 1981. november 26.)

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

KATONA GYULA
MTA MATEMATIKAI KUTATÓ INTÉZET
1053 BUDAPEST, REÁLTANODA U. 13—15.

EXTREMAL PROBLEMS OF THE RELATIONAL DATA BASES

J. DEMETROVICS and GY. KATONA

In our paper we investigate the problem how complicated can be combinatorically the functional dependencies of a matrix.

We prove that the system of minimal keys is a *Sperner-system* and every *Sperner-system* can be represented by a matrix as a system of minimal keys. We give estimations on the number of rows of the relation which represents a given *Sperner-system* and determine the asymptotically exact maximal number of the "essential" functional dependencies.

82-1100 — Szegedi Nyomda — Felelős vezető: Dobó József igazgató