# tanulmányok

109/1980

MTA Számítástechnikai és Automatizálási Kutató Intézet Budapest

MAGYAR TUDOMÁNYOS AKADÉMIA
SZÁMITÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZETE

# RELACIÓS ADATBÁZIS LOGIKAI SZINTÜ VIZSGÁLATA FUNKCIONÁLIS FÜGGŐSÉGEK SZEMPONTJÁBÓL

Irta:

*Békéssy András*
*Demetrovics János*
*Gyepesi György*

Tanulmányok 109/1980.

A kiadásért felelős:

DR VÁMOS TIBOR

ISBN 963 311 106 4
ISSN 0324-2951

Készült a
KSH Reprográfiai Üzemében 80/164

Tartalom

# I. Fejezet

## Bevezetés

### A funkcionális függőségek alaptulajdonságai

A relációs adatmodell adatbázisok logikai szerkezetének leirására alkalmas. A relációs adatbázis egy egységét, a Codd-féle relációt egy mátrix alakjában képzeljük el. A mátrix oszlopai az un. attributumokkal indexeltek: az attributumok az adatok "fajtái"; ha pl. emberek korát és nevét jegyezzük fel egy relációban, akkor itt attributum a kor és a név. A mátrix egy-egy sora egy-egy adategységen, az un. objektumon mért /vagy arra vonatkozó, feljegyzett, stb./ attributumok értékének "vektora". Az előbbi példánál maradva, (45, XX.YY) sora a reláció mátrixának akkor és csak akkor, ha létezik XX.YY nevü 45 éves ember abban az adathalmazban, amelyre az adatbázis vonatkozik.

Nem engedjük meg, hogy a relációban két sor azonos legyen. Ha például az adathalmazban két 45 éves XX.YY nevü ember volna, akkor megkülönböztethetőeknek kell lenniük, és azt az attributumot, amely megkülönbözteti őket, harmadik oszlopként be kell vennünk a relációba.

Minden egyes $a$ attributumhoz tartozik egy $T_a$

halmaz, azon dolgok halmaza, amelyek az $a$ attributum "értékei" lehetnek; pl.

$$T_{kor} = \{ 0 \text{ év}, 1 \text{ év}, \ldots, 150 \text{ év} \} ,$$

$T_{név}$ az emberek nevének halmaza: tulajdonképpen betűcsoportoknak tetszőleges halmaza. A $T_a$ halmazt az $a$ attributum tartományának vagy domainjének nevezzük. A reláció a domainek szorzatterének egy részhalmaza. A pontos definició a következő:

1.1 Definició Legyen $\Omega$ véges halmaz, és $a \in \Omega$ esetén $T_a$ egy nem üres halmaz; ekkor egy $R \subseteq \underset{a \in \Omega}{X} T_a$ halmazt $\Omega$ feletti relációnak nevezünk; $\Omega$ az attributumok halmaza.

Ha $R$ reláció $\Omega$ felett és $h \in R$ akkor $h$ egy függvény:

$$h : \Omega \to \bigcup_{a \in \Omega} T_a \quad ,$$

és világos, hogy

$$(\forall a)(a \in \Omega \Rightarrow h(a) \in T_a)$$

Ezekután kifejtjük, hogy mit értünk egy relációs adatmodell logikai strukturáján.

<u>1.2 Definició</u> Legyen $\Omega$ egy attributumhalmaz és R egy $\Omega$ feletti reláció. Ha $A \subseteq \Omega$, $B \subseteq \Omega$ és

$$(\forall h \in R)(\forall g \in R)\Big(\big((\forall a \in A)(h(a) = g(a))\big) \Rightarrow \big((\forall b \in B)(h(b) = g(b))\big)\Big)$$

akkor azt mondjuk, hogy B <u>funkcionálisan</u> függ A-tól az R relációban, jelekben $A \xrightarrow[R]{f} B$.

$A \xrightarrow[R]{f} B$ tehát azt jelenti, hogy az R bármely sorának B-beli attributumokhoz tartozó értékeit "meghatározzák" az A-beli attributumokhoz tartozó értékek.

Vezessük be a következő jelölést: ha X,Y két halmaz és $f$ egy $f: X \to Y$ függvény, továbbá $Z \subseteq X$, akkor $f{\restriction}Z$ jelölje azt a függvényt, amelynek értelmezési tartománya $Z$, és minden $z \in Z$ -re $f{\restriction}Z(z) = f(z)$. Az 1.2 definició szerint $A \xrightarrow[R]{f} B$ akkor és csak akkor, ha létezik egy $\varphi : \mathop{\times}_{a \in A} T_a \to \mathop{\times}_{a \in B} T_a$ leképezés úgy, hogy ha $h \in R$ akkor $h{\restriction}B = \varphi(h{\restriction}A)$

Az $R$ reláció <u>logikai strukturáján</u> az

$$\mathcal{F}_R = \{ (A,B) : A \subseteq \Omega,\ B \subseteq \Omega,\ A \xrightarrow[R]{f} B \}$$

halmazt értjük. Az $\mathcal{F}_R$ -et azért érdemes vizsgálni, mert ha adott egy reláció, melynek ismerjük funkcionális függőségeit, akkor ezeket felhasználhatjuk bizonyos, az adatbázisra vonatkozó információk tömöritésére [5], [6].

Megjegyzések: 1. A logikai struktura általában azoknak a megszoritásoknak összessége, amelyek a relációt a tartományok szorzatterének részhalmazára szükitik le. A jelen keretek között csak olyan megszoritásokkal foglalkozunk, amelyek funkcionális függőségek formájában fogalmazhatók meg, és ezért a logikai struktura fogalmát is csak a funkcionális függőségek keretében definiáltuk, nem pedig a legáltalánosabb formában.

2. Egy adatbázisban egy R relációnak általában mindig csak egy valódi részhalmaza van jelen: a reláció éppen érvényes sorai. Ugy gondoljuk azonban, hogy az adatbázis tartalma időről - időre változhat, a bázisbeli reláció sorokkal gyarapodhat, vagy egyes sorok törlődhetnek. A logikai struktura határozza meg, hogy milyen sorok kerülhetnek egyáltalán az adatbázisba, de ezek általában nincsenek mind jelen. Ennélfogva a bázis egy meghatározott időpontjában egy-egy relációban több funkcionális függés "látszik", mint amennyi a logikai strukturában van. Ezeket "látszólagos" függőségnek nevezzük és vizsgálatukkal általában nem foglalkozunk.

3. Ugy képzeljük, hogy egy-egy reláció bizonyos funkcionális függőségei eleve ismeretesek az adatbázis

tervezője előtt. Ezeket iniciális függőségeknek fogjuk nevezni, ezek képezik a logikai struktura alapjait. Az iniciális függőségek azonban általában más, újabb funkcionális függőségek létezését implikálják, és a lezárási müveletek tulajdonságait tartalmazza az alábbi definició.

1.3 Definició Legyen $\Omega$ az attributumok véges halmaza és

$$\mathfrak{F} \subseteq \{(A,B) : A \subseteq \Omega,\ B \subseteq \Omega\}$$

a funkcionális függőségek egy halmaza. Azt mondjuk, hogy $\mathfrak{F}$ a funkcionális függőségek teljes családja, ha rendelkezik a következő négy tulajdonsággal:

($A$, $B$, $A'$, $B'$ az $\Omega$ tetszőleges részhalmazai )

(F1) $(A,A) \in \mathfrak{F}$;

(F2) ha $(A,B) \in \mathfrak{F}$, $(B,B') \in \mathfrak{F}$, akkor $(A,B') \in \mathfrak{F}$;

(F3) ha $(A,B) \in \mathfrak{F}$ és $(A',B') \leq (A,B)^{*}$, akkor $(A',B') \in \mathfrak{F}$;

(F4) ha $(A,B) \in \mathfrak{F}$ és $(A',B') \in \mathfrak{F}$ akkor $(A \cup A',\ B \cup B') \in \mathfrak{F}$;

[ 1].

---

* $(A',B') \leq (A,B)$ akkor és csak akkor, ha $A' \supseteq A$, $B' \subseteq B$ és $(A',B') \geq (A,B)$ ha $(A,B) \leq (A',B')$.

Könnyen ellenőrizhető, hogy ha R reláció $\Omega$ fölött, akkor $\mathcal{F}_R$ funkcionális függőségek teljes családja.

Forditva is, ha $\mathcal{F}$ funkcionális függőségek teljes családja, akkor van olyan R reláció, hogy $\mathcal{F}_R = \mathcal{F}$. Ez az állitás nem triviális. A tétel Armstrongtól [Arm] származik. Itt egy másik, bizonyitást adunk [11] amely, hasonlóan Armstrong bizonyitásához konstruktiv, azonban számunkra ugy tünik, hogy a konstrukció természetesebb mint Armstrongé. Először azonban foglalkoznunk kell a teljes családok maximális elemeinek fogalmával és tulajdonságaival.

<u>1.4 Definició</u> Ha $\mathcal{F}$ teljes család, akkor jelölje $\hat{\mathcal{F}}$ az $\mathcal{F}$ <u>maximális elemeinek halmazát</u>. A maximális elemek definiciója:

$$\hat{\mathcal{F}} = \{(A,B) \in \mathcal{F} : (A,B) \leq (A',B'),\ (A',B') \in \mathcal{F} \Rightarrow (A,B) = (A',B')\}.$$

<u>1.1 Tétel</u> [1] Legyen $\mathcal{M} \subseteq \{(A,B) \mid A \subseteq \Omega,\ B \subseteq \Omega\}$

Az $\mathcal{M}$ halmaz egy teljes függőségi család maximális elemeinek halmaza akkor és csak akkor ha

(M1) Minden $A \subseteq \Omega$ -hoz létezik $(A',B') \in \mathcal{M}$ úgy hogy $(A,A) \leq (A',B')$

(M2) Ha $(A,B),(A',B')\in \mathcal{M}$ és $(A',B')\leq(A,B)$ akkor $A=A'$ és $B=B'$

(M3) Ha $(A,B),(A',B')\in \mathcal{M}$ és $A'\subseteq B$ akkor $B'\subseteq B$ is.

Bizonyitás 1. Tegyük fel először, hogy $\mathcal{F}$ teljes család és $\mathcal{M}=\hat{\mathcal{F}}$. Igazolni kell az M1-M3 tulajdonságokat.

(M1) következik (F1)-ből.

(M2) nyilvánvaló az 1.4 definicióból.

Az (M3) igazolására tegyük fel, hogy $(A,B),(A',B')\in\hat{\mathcal{F}}$, $A'\subseteq B$. Ekkor $(B,B')\leq(A',B')$ tehát (F3) szerint $(B,B')\in\mathcal{F}$ és igy (F2) szerint, mivel $(A,B)\in\mathcal{F}$, következik $(A,B')\in\mathcal{F}$ is. De akkor (F4) szerint $(A,B\cup B')\in\mathcal{F}$ is. Azonban $(A,B)\leq(A,B\cup B')$ ennélfogva (A,B) maximális volta miatt $B\cup B'=B$ tehát $B'\subseteq B$.

2. Tegyük fel most, hogy $\mathcal{M}$ kielégiti az M1-M3 feltételeket. Tekintsük a funkcionális függőségeknek azt az $\mathcal{F}$ halmazát, amelyet a következőképpen definiálunk:

$$\mathcal{F}=\{(A,B):(\exists(A',B')\in\mathcal{M})((A,B)\leq(A',B'))\}$$

Be kell bizonyitanunk, hogy $\mathcal{F}$ teljes család, melynek maximális elemei az $\mathcal{M}$ halmazbeli függőségek és csak azok.

A,B,C, B' stb. legyenek $\Omega$ részhalmazai.

(F1): (M1) miatt van olyan $(A',B')\in\mathcal{M}$ hogy $(A',B')\geq(A,A)$ tehát $(A,A)\in\mathcal{F}$.

(F2): Tegyük fel, hogy (A,B) és (B,C) elemei az $\mathcal{F}$-nek. Akkor létezik $(A',B')\in\mathcal{M}$ és $(B'',C')\in\mathcal{M}$ úgy, hogy $(A',B')\geq(A,B)$ és $(B'',C')\geq(B,C)$ tehát $B'\supseteq B\supseteq B''$ ; továbbá mivel $B''\subseteq B'$ , azért (M3) szerint $C'\subseteq B'$ is igaz. Látjuk tehát, hogy $A\supseteq A'$ és $C\subseteq C'\subseteq B'$ tehát $(A,C)\leq(A',C')\leq(A',B')$ ezért $(A,C)\in\mathcal{F}$ az $\mathcal{F}$ értelmezése alapján.

(F3): Evidens a $\leq$ reláció tranzitivitása miatt.

(F4): Legyen $(A,B)\in\mathcal{F}$, $(C,D)\in\mathcal{F}$. Akkor $\mathcal{F}$ definiciója szerint létezik $(A',B')\in\mathcal{M}$, $(C',D')\in\mathcal{M}$ úgy, hogy $(A',B')\geq(A,B)$, $(C',D')\geq(C,D)$. (M1) miatt van olyan $(X,Y)\in\mathcal{M}$ hogy $(X,Y)\geq(A'\cup C',A'\cup C')$ és igy $A'\subseteq Y$, $C'\subseteq Y$. De akkor (M3) szerint $B'\subseteq Y$, $D'\subseteq Y$ tehát

$(X,Y)\geq(A'\cup C',B'\cup D')\geq(A\cup C,B\cup D)$ tehát

$(A\cup C,B\cup D)\in\mathcal{F}$ .

$\hat{\mathcal{F}}=\mathcal{M}$ nyilvánvaló. $\square$

<u>Megjegyzés</u> Ha adva van $\mathcal{M}$ akkor a hozzá tartozó teljes család

$$\mathcal{F}=\{(A,B):(\exists(A',B')\in\mathcal{M})((A,B)\leq(A',B'))\}$$

Tekintsük a következő halmazt:

$$\mathcal{B} = \{ B : (\exists A)((A, B) \in \mathcal{M}) \} \quad ,$$

amelyet a "maximális jobboldalak halmazának" fogunk nevezni.

Armstrong érdekes eredménye, hogy a maximális jobboldalak halmaza egyértelmüen karakterizálja $\mathcal{F}$ -et és $\mathcal{M}$ -et. Ez a tény nagyon hasznos, mert lehetővé teszi a teljes családok viszonylag kisebb információval való leirását. A $\mathcal{B}$ halmazra a következő tétel érvényes:

<u>1.2 Tétel</u> Legyen $\mathcal{F}$ funkcionális függőségek teljes családja, $\mathcal{M}$ az $\mathcal{F}$ maximális függőségeinek halmaza $\Omega$ felett, és $\mathcal{B}$ a maximális jobboldalak halmaza. Akkor $\mathcal{B}$ metszetfélháló és $\Omega \in \mathcal{B}$ azaz

(B1) $\Omega \in \mathcal{B}$

(B2) Ha $B_1 \in \mathcal{B}$ , $B_2 \in \mathcal{B}$ akkor $B_1 \cap B_2 \in \mathcal{B}$

Forditva, ha $\mathcal{B}$ kielégiti a B1-B2 feltételeket, akkor létezik a funkcionális függéseknek egy és csakis egy teljes családja $\Omega$ felett, amelynek maximális jobboldalhalmaza éppen $\mathcal{B}$ . [1]

Bizonyitás

(B1) : Teljesül, hiszen (F1) szerint $(\Omega,\Omega)\in\mathcal{F}$ és (M1) szerint ezért létezik $X\subseteq\Omega$ , hogy $(X,\Omega)\in\mathcal{M}$

(B2) : legyen $B_1,B_2\in\mathcal{B}$. Ekkor létezik $X,Y\subseteq\Omega$ úgy, hogy $(X,B_1),(Y,B_2)\in\mathcal{M}$. (F1) szerint $(B_1\cap B_2, B_1\cap B_2)\in\mathcal{F}$ és igy (M1) szerint van $(Z,B')\in\mathcal{M}$ úgy, hogy $(Z,B')\geqslant(B_1\cap B_2, B_1\cap B_2)$, azaz $B'\supseteq$ $\supseteq B_1\cap B_2\supseteq Z$ . Igy persze $Z\subseteq B_1$ , ezért (M3) szerint $B'\subseteq B_1$. Hasonlóan $B'\subseteq B_2$ és tehát $B\subseteq B_1\cap B_2$ és igy $B'=B_1\cap B_2$ .

Forditva, tegyük fel, hogy $\mathcal{B}\subseteq\mathcal{P}(\Omega)$* teljesiti B1 -t és (B2)-t. Legyen $\mathcal{F}=\{(A,B):$ minden $B'\in\mathcal{B}$ -re $A\subseteq B'\Rightarrow B\subseteq B'\}$ . Könnyü ellenőrizni, hogy $\mathcal{F}$ teljes család .

Bebizonyitjuk, hogy $\hat{\mathcal{F}}=\mathcal{M}$ . Legyen $(A,B)\in\hat{\mathcal{F}}$

Legyen $\mathcal{J}=\{B'\in\mathcal{B}: A\subseteq B'\}$ ; ekkor $B=\cap\mathcal{J}$ .

véges volta és B1 , B2 miatt igy $B\in\mathcal{B}$ .

Tegyük fel most, hogy $B\in\mathcal{B}$ . $\mathcal{F}$ teljes család, ezért létezik $(A_1,B_1)$ maximális eleme $\mathcal{F}$ -nek úgy, hogy $A_1\subseteq B\subseteq B_1$ . Ekkor $\mathcal{F}$ definiciója miatt $A_1\subseteq B$ maga után vonja $B_1\subseteq B$ teljesülését; azaz $B=B_1\in\mathcal{B}$.

* $\mathcal{P}(\Omega)$ jelöli az $\Omega$ hatványhalmazát, azaz $\mathcal{P}(\Omega)=\{A: A\subseteq\Omega\}$

Azt kell még megmutatnunk, hogy $\mathcal{F}$ az egyedüli teljes család, amelynek maximális jobboldalhalmaza $\mathcal{B}$ .

Tegyük fel, hogy $\mathcal{F}_1$ teljes család, melynek maximális jobboldalhalmaza $\mathcal{B}$ . Legyen $\mathcal{M}_1$ az $\mathcal{F}_1$ maximális elemeinek halmaza. Először bebizonyitjuk, hogy $\mathcal{F}_1 \subseteq \mathcal{F}$ Minden $B' \in \mathcal{B}$ -re létezik $(A', B') \in \mathcal{M}_1$ és ha $(A, B) \in \mathcal{F}_1$, $A \subseteq B'$ akkor létezik $(A'', B'') \in \mathcal{M}_1$ úgy, hogy $(A'', B'') \geqslant (A, B)$ és $A'' \subseteq B'$. (M3) miatt ekkor $B'' \subseteq B'$ ; igy $B \subseteq B'$. Ezért $(A, B) \in \mathcal{F}$, $\mathcal{F}$ definiciója szerint.

Végül megmutatjuk, hogy $\mathcal{F} \subseteq \mathcal{F}_1$

Ha $(A, B) \in \mathcal{F}$ akkor (M1) szerint $(A', B') \in \mathcal{M}_1$ valamely $A' \subseteq A \subseteq B'$ -re hiszen $(A, A) \in \mathcal{F}_1$ (F1) miatt. $\mathcal{F}$ definiciója miatt $B \subseteq B'$ ; igy $(A', B') \geqslant (A, B)$ és igy $(A, B) \in \mathcal{F}_1$ (F3) alapján. $\square$

<u>Megjegyzés:</u> Ha $\mathcal{B}$ -ra igaz (B1) és (B2) , akkor a $\mathcal{B}$ -hez tartozó teljes család

$$\mathcal{F} = \{(A, B) : (\forall B' \in \mathcal{B})(A \subseteq B' \Rightarrow B \subseteq B')\}$$

<u>1.4 Definició</u> Legyen $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ metszetfélháló, azaz $B_1, B_2 \in \mathcal{B} \Rightarrow B_1 \cap B_2 \in \mathcal{B}$. Az <u>$\mathcal{J} \subseteq \mathcal{B}$ halmazt $\mathcal{B}$ generátorrendszerének nevezzük</u>, ha

$$\mathcal{B} = \{\cap \mathcal{J}' : \mathcal{J}' \subseteq \mathcal{J}\}$$

Ha $\mathcal{J}$ tetszőleges részhalmaza $\Omega$ hatványhalmazának, $\mathcal{P}(\Omega)$-nek, akkor az $\mathcal{J}$ által generált metszetfélháló $\mathcal{B}$, ha $\mathcal{B}$ minden $B$ halmaza metszeteként áll elő $\mathcal{J}$ bizonyos elemeinek, jelekben

$$\mathcal{B} = \{ \cap \mathcal{J}' : \mathcal{J}' \subseteq \mathcal{J} \}$$

Nyilvánvaló, hogy minden $\mathcal{F}$ teljes családot egyértelműen meghatározza maximális jobboldalhalmazának egy generátorrendszere és minden $\mathcal{J} \subseteq \mathcal{P}(\Omega)$-ra létezik egyetlen $\mathcal{F}$ teljes család, amelynek maximális jobboldalainak halmazát $\mathcal{J}$ generálja, és ez az $\mathcal{F} = \{(A,B) : (\forall G \in \mathcal{J})(A \subseteq G \Rightarrow B \subseteq G)\}$

A következő tétel azt mutatja, hogy (F1)-(F4)-gyel axiomatizáltuk a relációk funkcionális függésrendszereit.

**1.3 Tétel [ 1 ]** Legyen $\mathcal{F}$ funkcionális függések teljes családja $\Omega$-n. Ekkor létezik olyan R reláció $\Omega$-n, hogy $\mathcal{F} = \mathcal{F}_R$.

Bizonyitás legyen $\mathcal{M} = \hat{\mathcal{F}}$, $\mathcal{B} = \{ B : (\exists A)((A,B) \in \mathcal{M})\}$, $\mathcal{J}$ generátorrendszere $\mathcal{B}$-nek, $\mathcal{J} = \{G_1, \dots, G_\ell\}$ R relációt konstruálunk* $\Omega$ felett. A reláció sorai, $h_0, \dots, h_\ell$ a következő definicióval adottak:

(1) ha $h_0(a) = 0$ minden $a \in \Omega$-ra

(2) ha $1 \le i \le \ell$ akkor $h_i(a) = \begin{cases} 0, \text{ ha } & a \in G_i \\ i, \text{ ha } & a \notin G_i \end{cases}$

Állitjuk, hogy $\mathcal{F} = \mathcal{F}_R$. $\mathcal{J}$ definiciója miatt

$$(A,B) \in \mathcal{F} \Leftrightarrow (\forall i)(1 \le i \le \ell \Rightarrow (A \subseteq G_i \Rightarrow B \subseteq G_i))$$

Tegyük fel, hogy $(A,B) \in \mathcal{F}$ és $h \ne g \in R$ olyanok, hogy

$$h \restriction A = g \restriction A,\ h = h_i,\ g = h_j,\ i \ne j,\ 0 \le i < j \le \ell.$$

Két eset van:

a/ i=0 R definiciója miatt ekkor $A \subseteq G_j$, igy $(A,B) \in \mathcal{F}$ miatt $B \subseteq G_j$ és ezért $h \restriction B = g \restriction B$

b/ i$>$0 ekkor $A \subseteq G_i \cap G_j$ és igy $B \subseteq G_i \cap G_j$, tehát $g \restriction B = h \restriction B$.

Tehát $(A,B) \in \mathcal{F} \Rightarrow (A,B) \in \mathcal{F}_R$.

Legyenek most $A, B \subseteq \Omega$ olyanaok, hogy $(A,B) \notin \mathcal{F}$ .
Ekkor létezik $i$ úgy, hogy $1 \le i \le \ell$ és $A \subseteq G_i$, $B \nsubseteq G_i$; igy

$$h_0 \restriction A = h_i \restriction A \text{ és } h_0 \restriction B \ne h_i \restriction B$$

Igy $(A,B) \notin \mathcal{F} \Rightarrow (A,B) \notin \mathcal{F}_R$.
Végül is $\mathcal{F} = \mathcal{F}_R$

$\square$

---

* Ez a konstrukció és a bizonyitás új eredmény.

Kulcsok

Ha R egy reláció $\Omega$ -n, akkor különösen fontosak $\mathcal{F}_R$ -nek azon elemei, melyek $(X,\Omega)$ alakúak; hiszen ezek ismerete teszi lehetővé a reláció sorainak kevés oszloppal való karakterizálását.

1.5 Definició Ha $\mathcal{F}$ funkcionális függések teljes családja, akkor $\mathcal{C} = \{A : (A,\Omega) \in \hat{\mathcal{F}}\}$ elemeit $\mathcal{F}$ kulcsainak nevezzük.

(B1) miatt

(C1) $\mathcal{C} \neq \emptyset$

Továbbá nyilvánvaló, hogy ha

(C2) $K_1, K_2 \in \mathcal{C}$ akkor $K_1 \neq K_2 \Rightarrow K_1 \not\subset K_2$ és $K_2 \not\subset K_1$

A kulcsok halmaza tehát Sperner-család $\Omega$ felett.

1.4 Tétel [ 1 ] Ha $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ teljesiti (C1) -t és (C2) -t, akkor létezik $\mathcal{F}$ funkcionális függések olyan teljes családja, amelyre $\mathcal{C}$ az $\mathcal{F}$ kulcsainak halmaza.

Bizonyitás Legyen $\mathcal{B} = \{B : K \in \mathcal{C} \Rightarrow K \notin B\} \cup \{\Omega\}$

ekkor $\mathcal{B}$ -re nyilván igaz (B1) és (B2). Az 1.2 Tétel szerint létezik olyan $\mathcal{F}$ teljes család, amelynek $\mathcal{B}$ a maximális jobboldalhalmaza. Ez az $\mathcal{F}$ igy jellemezhető:

$\mathcal{F} = \{(A, B):$

ha $G \subseteq \Omega$ nem tartalmaz egyetlen $K \in \mathcal{C}$ -t sem

és $A \subseteq G$, akkor $B \subseteq G\}$.

Igy azon maximális $(A, B) \in \mathcal{F}$ elemek, melyekre $B = \Omega$, éppen a $(K, \Omega)$ alakúak, ahol $K \in \mathcal{C}$; azaz $\mathcal{F}$ kulcsainak halmaza $\mathcal{C}$. $\square$

<u>Megjegyzés</u> 1.4 Tétel nem erősithető úgy, hogy minden $\mathcal{C} \subseteq \mathcal{P}(\Omega)$-hoz ami teljesiti (C1) -t és (C2) -t, <u>egyetlen</u> $\mathcal{F}$ létezik, mert ha $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ metszetre zárt és $\Omega \in \mathcal{B}$ olyan, hogy $B \in \mathcal{B} \setminus \{\Omega\}$ és $K \in \mathcal{C} \Rightarrow B \not\supseteq K$, akkor $\mathcal{B}$ -hez létezik pontosan egy $\mathcal{F}$ úgy, hogy $\mathcal{B}$ az $\mathcal{F}$ maximális jobboldalainak rendszere, és igy $\mathcal{C}$ az $\mathcal{F}$ kulcsrendszere. $\mathcal{C}$ -hez ilyen $\mathcal{B}$ azonban több is lehet: pl. ha $\Omega = \{1,2,3\}$; $\mathcal{C} = \{\{1,3\}\}$, akkor $\mathcal{B}_0 = \{\{1,2\},\{2,3\},\{2\},\phi,\{1,2,3\}\}$ és $\mathcal{B}_1 = \mathcal{B}_0 \cup \{\{1\}\}$ is olyan, hogy $\Omega$ -t kivéve, egyik elemük sem tartalmazza $\mathcal{C}$ egyetlen elemét sem, igy ehhez a $\mathcal{C}$ -hez több teljes család is tartozik. Az $\Omega$ feletti Sperner rendszerek tehát, mint kulcsrendszerek, a teljes funkcionális függőségi családokat ekvivalenciaosztályokra bontják /ld. 2. Fejezet/.

## Lineáris relációk

Jelölje $\mathbb{Q}$ a racionális számtestet; $\mathbb{Q}^n$ a racionális számokból álló n hosszúságú sorozatok halmaza. Ekkor $\mathbb{Q}^n$ természetes módon n-dimenziós vektortér felett.

1.6 Definició Egy R reláció $\Omega$ felett lineáris, ha R sorainak halmaza altér $\mathbb{Q}^n$ -ben.

1.7 Definició Legyen R reláció $\Omega$ felett, $R \subseteq \mathbb{Q}^n$, $(A,B) \in \mathcal{F}_R$ ekkor az (A,B) funkcionális függés lineáris R-ben, ha minden $b \in B$-re léteznek $\alpha_{a,b} \in \mathbb{Q} : a \in A$ együtthatók úgy, hogy ha $h \in R$, akkor, $h(b) = \sum\limits_{a \in A} \alpha_{a,b} h(a)$

1.5 Tétel Ha R lineáris reláció $\Omega$ felett és $(A,B) \in \mathcal{F}_R$ akkor (A,B) lineáris függés R-ben. /Azaz lineáris reláció minden függése lineáris./

Bizonyitás feltehető, hogy $|B|=1$, azaz $B=\{b\}$, mert $(A,B) \in \mathcal{F}_R$ akkor és csak akkor, ha $(A,\{b\}) \in \mathcal{F}_R$ minden $b \in B$-re és (A,B) lineáris R-ben akkor és csak akkor, ha $(A, \{b\})$ lineáris R-ben minden $b \in B$-re.

Legyen S az R-nek a $\mathbb{Q}^A$ -ra való vetülete. Nyilvánvaló, hogy S altere $\mathbb{Q}^A$ -nak.
Ha $\underline{x} \in R$, akkor $pr_A(\underline{x}) \in S$ az $\underline{x}$ vetülete.
$(A, \{b\}) \in \mathfrak{F}_R$ miatt $f : S \to \mathbb{Q}$ a következő definicióval függvény: legyen $\underline{x} \in R$ esetén $f(pr_A(\underline{x})) = \underline{x}(b)$ /azaz az $\underline{x}$ vektor b-dik koordinátája/. Állitjuk, hogy $f$ lineáris leképezés.

(1) $\underline{x}$, $\underline{y} \in R$; akkor $\underline{x}+\underline{y} \in R$; igy, $pr_A$ lineáris volta miatt $pr_A(\underline{x}+\underline{y}) = (pr_A(\underline{x}) + pr_A(\underline{y})) \in S$.
$(\underline{x}+\underline{y})(b) = \underline{x}(b) + \underline{y}(b)$; igy
$f(pr_A(\underline{x}) + pr_A(\underline{y})) = f(pr_A(\underline{x}+\underline{y})) = \underline{x}(b) + \underline{y}(b) =$
$= f(pr_A(\underline{x})) + f(pr_A(\underline{y}))$

(2) $\underline{x} \in R$, $\alpha \in \mathbb{Q}$; ekkor $pr_A(\alpha\underline{x}) = \alpha \cdot pr_A(\underline{x}) \in S$ és
$f(\alpha \cdot pr_A(\underline{x})) = f(pr_A(\alpha \cdot \underline{x})) = (\alpha\underline{x})(b) = \alpha \cdot \underline{x}(b) =$
$= \alpha \cdot f(pr_A(\underline{x}))$

S lineáris altér $\mathbb{Q}^A$ -ban, ezért $f$ kiterjeszthető $\mathbb{Q}^A$ -ra egy $\tilde{f}$ lineáris leképezéssé; igy van $\{\alpha_{a,b} : a \in A\} \subseteq \mathbb{Q}$ úgy, hogy minden $\underline{x} \in \mathbb{Q}^A$ -ra $\tilde{f}(\underline{x}) = \sum_{a \in A} \alpha_{a,b}\, \underline{x}(a)$. $\tilde{f}$ kiterjeszti $f$ -t; tehát $f$ a kivánt alakú $\square$

Ha $R \subseteq \mathbb{Q}^n$ lineáris reláció, akkor R-nek vagy egyetlen sora van, vagy végtelen sok. R logikai strukturáját azonban R-nek már véges sok sora is tükrözi; hiszen $\mathcal{P}(\Omega)^2 = \{(A,B) : A \subseteq \Omega, B \subseteq \Omega\}$ véges és ha $(A,B) \in \mathcal{P}(\Omega)^2$ nem funkcionális függés R-ben, akkor ezt R-nek két sora mutatja.

Mivel a lineáris függések gyakorlati szempontból fontosak, sürün előfordulnak, érdemes megvizsgálni, hogy melyek azok a függésrendszerek amelyek realizálódhatnak lineáris relációban.

<u>1.1 lemma</u> Ha $R \subseteq \mathbb{Q}^m$ lineáris reláció $\Omega$ felett, és $K_1, K_2$ kulcsa $\mathcal{F}_R$ -nek, akkor $|K_1| = |K_2|$

<u>Biz.:</u> legyen $R' \subseteq R$ olyan reláció $\Omega$-n, amelynek véges sok sora van és amelyre $\mathcal{F}_{R'} = \mathcal{F}_R$. Az 1.5 Tétel szerint $\mathcal{F}_{R'}$ minden eleme lineáris függés. Könnyen meggondolható, hogy $K \subseteq \Omega$ kulcsa $\mathcal{F}_{R'}$ -nek, akkor és csak akkor, ha a $\{(h(a) : h \in R') : a \in K\}$ vektorhalmaz maximális lineárisan független részhalmaza $\{(h(a) : h \in R') : a \in \Omega\}$ -nak. Igy ha $K \subseteq \Omega$ kulcs, akkor K elemszáma egyenlő az R' mátrixának rangjával $\square$

<u>1.1 Következmény</u> Ha $R \subseteq \mathbb{Q}^m$ lineáris reláció, akkor $\mathcal{F}_R$ minden kulcsának pontosan annyi eleme van, amekkora R dimenziója.

1.1 lemma szerint tehát nem igaz, hogy minden teljes család lineáris reláció függésrendszere lehet.

A lineáris relációk függésrendszereinek axiomatizálása nagyon nehéz feladat, mert egy $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ lineáris reláció függésrendszerének kulcsrendszere akkor és csak akkor, ha $\mathcal{C}$ a $\mathbb{Q}$ felett koordinátázható matroid; és a koordinátázható matroidok belső jellemzése a véges kombinatorika régóta vizsgált, máig megoldatlan problémája.

Ha $|\Omega| = n$, $k \leq n$, $\mathcal{C} = \{A \subseteq \Omega : |A| = k\}$, akkor $\mathcal{C}$ egy lineáris reláció kulcsrendszere.[8]

## Néhány kombinatorikus eredmény

A következőkben a funkconális függések vizsgálatánál felmerülő kombinatorikai jellegü problémákkal foglalkozunk. $|\Omega| = n$, $\mathcal{F}$ funkcionális függések teljes családja $\Omega$ felett, ekkor $\mathcal{F} \subseteq \mathcal{P}(\Omega) \times \mathcal{P}(\Omega)$ ( $\mathcal{P}(\Omega)$ jelöli az $\Omega$ hatványhalmazát, azaz $\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$ ) ; igy $|\mathcal{F}| \leq 2^n \cdot 2^n = 2^{2n} = 4^n$. Ha $X \subseteq Y \subseteq \Omega$ akkor $(Y, X) \in \mathcal{F}$, ezért $|\mathcal{F}| \geqslant \sum_{k=0}^{n} \binom{n}{k} \cdot 2^k = 3^n$. Végül is $3^n \leq |\mathcal{F}| \leq 4^n$.
Relációs adatmodell gyakorlati alkalmazásakor általában az a helyzet, hogy adott funkcionális függéseknek egy $\mathcal{G}$

halmaza / $\mathcal{G}$ nem feltétlenül teljes család/ és szeretnénk tudni, hogy mik azok a további függések, amelyek biztosan teljesülnek. A funkcionális függések körül nyilván a nem triviálisak érdekesek csak.

Mivel minden $(A,B)$ függés B jobboldalát tartalmazza egy $(A',B')$ függés B' jobboldala, azért a nem triviálisak közül is csak a maximálisak megismerése fontos.

Hány maximális nem triviális függése lehet egy teljes családnak? Ha csak kevés, akkor "jó lenne". Azonban sok lehet.

<u>1.7 Definició</u> $\mathfrak{f}:\mathcal{P}(\Omega)\to\mathcal{P}(\Omega)$ lezárás, ha minden $A, B \subseteq \Omega$-ra

a/ $A \subseteq \mathfrak{f}(A)$

b/ $\mathfrak{f}(\mathfrak{f}(A)) = \mathfrak{f}(A)$

c/ $A \subseteq B \Rightarrow \mathfrak{f}(A) \subseteq \mathfrak{f}(B)$

Ha $\mathcal{F}$ teljes család, akkor az

$$\mathfrak{f}_{\mathcal{F}}(A) = \{ b : (A, \{b\}) \in \mathcal{F} \}, \text{ ha } A \subseteq \Omega$$

definició lezárást definiál és ha $\mathfrak{f}:\mathcal{P}(\Omega)\to\mathcal{P}(\Omega)$ lezárás akkor $\mathcal{F}_{\mathfrak{f}} = \{(A,B) : B \subseteq \mathfrak{f}(A)\}$ teljes család és $\mathcal{F}_{\mathfrak{f}_{\mathcal{F}}} = \mathcal{F}$ és $\mathfrak{f} = \mathfrak{f}_{\mathcal{F}_{\mathfrak{f}}}$. Világos továbbá, hogy

$\{ f_{\mathcal{F}}(A) : A \subseteq \Omega \} = \mathcal{B}$ az $\mathcal{F}$ maximális jobboldalainak halmaza, ha $\mathcal{F}$ teljes család.

<u>1.8 Definició</u> Ha $f$ lezárás $\Omega$ -n, akkor az $(A, f(A))$ pár <u>nem triviális</u>, ha $A \neq f(A)$ és <u>maximális</u>, ha

$B \subseteq A,\ f(B) = f(A) \Rightarrow A = B.$

Jelölje $N(f)$ a maximális nem triviális párok számát,

$$N_n = \max \{ N(f);\ f \text{ lezárás} \}$$

Ekkor $N_n \geqslant 2^{n-1}$, mert ha $a \in \Omega$ rögzitett attributum és $f(A) = A \cup \{a\}$ akkor $(A, A \cup \{a\})$ maximális nem triviális, hacsak $a \notin A$; tehát $N(f) = 2^{n-1}$. $N_n < 2^n$ nyilvánvaló; lényeges eredmény viszont, hogy $N_n$ aszimptotikusan egyenlő $2^n$-nel[4].

Tehát ha egy $\mathcal{F}$ teljes családot a maximális elemeit használva kivánjuk leirni, akkor általában $2^{|\Omega|}$ sok függést kell megismernünk.

Ha $\mathcal{C}$ egy teljes család kulcsainak rendszere, akkor $K_1, K_2 \in \mathcal{C},\ K_1 \subseteq K_2 \Rightarrow K_1 = K_2$; azaz $\mathcal{C}$ un. Sperner-rendszer. Sperner 1928-ban bizonyitotta [Sp], hogy ha $|\Omega| = m$ és $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ Sperner-rendszer, akkor $|\mathcal{C}| \leq \binom{m}{[\frac{m}{2}]}$

Láttuk, hogy ha $\mathcal{C}$ Sperner-rendszer, akkor $\mathcal{C}$ több teljes családnak is lehet kulcsrendszere.

Ha $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ metszetfélháló, akkor létezik egyetlen teljes család, aminek $\mathcal{B}$ a maximális jobboldal-rendszere; azaz a maximális jobboldalak már egyértelmüen meghatározzák $\mathcal{F}$-t.

Metszetfélhálót tetszőleges generátorrendszere meghatározza. Ha $\mathcal{B}$ metszetfélháló, akkor van $\mathcal{J} \subseteq \mathcal{B}$ úgy, hogy $\mathcal{B} = \{\cap \mathcal{J}' : \mathcal{J}' \subseteq \mathcal{J}\}$ és $(*)$: ha $G \in \mathcal{J}$ akkor

$$G \neq \cap\{H \in \mathcal{J} : H \neq G, G \subset H\}$$

<u>1.9 Definició</u> $\mathcal{J} \subseteq \mathcal{P}(\Omega)$ metszettulajdonságú, ha $(*)$-t teljesiti

<u>1.6 Tétel</u> Van $c$ $n$-től független konstans úgy, hogy ha $|\Omega| = n$, $\mathcal{J} \subseteq \mathcal{P}(\Omega)$ metszettulajdonságú, akkor

$$|\mathcal{J}| \leq c \cdot \binom{n}{[\frac{n}{2}]}$$

<u>Bizonyitás:</u> Legyen

$$S = \left\{A \subseteq \Omega : \frac{n}{3} \leq |A| \leq \frac{2n}{3}\right\}$$

S-n a tartalmazás parciális rendezés; legyen L az S maximális láncainak száma; $\mathcal{M} \subseteq \mathcal{P}(S)$ olyan, hogy $A, B, C \in \mathcal{M}$ különbözők $\Rightarrow$ $A \cup B \neq C$. Legyen $M \in \mathcal{M}$ tetszőleges, $|M| = i$ ; ekkor olyan maximális lánca S-nek, ami átmegy M-en, $\left(1/\binom{m}{i}\right)\cdot L$ sok van. Tekintsük azon maximális láncait S-nek, melyek átmennek M-en és amelyek nem tartalmaznak $M'\in\mathcal{M}$-t, $M\neq M'$, amelyre $M'\subseteq M$. Azok a maximális S-láncok, melyek M-en átmennek és nem ilyenek, legfeljebb $\frac{m}{3i}\cdot L\cdot 1/\binom{m}{i}$ sokan vannak, mert ha minden „rossz" lánchoz a benne levő $\frac{m}{3}$ -as halmazt rendeljük, akkor ezen $\frac{m}{3}$ -asok közül bármely kettő uniója nem $M$; így ezek száma az Erdős-Chao Ko-Rado[13] tétel szerint legfeljebb $\binom{i-1}{\frac{m}{3}-1}$. $\frac{m}{3}$ -as része M-nek $\binom{i}{\frac{m}{3}}$ van ; tehát legfeljebb $L/\binom{m}{i}\cdot\left(1-\binom{i-1}{\frac{m}{3}-1}\Big/\binom{i}{\frac{m}{3}}\right)$ sok rossz lánc van.

Jelölje $\mathcal{M}_i$ az $\mathcal{M}$ $i$ számosságú elemeinek halmazát; ekkor így

$$\sum_{i=\frac{m}{3}}^{\frac{2m}{3}}|\mathcal{M}_i|\cdot\frac{m}{3i}\frac{L}{\binom{m}{i}}\leq L\text{, azaz }\sum_{i=\frac{m}{3}}^{\frac{2m}{3}}|\mathcal{M}_i|\cdot\frac{m}{3i}\frac{1}{\binom{m}{i}}\leq 1.\quad i\leq\frac{2m}{3},$$

tehát $\frac{m}{3i}\geq\frac{m}{\frac{2m}{3}\cdot 3}=\frac{1}{2}$ és $\frac{1}{\binom{m}{i}}\geq\frac{1}{\binom{m}{[\frac{m}{2}]}}$; tehát $\sum_{i=\frac{m}{3}}^{2m/3}|\mathcal{M}_i|\leq 2\binom{m}{[\frac{m}{2}]}$;

azaz $|\mathcal{M}|\leq 2\cdot\binom{m}{[\frac{m}{2}]}$.

A Stirling-formulát használva könnyen látható, hogy

$$\left|\left\{A\subseteq\Omega : |A|<\frac{m}{3}\text{ vagy }|A|>\frac{2m}{3}\right\}\right|<\frac{1}{m}\binom{m}{[\frac{m}{2}]}\text{ ; így}$$

$c=2+\varepsilon$, ahol $\varepsilon$ tetszőlegesen kicsi, ha n elég nagy

$\square$

Ez a bizonyitás Pach Jánostól származik; nem nehéz a bizonyitást úgy módositani, hogy a $c\cdot\binom{n}{[\frac{n}{2}]}$ helyett $\left(1+\frac{c\log n}{\sqrt{n}}\right)\binom{n}{[\frac{n}{2}]}$ felső becslést kapjuk; a binomiális eloszlás közismert tulajdonságait használva az $\frac{n}{3}$ és $\frac{2n}{3}$ helyett kell jobb határokat választani.

Ez az eredmény az, amelyre utaltunk az 1.2 Tétel előtt. Az ott igért becslés egy teljes család leirásához szükséges információmennyiségre $\left(1+\frac{c\log n}{\sqrt{n}}\right)\binom{n}{[\frac{n}{2}]}$.

Érdekes és valószinüleg nehéz probléma a következő:

<u>1. Probléma</u> Mekkora $|\{\mathcal{F} : \mathcal{F}$ teljes család $\Omega$-n$\}|$, ha $|\Omega|=n$

<u>2. Probléma</u> Igaz-e, hogy létezik n-től független $c<1$ konstans úgy, hogy ha $\mathcal{C}\subseteq\mathcal{P}(\Omega)$ Sperner rendszer, akkor legfeljebb $2^{n\cdot c}$ db olyan teljes család létezik, melynek $\mathcal{C}$ a kulcsrendszere?

Az 1.6 Tétellel kapcsolatban megjegyezzük, hogy létezik $\mathcal{M}$ metszettulajdonságú rendszer, amelyre $|\mathcal{M}|=\left(1+\frac{1}{n}\right)\binom{n}{[\frac{n}{2}]}$. Legyen ugyanis $n=2m+1$, $|\Omega|=n$, $a\in\Omega$. Legyen $\mathcal{M}=\{A\subseteq\Omega : a\in A, |A|=m+1\}\cup\{A\subseteq\Omega : a\notin A, |A|=m\}$. Ekkor könnyen ellenőrizhető, hogy $\mathcal{M}$ metszettulajdonságú és $|\mathcal{M}|=\left(1+\frac{1}{n}\right)\binom{n}{[\frac{n}{2}]}$, mert $|\mathcal{M}|=2\cdot\binom{2m}{m}$.

1.6-t uniótulajdonságú rendszerre bizonyitottuk, de mivel $\mathcal{M} \subseteq \mathcal{P}(\Omega)$ uniótulajdonságú akkor és csak akkor, ha $\{\Omega - A : A \in \mathcal{M}\}$ metszettulajdonságú, ezért az eredmény valóban igaz metszettulajdonságú rendszerekre is.

# II. Fejezet

## Eredmények

Mint már azt korábban emlitettük, a gyakorlatban általában az a helyzet, hogy az adatbázis alaprelációinak néhány funkcionális függését ismerjük; a relációs adatmodell elméletet arra szeretnénk használni, hogy minél rövidebb úton tudjunk következtetni további függésekre; pl. az $(A, \Omega)$ alakú kulcsokra vagy az egyéb maximálisakra. A funkcionális függések által generált teljes család fogalmát bevezettük az I. Fejezetben, ezen hallgatólag az adott függéseket tartalmazó legszükebb teljes családot értettük. Ilyen létezik, mert könnyen ellenőrizhető, hogy teljes családok metszete teljes család. A [3] alapján a következőkben megteremtjük azon eszközöket, melyekkel a gyakorlatban hatékony módon tárgyalhatók a függőségek.

A továbbiakban kényelmes lesz $\{(F1),(F2),(F3),(F4)\}$ helyett a következő axiómarendszerrel dolgozni:

$$(A1)\quad A \subseteq \Omega \Rightarrow (A, A) \in \mathcal{F}$$

$$(A2)\quad (A,B) \in \mathcal{F}, (C,D) \in \mathcal{F} \Rightarrow (A \cup (\overline{B} \cap C), A \cup B \cup C \cup D) \in \mathcal{F}$$

$$(A3)\quad (A,B) \in \mathcal{F},\ C \supseteq A,\ D \subseteq B \Rightarrow (C, D) \in \mathcal{F}$$

Alternativ jelölésmóddal: $(A,B)\in\mathcal{F}$, $(C,D)\le(A,B)\Rightarrow(C,D)\in\mathcal{F}$

2.1 lemma $\mathcal{F}$ teljes család akkor és csak akkor, ha $\mathcal{F}$-re teljesül (A1), (A2) és (A3)

Biz.: (A2) következik (F1)-(F4) -ből; mert legyen $(A,B),(C,D)\in\mathcal{F}$; ekkor (F1) miatt $(\bar{B}\cap C,\bar{B}\cap C)\in\mathcal{F}$; igy (F4) szerint $(A,A\cup B)\in\mathcal{F}$ és $(A\cup(\bar{B}\cap C),A\cup B\cup(\bar{B}\cap C))=(A\cup(\bar{B}\cap C),A\cup B\cup C)\in\mathcal{F}$, (F4) alapján $(A\cup B\cup C,A\cup B\cup C\cup D)\in\mathcal{F}$; igy (F2) miatt végül $(A\cup(\bar{B}\cap C),A\cup B\cup C\cup D)\in\mathcal{F}$.
Forditva, (F2) következik (A2) -ből $B=C$ -vel és (A3)-t használva, és $A\cup(\bar{B}\cap C)\subseteq A\cup C$ miatt (F4) következik (A2), (A3) -ból. □

2.1 Definició $*$ a következő kétváltozós müvelet $\mathcal{P}(\Omega)\times\mathcal{P}(\Omega)$-n; ha $e_1=(A_1,B_1)$, $e_2=(A_2,B_2)\in\mathcal{P}(\Omega)\times\mathcal{P}(\Omega)$, akkor legyen

$$e_1*e_2=\left(A_1\cup(\bar{B}_1\cap A_2),\ B_1\cup B_2\right)$$

2.1 lemma miatt ha $\mathcal{F}$ teljes család, akkor $\mathcal{F}$ zárt $*$ -ra, azaz $e_1,e_2\in\mathcal{F}\Rightarrow e_1*e_2\in\mathcal{F}$
Néhány észrevételt teszünk a $*$ müveletről;

1. $*$ asszociativ, azaz $e_1*(e_2*e_3)=(e_1*e_2)*e_3$
2. $e_i=(A_i,B_i)$, $B_i\supseteq A_i$ : $i=1,\dots k$ ; ekkor 1. szerint $e_1*e_2*\dots*e_k$ zárójelezés nélkül is egyértelmü;

$e = e_1 * e_2 * \ldots * e_k = \left( L(e), \bigcup_{i=1}^{k} B_i \right)$ ahol

$L(e) = A_1 \cup (\overline{B}_1 \cap A_2) \cup (\overline{B}_1 \cap \overline{B}_2 \cap A_3) \cup \ldots \cup (\overline{B}_1 \cap \overline{B}_2 \cap \ldots \cap \overline{B}_{k-1} \cap A_k) =$

$= (A_1 \cup \overline{B}_1) \cap (A_1 \cup A_2 \cup \overline{B}_2) \cap \ldots \cap (A_1 \cup A_2 \cup \ldots \cup A_{k-1} \cup \overline{B}_{k-1}) \cap (A_1 \cup \ldots \cup A_k)$

3. ha $e_i = (A_i, A_i)$ ; $i = 1, \ldots k$ ; akkor

$e_1 * e_2 * \ldots * e_k = \left( \bigcup_{i=1}^{k} A_i , \bigcup_{i=1}^{k} A_i \right)$ azaz triviális függésekből a $*$ müvelet triviális függést eredményez.

4. ha $e = (A, B)$, $B \supseteq A$ és $f = (C, C)$ ; akkor

$L(e * f) \subseteq L(f * e)$ , hiszen

$e * f = (A \cup (\overline{B} \cap C), B \cup C)$ és

$f * e = (C \cup (\overline{C} \cap A), B \cup C) = (C \cup A, B \cup C)$

Tehát ilyen $e, f$ mellett $e * f$ több információt adó függés, mint $f * e$ /pontosabban $e * f$ legalább annyi információt ad, mint $f * e$ /

5. a $*$ müvelet idempotens, sőt

$e_1 * e_2 * \ldots * e_k * e_j = e_1 * e_2 * \ldots * e_k$ minden $1 \leq j \leq k$ -re

A továbbiakban egy $(A, B)$ függést triviálisnak nevezünk, ha $B \subseteq A$ . Legyen $\mathcal{F}'' \subseteq \mathcal{P}(\Omega) \times \mathcal{P}(\Omega)$ olyan, hogy $A \subseteq \Omega \Rightarrow (A, A) \in \mathcal{F}''$ . Legyen $\mathcal{G}$ az $\mathcal{F}''$ $*$-zárt burka /ilyen van, mert $*$-zárt halmazok metszete is $*$-zárt; ahol egy $\mathcal{C} \subseteq \mathcal{P}(\Omega) * \mathcal{P}(\Omega)$ $*$-zárt, ha $(\forall k)(\forall e_1, \ldots e_k \in \mathcal{C})(e_1 * e_2 * \ldots * e_k \in \mathcal{C})$ /

Ekkor $\mathcal{G}$ -re nyilván igaz (A1) és (A2); mig (A3) nem feltétlenül. Jelölje $\mathcal{F}$ az $\mathcal{F}''$ által generált teljes családot. $\mathcal{G} \subseteq \mathcal{F}$ nyilvánvaló, de általában $\mathcal{G} \neq \mathcal{F}$, azaz $\mathcal{G}$ nem feltétlenül teljes család; de igaz a következő

<u>2.2 lemma</u> $\hat{\mathcal{F}} = \hat{\mathcal{G}}$, és igy $\mathcal{F}$ kulcsai éppen a $\mathcal{G}$ kulcsai

<u>Biz.:</u> legyen $\mathcal{G}' = \{(C, D) : \text{van } (A, B) \in \mathcal{G}$, hogy $C \supseteq A$ és $D \subseteq B\}$
Nyilván elég bizonyitani, hogy $\mathcal{G}' = \mathcal{F}$. $\mathcal{G}' \subseteq \mathcal{F}$ világos.
Mivel $\mathcal{F}$ az $\mathcal{F}''$ által generált teljes család és $\mathcal{G}'$ -re igaz (A1) és (A3) és $\mathcal{F}'' \subseteq \mathcal{G}'$, azért elegendő megmutatni, hogy $\mathcal{G}'$ -re igaz (A2).
Legyen tehát $(A_1, B_1) \in \mathcal{G}'$, $(A_2, B_2) \in \mathcal{G}'$. Ekkor létezik $(A_1', B_1'), (A_2', B_2') \in \mathcal{G}$ úgy, hogy $A_i \supseteq A_i'$, $B_i \subseteq B_i'$, ha $i = 1, 2$. $(A_1', B_1') * (A_2', B_2') = (A_1' \cup (\overline{B}_1' \cap A_2'), B_1' \cup B_2') \in \mathcal{G}$, hiszen $\mathcal{G}$ $*$-zárt.
$A_1 \cup (\overline{B}_1 \cap A_2) \supseteq A_1' \cup (\overline{B}_1' \cap A_2')$, hiszen $A_1 \supseteq A_1'$ és $A_2 \supseteq A_2'$; és $B_1 \subseteq B_1'$ miatt $\overline{B}_1' \subseteq \overline{B}_1$.
Igy $(A_1 \cup (\overline{B}_1 \cap A_2)), A_1 \cup A_2 \cup B_1 \cup B_2) \in \mathcal{G}'$, mert $(A, A) * (A, B) = (A, A \cup B)$ és $B_i \subseteq B_i'$ ha $i = 1, 2$ $\square$

<u>2.3 lemma</u> $e_i = (A_i, B_i)$: $i = 1, \ldots k$ nem triviális funkcionális függések; $\mathcal{F}$ az általuk generált teljes család;

$\omega = \bigcup_{i=1}^{k} B_i$, $\Gamma = \Omega - \omega$. Ekkor minden $(X,\Omega) \in \mathcal{F}$-hez létezik $(\pi_1,\dots,\pi_k)$ permutációja $\{1,2,\dots k\}$-nak úgy, hogy $X \supseteq L(e_{\pi_1} * e_{\pi_2} * \dots * e_{\pi_k} * g)$ ahol $g = (\Gamma,\Gamma)$

Biz.: ha $(X,\Omega) \in \mathcal{F}$ akkor 2.2 lemma szerint létezik $X' \subseteq X$ úgy, hogy $(X',\Omega) \in \mathcal{G} (\subseteq \mathcal{F})$. Tehát $(X,\Omega) \in \mathcal{G}$ feltehető. Ekkor, $\mathcal{G}$ definiciója szerint valamely $d = (D,D)$-re $X = L(e_{\pi_1} * e_{\pi_2} * \dots * e_{\pi_s} * d) = A_{\pi_1} \cup \dots \cup (\overline{B}_{\pi_1} \cap \dots \cap \overline{B}_{\pi_s} \cap D)$

Tehát (1) $L(e_{\pi_1} * \dots * e_{\pi_k} * g) \subseteq A_{\pi_1} \cup (\overline{B}_{\pi_1} \cap A_{\pi_2}) \cup \dots \cup (\overline{B}_{\pi_1} \cap \dots \cap D)$ fennállását kell bizonyitanunk.

Persze $B_1 \cup B_2 \cup \dots \cup B_s \cup D = \Omega$ ; igy

$D \supseteq (B_{s+1} \cup \dots \cup B_k \cup \Gamma) \cap \overline{B}_1 \cap \overline{B}_2 \cap \dots \cap \overline{B}_s$; továbbá,

mivel $A_i \subset B_i$, azért

$$D \supseteq (A_{s+1} \cup \dots \cup A_k \cup \Gamma) \cap \overline{B}_1 \cap \dots \cap \overline{B}_s \text{ ; tehát}$$

$$\overline{B}_1 \cap \overline{B}_2 \cap \dots \cap \overline{B}_s \cap D \supseteq (A_{s+1} \cap \overline{B}_1 \cap \overline{B}_2 \cap \dots \cap \overline{B}_s) \cup$$
$$\cup \dots \cup (A_k \cap \overline{B}_1 \cap \dots \cap \overline{B}_{k-1}) \cup (\Gamma \cap \overline{B}_1 \cap \dots \cap \overline{B}_k),$$

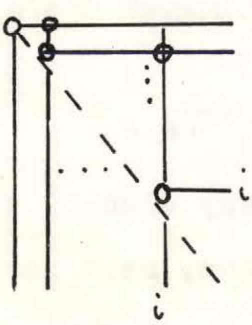
ami mutatja, hogy (1) igaz. □

Következmény: k db funkcionális függésből legfeljebb k! sok kulcs vezethető le. /Osborne és Tompa tétele/ Igy ha az $\mathcal{F}$ teljes családot k db nem triviális eleme generálja, akkor $\mathcal{F}$ kulcsainak száma legfeljebb k!

2.2 Tétel [15] Minden k-ra létezik k(k-1) számosságú $\Omega$ és k nemtriviális függés felette, amelyek k! sok kulcsot generálnak.

Biz.: legyen $|\Omega| = k\cdot(k-1)$; $\Omega$ elmeit soroljuk fel egy $k \times k$ -as $M$ mátrixban, a főátlót üresen hagyva. Legyen $A_i$ a mátrix i-edik sorában levő elemek halmaza, $B_i$ az i-edik oszlopbelieké. Ekkor a mátrix főátló feletti elemei kulcsot alkotnak, hiszen az első sortól függ az első oszlop; igy a második sor első eleme függ az első sortól, hiszen az első oszlopban van; igy a felső háromszög első két sorának uniójától függ az első két oszlop uniója, stb.; a felső háromszög első k sorának uniójától függ az első k oszlop uniója, ez k-ra indukcióval könnyen igazolható.

M



Ha $\pi = (\pi_1, \dots, \pi_k)$ az $\{1, .., k\}$ -nak egy tetszőleges permutációja, akkor legyen $K_\pi$ annak a mátrixnak felső háromszögmátrixának elemhalmaza, amely M-ből a sorok és oszlopok $\pi$ szerinti permutációjával keletkezik. Könnyü belátni, hogy $\pi \neq \pi' \Rightarrow K_\pi \neq K_{\pi'}$ és $K_\pi$ az $\{(A_i, B_i) : i = 1, \dots, k\}$ által generált teljes családnak kulcsa. Tehát $(A_i, B_i)$ - k k! sok kulcsot generálnak. $\square$

$\Omega$ elemszáma a 2.2 Tételben éles, amint azt a következő tétel mutatja

<u>2.3 Tétel</u> Ha k funkcionális függésből k! kulcs keletkezik, akkor az attributumok száma legalább $k\cdot(k-1)$

<u>Biz.:</u> legyenek $(A_i, B_i): i=1,\dots k$ a kezdő funkcionális függések. Ekkor 2.3 lemma szerint az általuk meghatározott kulcsok halmaza $\mathcal{C}$ a

$$\{L_\pi = L(e_{\pi_1} * \dots * e_{\pi_k} * g) : \pi = (\pi_1, \dots, \pi_k) \text{ permutációja } \{1, \dots k\}\text{-nek}\}.$$

Ha $|\mathcal{C}| = k!$ akkor $\pi \neq \pi' \Rightarrow L_\pi \nsubseteq L_{\pi'}, L_{\pi'} \nsubseteq L_\pi$. Legyen $\pi = (\pi_1, \dots, \pi_k)$-ra

$$Z(\pi_{k-1}, \pi_k) = \overline{B}_{\pi_1} \cap \overline{B}_{\pi_2} \cap \dots \cap \overline{B}_{\pi_{k-2}} \cap A_{\pi_{k-1}}$$

Minden $Z(i,j)$ tartalmazza $\Omega$-nak egy elemét, amelyet $(i',j') \neq (i,j)$ esetén $Z(i',j')$ nem tartalmaz, mert ha nem, akkor $\overline{B}_1 \cap \overline{B}_2 \cap \dots \cap \overline{B}_{k-2} \cap A_{k-1} = (\overline{B}_1 \cap \overline{B}_2 \cap \dots \cap \overline{B}_{k-2} \cap \overline{B}_k \cap A_{k-1}) \cup$

$\cup (\overline{B}_1 \cap \overline{B}_2 \cap \dots \cap \overline{B}_{k-2} \cap A_{k-1} \cap A_k)$ állna; és igy

$L(e_1 * \dots * e_{k-1} * e_k) \subseteq L(e_1 * \dots * e_k * e_{k-1})$ lenne, ami ellentmondás.

Mivel $|\{Z(i,j) : i \neq j\}| = k(k-1)$, azért igy $|\Omega| \geq k(k-1)$ valóban $\square$

<u>2.2 Definició</u> $Z(i,j)$ <u>kritikus halmaza</u>

$$S(i,j) = Z(i,j) - \bigcup_{(i',j') \neq (i,j)} Z(i',j')$$

Ekkor 2.3 Tétel feltételei mellett egyetlen kritikus halmaz sem üres.

2.4 Tétel Ha az $(A_i, B_i)$: $i=1\dots k$ funkcionális függések $k!$ kulcsot generálnak, akkor

(i) $(\forall i \le k)(|A_i| \geq k-1)$

(ii) minden kulcs elemszáma legalább $k(k-1)/2$

Bizonyitás (i) $A_{\pi_1} \supseteq A_{\pi_1} \cap \overline{B}_{\pi_2} \cap \dots \cap \overline{B}_{\pi_{k-1}}$ ;
azaz legalább $(k-1)$ kritikus halmazt tartalmaz $A_i$ ;
előző megjegyzésünk szerint egyetlen kritikus halmaz sem üres, tehát $|A_i| \geq k-1$

(ii) Ha $K = A_{\pi_1} \cup (\overline{B}_{\pi_1} \cap A_{\pi_2}) \cup \dots \cup (\overline{B}_{\pi_1} \cap \overline{B}_{\pi_2} \cap \dots \cap \overline{B}_{\pi_{k-1}} \cap A_{\pi_k})$,
akkor a K-t képező unió tagjai nyilvánvalóan diszjunktak;
$A_{\pi_1}$ $(k-1)$ kritikus halmazt tartalmaz; $(\overline{B}_{\pi_1} \cap A_{\pi_2})$ $(k-2)$-t, stb.; tehát $|K| \geq \frac{k(k-1)}{2}$

$\square$

Vegyük észre, hogy 2.3 lemma ad egy algoritmust adott funkcionális függések által generált teljes család kulcsainak megtalálására. Az algoritmus általában k! sok lépést igényel, ha k db függés adott; a 2.2 Tétel mutatja, hogy általában ennél jobb nincs.

Levezethetőség, teljes család generálása

A következőkben teljes családok generátorrendszereiről lesz szó. Megjegyezzük, hogy a levezethetőség fogalma lényegében ugyanaz, mint a $*$ müvelettel való előállithatóság; az egyetlen különbség az, amire a 2.2 lemma mutat rá; azaz, hogy a $*$ müvelettel általában csak a maximális függéseket tudjuk előállitani.

Meg akarjuk válaszolni azt a kérdést, hogy egy adott $\mathcal{F}$ teljes család generálásához hány eleme szükséges.

Az I. Fejezet utolsó tétele válaszol erre; hiszen, $\mathcal{B}$ -vel jelölve $\mathcal{F}$ maximális jobboldalainak rendszerét, $\mathcal{B}$ meghatározza $\mathcal{F}$ -et. $\mathcal{B}$ -nek egyetlen minimális számosságú generátorrendszere van, amely $\mathcal{B}$ un. metszetirreducibilis elemeiből áll, azaz

$$\mathcal{J} = \{ B \in \mathcal{B} : (\forall \mathcal{B}' \subseteq \mathcal{B})(B = \cap \mathcal{B}' \Rightarrow B \in \mathcal{B}') \}$$

$\mathcal{J}$ generálja $\mathcal{B}$ -t, mint metszetfélhálót. Tegyük fel ugyanis, hogy $B \in \mathcal{B}$ maximális számosságú, amely nincs az $\mathcal{J}$ által generált metszetfélhálóban. Ekkor természetesen $B \in \mathcal{B} - \mathcal{J}$ ; igy $B = \cap \{ C : C \supseteq B, C \in \mathcal{B} \setminus \{B\} \}$; $C \supsetneqq B \Rightarrow |C| > |B|$, ezért ha $C \in \mathcal{B}$, $C \supsetneqq B$ , ugy létezik $\mathcal{J}_C \subseteq \mathcal{J}$ úgy, hogy $C = \cap \mathcal{J}_C$ ; ekkor $B = \cap \{ \cap \mathcal{J}_C : C \in \mathcal{B} \setminus \{B\}, C \supseteq B \}$; tehát $\mathcal{J}_B = \cup \{ \mathcal{J}_C : C \supseteq B, C \in \mathcal{B} \setminus \{B\} \} \subseteq \mathcal{J}$ -re $B = \cap \mathcal{J}_B$ és igy $B$ az $\mathcal{J}$ által generált metszetfélhálóban van, ami ellentmondás.

Másrészt nyilvánvaló, hogy ha $\mathcal{J}'$ generálja $\mathcal{B}$ -t, akkor $\mathcal{J} \subseteq \mathcal{J}'$ .

Tehát ha egy $\mathcal{F}' \subseteq \mathcal{F}$ függéshalmaz generálja $\mathcal{F}$ -et, akkor $|\mathcal{F}'| \geqslant |\mathcal{J}|$ . Legyen $\mathcal{J} = \{B_i : i = 1, \dots k\}$ ; legyenek $A_i \subseteq \Omega : i = 1, \dots k$ olyanok, hogy $(A_i, B_i) \in \mathcal{F}$. Ekkor $\{(A_i, B_i) : i = 1, \dots k\}$ nyilvánvalóan generálja $\mathcal{F}$ -t. Tehát $\mathcal{F}$ minimális számosségú generátorrendszerei $|\mathcal{J}|$ számosságúak; és 1. tétel szerint $|\mathcal{J}| \leq 2\binom{m}{[\frac{m}{2}]}$, ahol $|\Omega| = m$.

Szeretnénk leirni a minimális számosságú generátorrendszereket. Ehhez a levezethetőség fogalma lesz hasznos.

2.2 Definició [2] $\mathcal{F}'$ funkcionális függések halmaza; ekkor egy (A,B) funkcionális függés levezethető $\mathcal{F}'$ -ből, ha létezik $m \geqslant 0$ és léteznek $(A_i, B_i) \in \mathcal{F}' : i = 1, \dots, m$ úgy, hogy ha $X_0 = A$ ; $X_i = X_{i-1} \cup B_i : i = 1, \dots m$ ; akkor $X_m \supseteq B$ és $X_{i-1} \supseteq A_i$ , ha $i = 1, \dots m$

2.3 Definició [8] $\mathcal{F}'$ funkcionális függések halmaza; ekkor egy (A,B) funkcionális függés levezethető $\mathcal{F}'$ -ből, ha a következő négy szabály véges sokszori alkalmazásával nyerhető;

(1) ha $A' \subseteq A \subseteq \Omega$, akkor $(A, A')$ levezethető $\mathcal{F}'$ -ből;

2 ha $(A, B)$ és $(B, C)$ levezethetők $\mathcal{F}'$ -ből akkor (A,C) is levezethető $\mathcal{F}'$ -ből

(3) $(A,B)$ és $(A,C)$ levezethetők $\mathcal{F}'$-ből akkor $(A,B\cup C)$ is levezethető $\mathcal{F}'$-ből;

(4) $\mathcal{F}'$ minden eleme levezethető $\mathcal{F}'$-ből.

2.4 lemma A 2.2 Definició szerinti levezethetőség fogalma megegyezik a 2.3 Definició szerinti levezethetőség fogalmával.

Bizonyitás legyen $\mathcal{F}'$ funkcionális függések halmaza $\Omega$ felett és tegyük fel, hogy egy $(A,B)\in \mathcal{P}(\Omega)\times\mathcal{P}(\Omega)$ a 2.2 Definició szerint levezethető $\mathcal{F}'$-ből. Ekkor létezik $\{(A_i,B_i)\in\mathcal{F}' : i=1,\dots m\}$ úgy, hogy $X_0=A$, $X_i=X_{i-1}\cup B_i$ $\Rightarrow X_{i-1}\supseteq A_i$, ha $i=1,\dots m$ és $X_m\supseteq B$. $X_0\supseteq A_1$; $(A_1,B_1)$ 2.3-levezethető $\mathcal{F}'$-ből (4) szerint és $(X_0,A_1)$ 2.3-levezethető $\mathcal{F}'$-ből (1) szerint. Igy, (2) szerint $(X_0,B_1)$ 2.3-levezethető $\mathcal{F}'$-ből. $(X_0,X_0)$ (1) szerint 2.3-levezethető $\mathcal{F}'$-ből; igy, (3) miatt $(X_0,X_0\cup B_1)$ 2.3-levezethető $\mathcal{F}'$-ből. Tegyük fel, hogy $k\leq m$ és $(X_0,X_{k-1})$ 2.3-levezethető $\mathcal{F}'$-ből. $X_k=X_{k-1}\cup B_k$; $A_k\subseteq X_{k-1}$, igy ekkor $(X_{k-1},A_k)$ (1) miatt 2.3-levezethető $\mathcal{F}'$-ből; igy, $(A_k,B_k)\in\mathcal{F}'$ miatt $(X_{k-1},B_k)$ 2.3-levezethető $\mathcal{F}'$-ből, (2) szerint; újra (2)-t alkalmazva $(X_0,X_{k-1})$ és $(X_{k-1},B_k)$-ra, kapjuk, hogy $(X_0,B_k)$ 2.3-levezethető $\mathcal{F}'$-ből. $(X_0,B_k)$

és $(X_0, X_{\ell-1})$ tehát 2.3-levezethetők $\mathcal{F}'$ -ből; igy (3) miatt $(X_0, X_\ell)$ is 2.3-levezethető $\mathcal{F}'$-ből. Végül $(X_0, X_m)$ 2.3-levezethető $\mathcal{F}'$ -ből; $B \subseteq X_m$ és (1) miatt $(X_m, B)$ 2.3-levezethető $\mathcal{F}'$-ből és igy, (2) szerint $(X_0, B) = (A, B)$ 2.3-levezethető $\mathcal{F}'$ -ből.

Forditva tegyük fel, hogy $(A,B)$ 2.3-levezethető $\mathcal{F}'$-ből. A bizonyitáshoz szükség lesz az $\tau(A, B)$ halmazfüggvényre, amelyet előzetesen definiálunk.

(1) Ha $(A', B')$ 2.3-levezethető $\mathcal{F}'$ -ből, akkor legyen $\tau(A', B')$ a következő; ha $(A', B') \in \mathcal{F}'$, valamint ha $A' \supseteq B'$ akkor legyen $\tau(A', B') = 0$.

(2) Legyen $(A', B')$ 2.3-levezethető $\mathcal{F}'$ -ből, $\tau(A', B')$ még definiálatlan és létezzenek $(A'_0, B'_0)$ és $(A'_1, B'_1)$ az $\mathcal{F}'$ -ből 2.3 levezethető függések úgy, hogy $\tau(A'_i, B'_i)$: $i = 0, 1$ már definiált, és $(A', B')$ előállitható $(A'_0, B'_0), (A'_1, B'_1)$ -ból (1), (2), (3), (4) valamelyikét alkalmazva. Legyen

$$\tau(A', B') = \max_{i=0,1} \left( \tau(A'_i, B'_i) \right) + 1$$

$\tau(A, B)$ -re vonatkozó indukcióval bizonyitjuk, hogy ha $(A, B)$ 2.3-levezethető $\mathcal{F}'$ -ből, akkor 2.2-levezethető is.

Legyenek $(A_0, B_0), (A_1, B_1)$ olyanok, hogy $\tau(A_i, B_i) < \tau(A, B)$, ha $i = 0, 1$; mutassa $(A_i, B_i)$ 2.2-levezethető voltát $\mathcal{F}'$-ből

$(A^i_j, B^i_j) \in \mathcal{F}'$: $j = 1, \dots, m_i$; $X^i_j$ : $j = 0, \dots m_i$; $i = 0, 1$

Ha $(A,B)$ (2) alapján keletkezik $(A_0,B_0),(A_1,B_1)$ -ből, akkor $X_0^1=A_1=B_0\subseteq X_{m_0}$ ; legyen $X_{m_0+i}=X_i^1\cap B_0$, ha $i=0,\dots m_0$; $X_i=X_i^0$, ha $i=0,\dots,m_0-1$; nyilvánvaló, hogy ez mutatja, hogy $(A,B)$ 2.2-levezethető $\mathcal{F}'$ -ből. Ha $(A,B)$ (3) alapján adódik $(A_0,B_0),(A_1,B_1)$ -ből, akkor $A_0=A_1$; azaz $X_0^1=X_0^0$; $B_0\cup B_1\subseteq X_{m_0}^0\cup X_{m_1}^1$; tekintsük az $X_0^0,\dots X_{m_0}^0, X_1^1\cup X_{m_0}^0,\dots X_{m_1}^1\cup X_{m_0}^0$ sorozatot; könnyen látható, hogy ez mutatja $(A,B)$ 2.2-levezethetőségét $\mathcal{F}'$ -ből. $\square$

2.4 lemma alapján a továbbiakban egyszerüen levezethetőséget mondunk. A 2.3 definicióból nyilvánvaló, hogy ha $\mathcal{F}'$ funkcionális függések halmaza, akkor az $\mathcal{F}'$ által generált teljes család

$$\mathcal{F}=\{(A,B):(A,B)\text{ levezethető } \mathcal{F}'\text{-ből}\}$$

Megjegyezzük, hogy a levezethetőség mindkét alakja hasznos; a 2.2 Definició a levezethetőség algoritmikus voltát mutatja, mig a 2.3 Definiciót az (F1)-(F4) axiómákkal összevetve világos, hogy az $\mathcal{F}'$-ből levezethetők éppen az $\mathcal{F}'$ által generált teljes család. Ha $\mathcal{F}'$ minimális generátorrendszere az $\mathcal{F}$ teljes családnak, akkor nyilvánvaló, hogy $(A,B)\in\mathcal{F}'$ esetén $(A,B)$ nem vezethető le $\mathcal{F}'-\{(A,B)\}$ -ből. Ha funkcionális függések egy $\mathcal{F}$ rendszere ilyen, akkor azt mondjuk hogy $\mathcal{F}'$ redundanciamentes.

Legyen $\mathcal{F}' = \{(A_i, B_i) : i = 1, \dots q\}$ és legyen $\mathcal{F}$ az $\mathcal{F}'$ által generált teljes család. Ekkor $\mathcal{F}$ minden $(A_i, B_i)$ elemét kicserélhetjük $\mathcal{F}$ -nek egy maximális $(A'_i, B'_i)$ elemére úgy, hogy először megkeressük azt a maximális $B'_i$-t, amelyre $(A_i, B'_i)$ levezethető $\mathcal{F}'$ -ből és $B_i \subseteq B'_i$, majd ugyanigy minimalizáljuk $A_i$ -t egy $A'_i \subseteq A_i$ -vé; $\{(A'_i, B'_i) : i = 1, \dots q\}$ ekkor szintén $\mathcal{F}$ -et generálja. /Előfordulhat természetesen, hogy két különböző $(A, B) \in \mathcal{F}'$ ugyanahhoz az $(A', B') \in \mathcal{F}$ maximális függéshez vezet/.

Igy $\mathcal{F}$ minden $\mathcal{F}'$ generátorrendszere helyettesithető egy legfeljebb $|\mathcal{F}'|$ számosságú, maximális függésekből álló generátorrendszerrel. Ha $(A, B) \in \mathcal{F}'$ triviális függés, azaz $A' \supseteq B'$, akkor $\mathcal{F}' \setminus \{(A', B')\}$ ugyanazt a teljes családot generálja, mint $\mathcal{F}'$.

Jelölje $\mathcal{F}$ maximális elemeinek halmazát /mint az I. Fejezetnél/ $\hat{\mathcal{F}}$. $\hat{\mathcal{F}}$ nyilvánvalóan generálja $\mathcal{F}$ -et.

<u>2.4 Definició</u>[2] $\mathcal{F}$ teljes család; $(A, B) \in \hat{\mathcal{F}}$ nemtriviális függés, azaz $A \not\supseteq B$; ekkor $(A, B)$ <u>egyszerü</u>, ha $(A, B)$ <u>nem vezethető le</u> $(\hat{\mathcal{F}} \setminus \{(A', B') : B' = B\})$ -ből.

Pl. tekintsük az $\Omega = \{a, b, c, d, e, f, g\}$ felett az $\mathcal{F}' = \{(\{a, b\}, \{c\}), (\{c, d\}, \{e\}), (\{e, f\}, \{g\})\}$ által generált $\mathcal{F}$ teljes családot. Ekkor $(\{a, b, d, f\}, \{a, b, c, d, e, f, g\}) \in \hat{\mathcal{F}}$ nem egyszerü, mig $(\{a,b\},\{a,b,c\}),(\{c,d\},\{c,d,e\}),(\{e,f\},\{e,f,g\})$ egyszerüek.

2.4 Definicióban az a fontos, hogy egyszerü $(A,B)$ nem vezethető le olyan maximális függésekből, melyek jobboldala valódi része $B$ -nek.

<u>2.5 Tétel</u>[2] $\mathcal{F}$ -t generálják az egyszerü függései

<u>Bizonyitás</u> $\mathcal{F}$ -t $\hat{\mathcal{F}}$ generálja. Tegyük fel, hogy $\hat{\mathcal{F}}$ -nek létezik olyan eleme, amely nem vezethető le $\mathcal{F}$ egyszerü függéseiből; legyen $(A,B)\in\hat{\mathcal{F}}$ egy ilyen úgy, hogy $|B|$ minimális. Ekkor $(A,B)$ nyilvánvalóan nem egyszerü; igy $(A,B)$ levezethető $\mathcal{F}$ olyan maximális függéseiből, melyeknek jobboldala valódi része $B$ -nek és igy ezek mindegyike levezethető az $\mathcal{F}$ egyszerü függéseiből, $|B|$ minimalitása miatt. De ekkor nyilvánvaló, hogy $(A,B)$ is levezethető $\mathcal{F}$ egyszerü függéseiből, ami ellentmondás. $\square$

<u>2.5 Definició</u>[2] legyen $\mathcal{F}$ teljes család, $(A,B)$ egyszerü függése $\mathcal{F}$ -nek, $(A',B')\in\hat{\mathcal{F}}$. Ekkor $(A,B)$ <u>származtatható $(A',B')$ -ből</u>, ha $(A,B)$ levezethető $(\hat{\mathcal{F}}\setminus\{(A'',B''): B''=B\})\cup\{(A',B')\}$ -ből. Vegyük észre, hogy ha $(A,B)$ származtatható $(A',B')$ -ből, akkor $B'=B$ és $(A',B')$

egyszerü függése $\mathcal{F}$ -nek. Igy a <u>származtathatóság</u> az egyszerü függései között reláció; könnyen látható, hogy reflexiv és tranzitiv; azaz ha $(A,B)$ $(A',B')$-ből származtatható és $(A',B')$ $(A'',B'')$-ből származtatható, akkor $(A,B)$ $(A'',B'')$ -ből származtatható.

<u>2.6 Definició</u>[2] $\mathcal{F}$ teljes család egy <u>$(A,B)$ egyszerü függése generátorjelölt</u>, ha minden $(A',B')\in\mathcal{F}$-re igaz, hogy ha $(A,B)$ származtatható $(A',B')$ -ből, akkor $(A',B')$ származtatható $(A,B)$ -ből.

<u>2.6 Tétel</u>[2] $\mathcal{F}$ teljes család; ekkor $\mathcal{F}$ minden egyszerü függése származtatható valamely generátorjelöltből.

<u>Bizonyitás</u> legyen $(A,B)\in\mathcal{F}$ egyszerü. Ha $(A,B)$ nem generátorjelölt, akkor $(A,B)$ származtatható egy $(A_1,B)$ egyszerü függésből, amely nem származtatható $(A,B)$ -ből. $(A_1,B)$ vagy generátorjelölt, vagy származtatható egy $(A_2,B)$ egyszerü függésből, amely nem származtatható $(A_1,B)$ -ből. Ezt a processzust folytatva vagy találunk egy generátorjelöltet, amelyből, a származtathatóság tranzitiv volta miatt $(A,B)$ származtatható; vagy, az $\mathcal{F}$ véges volta miatt, találunk $i<j$ -t úgy, hogy $(A_i,B)$ származtatható $(A_j,B)$ -ből és $(A_j,B)$ nem származtatható $(A_{j-1},B)$ -ből

és $A_i = A_j$ ; de ez ellentmondás, mert igy $(A_i, B)$ -ből származtatható $(A_{i+1}, B)$ .

A diagram segiti áttekinteni a processzust: $(A,B) \Rightarrow (A', B)$ jelöli azt, hogy $(A, B)$ -ből származtatható $(A', B)$ .

$$\begin{array}{c}
(A,B) \\
\Uparrow \;\not\Downarrow \\
(A_1,B) \\
\Uparrow \\
\vdots \\
\Uparrow \\
(A_i,B) \\
\vdots \\
\Uparrow \\
(A_l,B) \\
\Uparrow \\
(A_{l+1},B) = (A_j,B)
\end{array}$$

□

Legyenek az $(A,B),(A',B') \in \mathcal{F}$ egyszerü függések ekvivalensek, ha egymásból származtathatók. Ez a reláció nyilvánvalóan ekvivalencia az $\mathcal{F}$ teljes család egyszerü függései között. Ekkor

2.7 Tétel[2] Ha $\mathcal{F}$ teljes család, akkor $\mathcal{F}' \subseteq \hat{\mathcal{F}}$ generálja $\mathcal{F}$ -et akkor és csak akkor, ha $\mathcal{F}'$ minden ekvivalenciaosztályból tartalmaz elemet.

Bizonyitás 1/ ha $\{(A_i, B_i) : i = 1, \dots l\} = \mathcal{F}' \subseteq \hat{\mathcal{F}}$ generálja $\mathcal{F}$ -et, akkor minden $(A, B)$ generátorjelölthöz létezik $\{(A_{j_i}, B_{j_i}) : i = 1, \dots m\} \subseteq \mathcal{F}'$ úgy, hogy ha $X_0 = A$ , $X_i = X_{i-1} \cup B_{j_i}$ : $i = 1, \dots m$ ; akkor $A_{j_i} \subseteq X_{i-1}$ , ha $i = 1, \dots m$ ; és $X_m = B$ .

$B_{j_i} \subset B$, ha $i = 1, \dots m-1$ feltehető, igy $B_{j_m} = B$. Tehát $(A, B)$ származtatható $(A_{j_m}, B_{j_m})$ -ből; ezért $(A_{j_m}, B_{j_m})$ egyszerü és igy az $(A, B)$ ekvivalencia-osztályából való.

2/ legyen $\{(A_i, B_i): i=1, \dots q\} = \mathcal{F}'$ olyan, hogy $\mathcal{F}'$ minden ekvivalenciaosztályt metsz. 2.5 Tétel alapján elég bizonyitani, hogy ekkor minden egyszerü függés levezethető $\mathcal{F}'$-ből. Tegyük fel, hogy ez nem igaz és legyen $(A, B)$ $\mathcal{F}'$-ből nem levezethető egyszerü függés és $|B|$ minimális. $(A, B)$ a 2.6 Tétel szerint származtatható egy $(A', B)$ generátor-jelöltből; $\mathcal{F}'$ választása miatt tehát $(A, B)$ származtatható valamely $(A_i, B_i) \in \mathcal{F}'$ -ből. Leirva ezt a származtatást, az abban szereplő $(A^i, B^i): i = 1, \dots m$ sorozat tagjai helyett nyilvánvalóan választhatók $(A^{i'}, B^{i'})$ egyszerü függések; persze $B^{i'} \subsetneq B$; igy $(A^{i'}, B^{i'})$ levezethető $\mathcal{F}'$ -ből és igy $(A, B)$ is levezethető $\mathcal{F}'$ -ből, ami ellentmondás $\square$

Láttuk korábban, hogy ha $\mathcal{F}$ teljes család, akkor $\mathcal{F}' \subseteq \hat{\mathcal{F}}$ minimális generátorrendszere $\mathcal{F}$ -nek akkor és csak akkor, ha $\{B : (A, B) \in \mathcal{F}'\}$ minimális generátorrendszere $\mathcal{B}$ -nek. Ebből és a 2.7 Tételből triviálisan következik, hogy $\mathcal{J}$ -vel jelölve a $\mathcal{B}$ minimális generátorrendszerét, a származtathatóság ekvivalenciaosztályai $\{(A, B): (A, B) \in \mathcal{F}\}$ alakúak valamely $B \in \mathcal{J}$ -re.

Jegyezzük meg azt is, hogy a 2.5 Tétel és 2.7 Tétel bizonyitásainak ugyanaz a gondolata, mint annak bizonyitásáé, hogy egy metszetfélhálót a metszetirreducibilis elemei generálnak.

A 2.7 Tétel értelme az, hogy logikai jellemzését adja teljes családok minimális számosságú generátorrendszereinek; mig a 2.2 Def. előtti megjegyzésben kombinatorikusan jellemeztük azokat.

Legyen $\mathcal{F}$ teljescsalád, $\mathcal{B}$ az $\mathcal{F}$ maximális jobboldalainak rendszere, $\mathcal{S}$ a $\mathcal{B}$ maximális elemeinek halmaza; azaz $\mathcal{S} = \{B \in \mathcal{B} : (\forall B' \in \mathcal{B})(B' \supseteq B \Rightarrow B = B')\}$. Ekkor $\mathcal{S}$ Sperner-rendszer. Könnyü végiggondolni, hogy ekkor $\mathcal{F}$ kulcsrendszere, $\mathcal{C}$ azon $K \subseteq \Omega$-k halmaza, melyek minimálisak arra a tulajdonságra, hogy egyetlen $B \in \mathcal{S}$ sem tartalmazza őkat, azaz

$$\mathcal{C} = \{K \subseteq \Omega : (\forall B \in \mathcal{S})(K \nsubseteq B) \text{ és } (\forall K' \subsetneq K)(\exists B \in \mathcal{S})(K' \subseteq B)\}$$

Forditva, ha $\mathcal{C}$ az $\mathcal{F}$ kulcsrendszere, akkor

$$\{B \subseteq \Omega : (\forall K \in \mathcal{C})(K \nsubseteq B) \text{ és } (\forall B' \supsetneq B)(\exists K \in \mathcal{C})(K \subseteq B')\} = \mathcal{S}$$

az $\mathcal{F}$ maximális jobboldalai közül a tartalmazásra nézve maximálisak halmaza.

## Teljes családok és Sperner rendszerek relációval való reprezentálása

Természetes kérdés [11], hogy adott kulcsrendszert illetve teljes családot hány soros relációval lehet reprezentálni? Legyen $r(n)$ az a minimális szám minden n-re, hogy ha $S \subseteq P(\Omega)$ Sperner-rendszer, $|\Omega| = n$ ; akkor létezik R reláció $\Omega$ felett, melynek legfeljebb $r(n)$ sok sora van és amelyre $\mathcal{F}_R$ kulcsrendszere éppen $S$ . Jelölje $R(n)$ azt a legkisebb számot, amelyre igaz a következő; ha $\mathcal{F} \subseteq P(\Omega) \times P(\Omega)$ teljes család, $|\Omega| = n$, akkor létezik egy legfeljebb $R(n)$ sok sorú R reláció $\Omega$ felett, amelyre $\mathcal{F} = \mathcal{F}_R$ .

2.8 Tétel (i) $r(n) \leq \binom{n}{[\frac{n}{2}]} + 1$

(ii) $R(n) \leq 2\binom{n}{[\frac{n}{2}]} + 1$

Bizonyitás i/ ha $S \subseteq P(\Omega)$ Sperner-rendszer, akkor

$$\{B \subseteq \Omega : (\forall S \in S)(S \nsubseteq B) \text{ és } (\forall B' \supsetneq B)(\exists S \in S)(B' \supseteq S)\} = S'$$

Sperner-rendszer;

$S' = \{B_i : i = 1, \dots \ell\}$ elemeire végezve az 1.3 Tétel bizonyitása szerinti konstrukciót, a kapott R relációnak éppen $S$ a kulcsrendszere. R-nek $|S'| + 1$ sora van: $S'$ nyilván Sperner rendszer, igy $|S'| \leq \binom{n}{[\frac{n}{2}]}$, tehát R-nek legfeljebb $\binom{n}{[\frac{n}{2}]} + 1$ sora van.

ii/ ha $\mathcal{F}$ teljes család $\Omega$ felett, akkor maximális jobboldalrendszerének létezik legfeljebb $2\cdot\binom{n}{[\frac{n}{2}]}$ számosságú generátorrendszere. Ezután 1.3 Tétel bizonyitását végezve az állitás adódik $\square$

<u>2.9 Tétel</u> i/ $\frac{1}{n}\cdot\binom{n}{[n/2]} \leq r(n)\cdot \log r(n)$

ii/ $\frac{1}{n}\cdot\left(1+\frac{1}{n}\right)\binom{n}{[n/2]} \leq R(n)\cdot \log R(n)$

<u>Bizonyitás</u> i/ a különböző Sperner-rendszerek száma legalább $2^{\binom{n}{[n/2]}}$. Ha egy $\mathcal{F}$ teljes családot legfeljebb r-soros relációval lehet reprezentálni, akkor lehet olyan relációval is, melynek pontosan r sora van és amely legfeljebb r+1 sok jelet használ. $(r \times n)$-es, legfeljebb (r+1) jelet használó mátrix $(r+1)^{r\cdot n}$ sok van. Igy

$$(r(n)+1)^{r(n)\cdot n} \geq 2^{\binom{n}{[n/2]}}, \quad \text{igy}$$

$$\frac{1}{n}\cdot\binom{n}{[n/2]} \leq r(n)\cdot \log r(n)$$

ii/ láttuk /ld. 24. oldal/, hogy létezik $\left(1+\frac{1}{n}\right)\binom{n}{[n/2]}$ számosságú metszettulajdonságú rendszer. Igy, (i) bizonyitásának mintájára, $2^{\binom{n}{[n/2]}}$ helyett $2^{(1+1/n)\binom{n}{[n/2]}}$ -t használva, adódik az állitás

$\square$

## Irodalom

[1] W.W. Armstrong: "Dependency Structures of Data Base Relationships",
Information Processing 74, 580-583
North-Holland Publishing Company, Amsterdam, 1974.

[2] W.W. Armstrong: "On the Generation of Dependendy Structures of Relational Data Bases",
Publication # 272,
Université de Montréal

[3] Békéssy András - Demetrovics János: "Contribution to the theory of data base relations",
Discrete Mathematics, 27, /1979/, 1-1o.

[4] Békéssy András - Demetrovics János - Hannák László - Katona Gyula - Frankl Péter: "On the number of maximal dependencies in a data relation of fixed order",
Discrete Mathematics 3o, /1980/, 83-88.

[5] E.F. Codd: "A Relational Model of Data for Large Shared Data Banks",
Communications of the ACM,
13, 6, (June 1970), 377-387.

[6] E.F. Codd: "Further Normalization of the Data Base Relational Model", Courant Inst.Comp.Sci. Symp. 6,
Data Base System, Prentice-Hall, Englewood Cliffs, N.J. (1971), 33-64.

[7] Demetrovics János: "Homogén file kulcsairól",
Alkalmazott Matematikai Lapok 3, /1977/, 185-19.

[8] Demetrovics János: "On the equivalence of candidate keys with Sperner systems",
Acta Cybernetica 4, 3, /1979/, 247-252.

[9] Demetrovics János: "Relációs adatbázis modell",
MTA SZTAKI Közlemények 2o, /1978/, 21-35.

[1o] Demetrovics János: "On the number of candidate keys",
Information Processing Letters 7, 6, /1978/, 266-269.

[11] Demetrovics János: "Candidate keys and antichains"
SIAM Journal on Algebraic and Discrete Methods 1, 1,/1980.

[12] Demetrovics János: "O klucsak odnorodnüh fájlov",
Bolgár Akadémiai Közlemények /198o/.

[13] P. Erdős - Chao Ko - R. Rado: "Intersection theorems for systems of finite sets,
The Quarterly Journal of Mathematics, 12, /1961/, Oxford

[14] E. Sperner: "Eine Satz über Untermengen einer endlichen Menge",
Math. Z. 27, /1928/, 544-548.

[15] C.T. Yu - D.T. Johnson: "On the complexity of finding the set of candidate keys for a given set of functional dependencies",
<u>Information Processing Letters</u>, 5, 4,/October 1976/ loo-lol.

## S u m m a r y

The presented article deals with functional dependencies of the relational data model.

The First Chapter is divided into four parts. The fundamental results of W.W. Armstrong about functional dependencies and full families are shown first and the notions "maximal element" and "maximal right hand side" are defined. Both are important for the subsequent analysis. In the second part of this Chapter the key and candidate key are defined. There is, to every Sperner system S a relation constructed such that S be the set of its candidate keys. Next we deal with linear relations and linear dependencies. It is proved that every functional dependency of a linear relation is a linear dependency, and that the sets of candidate keys of linear relations are exactly the matroids that can be coordinatized over the rational field $\mathbb{Q}$. Last we investigate the following combinatorial problem:

what is the maximum number $N_n$ of maximal elements of a full family over an n-element attribute-set?

It is proved that $N_n$ is assymptotically $2^n$.

The first part of the second Chapter contains a method worked out by the first two authors. By this method it is possible to deduce some relations between the number of initial functional dependencies and that of keys generated by them.

After this we deal with the notion of deductibility. As to this notion the equivalence of defintiions given by J. Demetrovics and W.W. Armstrong is proved. The main problem concerning this notion is the characterization of generating sets of minimal cardinality. We give a combinatorial characterization for these sets in terms of maximal right hand sides and compare this characterization with that of W.W. Armstrong's that is of rather logical nature.

At last, some results on representation of full families and Sperner systems by relations are shown.

A TANULMÁNYOK sorozatban 1979-ben megjelentek:

88/1979 Renner G. - Gaál B. - Hermann Gy. - Horváth L. - Várady T.: Szoborszerü felületek tervezése és megmunkálása

89/1979 Ruda Mihály: A SIS77 statisztikai információs rendszer /a felhasznált számitástechnikai eszközök, a rendszer szerkezete és programjai/

90/1979 Bányász Cs. - Keviczky L.: Optimum Insensitivity of the Linear-continuous Transformation

91/1979 Téli iskola /Szentendre/

92/1979 Bolla M. - Csáki P. - Fischer J. - Herodek S. - Hoffman Gy. - Kutas T. - Telegdi L. - Wittmann I.: A balatoni ökoszisztéma modellezése

93/1979 Andor László: Kisgépes adatbázis kezelő rendszer

94/1979 Gertler János: Egy statisztikus szürési eljárás számitógépes folyamatirányitásához

95/1979 Báthory M. - Galló V. - Kovács E. - Mérő L. - Siegler A. - Vajta L.: Festőrobot vezérlésére alkalmas alafelsimerési berendezés

96/1979 Mérő László: Konturkeresés zajos digitalizát képekben

97/1979 Pásztorné - Matavovszky T.: Boole-függvény kezelőrendszer

98/1979 Kecskés Zsuzsa: Három dimenziós tárgyak drótvázának ábrázolása vonalrajzoló grafikus berendezésekkel

99/1979 Ivics József: KGST Riga

100/1979 Téli iskola

1980-ban jelentek meg:

101/1980 Gerencsér László - Hangos Katalin: Diszkrét lineáris sztochasztikus rendszerek önhangoló szabályozása.

102/1980 Pásztorné Varga Katalin: Rekurziv eljárás

103/1980 Gerencsér Piroska - Szép Endre - Zilahy Ferenc Marton Zsolt: Robotmegfogók adaptivitása I.

104/1980 Knuth Előd - Radó Péter - Tóth Árpád: Az SDLA előzetes ismertetése

105/1980 E. Knuth - P. Radó - Á. Tóth: Preliminary description of SDLA

106/1980 Prékopa András: Sztochasztikus programozási modellek és alkalmazásuk

107/1980 Kelle Péter: Megbizhatósági készletmodellek és alkalmazásai

108/1980 Almásy Gedeon: Mérlegegyenletek és mérési hibák