

Adatbázisok 1.

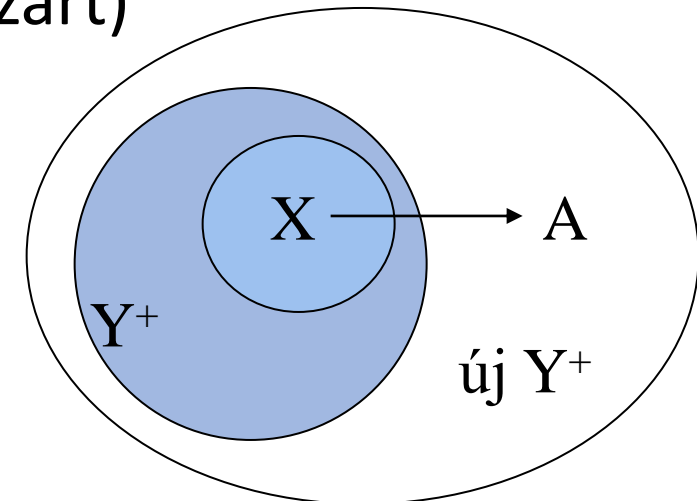
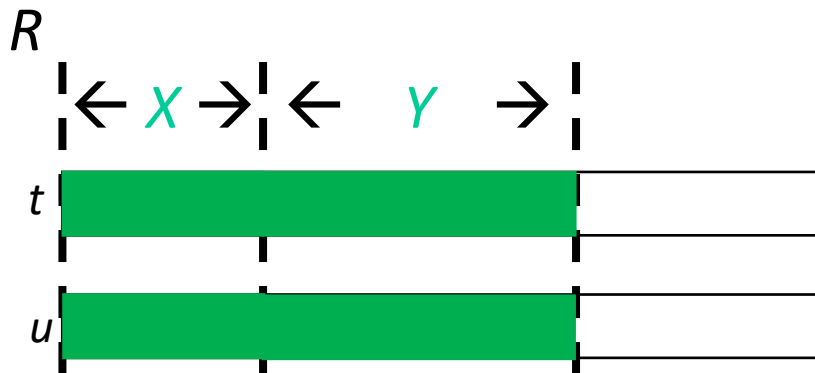
Többértékű függőségek

Negyedik normálforma

Funkcionális és többértékű függőségek következtetése

Ismétlés I.

- $X \rightarrow Y$ egy R relációra vonatkozó megszorítás, miszerint ha két sor megegyezik X összes attribútumán, Y attribútumain is meg kell, hogy egyezzenek.
- Szuperkulcs, kulcs; hogyan kaphatjuk meg?
- Armstrong axiómák, levezetés
- Lezárás (R reláció, F FF halmaz, Y attr. halmaz, Y^+ lezárt)



Ismétlés II.

- Anomáliák
- „Normalizálás”: relációséma több sémára való felbontása ($R(A_1, A_2, \dots, A_n)$ helyett $S(B_1, B_2, \dots, B_m)$ és $T(C_1, C_2, \dots, C_k)$)
- Következmény FF-ek megtalálása, exponenciális algoritmus
- Veszteségmentes felbontás ($r = \Pi_{R_1}(r) \bowtie \dots \bowtie \Pi_{R_k}(r)$ teljesül)
- BCNF, BCNF-re való felbontás
- Chase-teszt veszteségmentességhez
- 3NF

A	B	C	D
a	b ₁	c ₁	d
a	b ₂	c	d ₂
a ₃	b	c	d

Motiváció

Oktató	Város	Közterület	Tárgy	Szemeszter
Kovács István	Budapest	Rákóczi út 1.	Adatbázisok	2020/21/1
Kovács István	Debrecen	Kossuth tér 1.	Adatbázisok	2020/21/1
Kovács István	Budapest	Rákóczi út 1.	Hálózatok	2019/20/2
Kovács István	Debrecen	Kossuth tér 1.	Hálózatok	2019/20/2
Kovács István	Budapest	Rákóczi út 1.	Programozás	2018/19/1
Kovács István	Debrecen	Kossuth tér 1.	Programozás	2018/19/1

A TÉF definíciója

- A *többértékű függőség* (TÉF): az R reláció fölött $X \twoheadrightarrow Y$ teljesül: ha bármely két sorra, amelyek megegyeznek az X minden attribútumán, az Y attribútumaihoz tartozó értékek felcserélhetők, azaz a keletkező két új sor R -beli lesz.
- Más szavakkal: X minden értéke esetén az Y -hoz tartozó értékek függetlenek az R - X - Y értékeitől.
- Az előző példában: Oktató \twoheadrightarrow Város Közterület

Példa: TÉF

Alkesz(név, cím, tel, kedveltSörök)

- Az alkeszek telefonszámai függetlenek az általuk kedvelt söröktől.
 - $\text{név} \rightarrow \text{tel}$ és $\text{név} \rightarrow \text{kedveltSörök}$.
- Így egy-egy alkesz minden telefonszáma minden általa kedvelt sörrel kombinációban áll.
- Ez a jelenség független a funkcionális függőségektől.
 - itt a $\text{név} \rightarrow \text{cím}$ az egyetlen FF.

A név->->tel által implikált sorok

Ha ezek a soraink vannak:

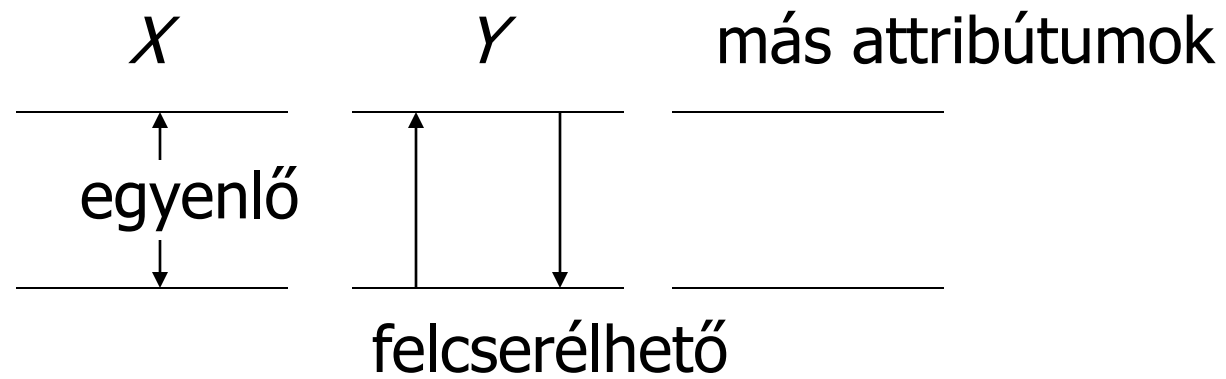
név	cím	tel	kedveltSörök
Sue	a	p1	b1
Sue	a	p2	b2
Sue	a	p2	b1
Sue	a	p1	b2

Akkor ezeknek a soroknak is szerepelnie kell

→ módosítási anomália!

(Pl. p2-t p3-ra módosítjuk, mindenütt meg kell tenni...)

Az $X \twoheadrightarrow Y$ TÉF képe



TÉF szabályok

- Minden FF TÉF.
 - Ha $X \rightarrow Y$ és két sor megegyezik X-en, Y-on is megegyezik, emiatt ha ezeket felcseréljük, az eredeti sorokat kapjuk vissza, azaz: $X \rightarrow\rightarrow Y$.
- *Komplementálás* : Ha $X \rightarrow\rightarrow Y$ és Z jelöli az összes többi attribútum halmazát, akkor $X \rightarrow\rightarrow Z$.

Oktató	Város	Közterület	Tárgy	Szemeszter	Oktató	Város	Közterület	Tárgy	Szemeszter
Kovács István	Budapest	Rákóczi út 1.	Adatbázisok	2020/21/1	Kovács István	Budapest	Rákóczi út 1.	Adatbázisok	2020/21/1
Kovács István	Debrecen	Kossuth tér 1.	Adatbázisok	2020/21/1	Kovács István	Budapest	Rákóczi út 1.	Hálózatok	2019/20/2
Kovács István	Budapest	Rákóczi út 1.	Hálózatok	2019/20/2	Kovács István	Budapest	Rákóczi út 1.	Programozás	2018/19/1
Kovács István	Debrecen	Kossuth tér 1.	Hálózatok	2019/20/2	...				
Kovács István	Budapest	Rákóczi út 1.	Programozás	2018/19/1					
Kovács István	Debrecen	Kossuth tér 1.	Programozás	2018/19/1					
...									

Nem tudunk darabolni

- Ugyanúgy, mint az FF-ek esetében, a baloldalakat nem „bánthatjuk” általában.
- Az FF-ek esetében a jobboldalakat felbonthattuk, míg ebben az esetben ez sem tehető meg.

Példa: többattribútumos jobboldal

Alkesz(név, tTársaság, tel, kedveltSörök, gyártó)

- Egy alkesznek több telefonja lehet, minden számot két részre osztunk: tTársaság (pl. Vodafone) és a maradék hét számjegy.
- Egy alkesz több sört is kedvelhet, mindegyikhez egy-egy gyártó tartozik.

Példa folytatás

- Mivel a tTársaság-tel kombinációk függetlenek a kedveltSörök-gyártó kombinációtól, azt várjuk, hogy a következő TÉF-ek teljesülnek:

név ->-> tTársaság tel

név ->-> kedveltSörök gyártó

Példa adat

Egy lehetséges előfordulás, ami teljesíti az iménti TÉF-et:

név	tTársaság	tel	kedveltS	gyártó
Sue	20	555-1111	Bud	A.B.
Sue	20	555-1111	WickedAle	Pete's
Sue	70	555-9999	Bud	A.B.
Sue	70	555-9999	WickedAle	Pete's

Ugyanakkor sem a **név->->tTársaság** sem a **név->->tel** függőségek nem teljesülnek.

Többértékű függőségek

- Axiomatizálás

Funkcionális függőségek	Többértékű függőségek	Vegyes függőségek
A1 (reflexivitás): $Y \subseteq X$ esetén $X \rightarrow Y$.	A4 (komplementer): $X \rightarrow \rightarrow Y$ és $Z = R - XY$ esetén $X \rightarrow \rightarrow Z$.	A7 (funkcionálisból többértékű): $X \rightarrow Y$ esetén $X \rightarrow \rightarrow Y$.
A2 (bővíthetőség): $X \rightarrow Y$ és tetszőleges Z esetén $XZ \rightarrow YZ$.	A5 (bővíthetőség): $X \rightarrow \rightarrow Y$ és tetszőleges $V \subseteq W$ esetén $XW \rightarrow \rightarrow YV$.	A8 (többértékűből és funkcionálisból funkcionális): $X \rightarrow \rightarrow Y$ és $W \rightarrow S$, ahol $S \subseteq Y$, $W \cap Y = \emptyset$ esetén $X \rightarrow S$.
A3 (transzitivitás): $X \rightarrow Y$ és $Y \rightarrow Z$ esetén $X \rightarrow Z$.	A6 (transzitivitás): $X \rightarrow \rightarrow Y$ és $Y \rightarrow \rightarrow S$ esetén $X \rightarrow \rightarrow S - Y$.	

Negyedik normálforma

- A TÉF-ek okozta redundanciát a BCNF nem szünteti meg.
- A megoldás: a negyedik normálforma!
- A negyedik normálformánál (4NF) amikor dekomponálunk, a TÉF-eket úgy kezeljük, mint az FF-eket, a kulcsok megtalálásánál azonban nem számítanak.

Triviális TÉF-ek

- Legyen R egy reláció, X és Y részhalmazai R attribútumainak, jelölés:
 $X, Y \subseteq R$
- $X \twoheadrightarrow Y$ TÉF automatikusan fennáll, ha $Y \subseteq X$
- $X \twoheadrightarrow Y$ TÉF automatikusan fennáll, ha $X \cup Y = R$

4NF definíció

- Egy R reláció **4NF**-ben van ha: minden $X \twoheadrightarrow Y$ nemtriviális **TÉF** esetén X superkulcs.
- **Nemtriviális TÉF** :
 1. Y nem részhalmaza X -nek,
 2. X és Y együtt nem adják ki az összes attribútumot.
- A superkulcs definíciója ugyanaz marad, azaz csak az FF-ektől függ.

BCNF kontra 4NF

- Kiderült, hogy minden $X \rightarrow Y$ FF
 $X \twoheadrightarrow Y$ TÉF is.
- Így, ha R 4NF-ben van, akkor BCNF-ben is.
 - Mert minden olyan FF, ami megsérti a BCNF-t, a 4NF-t is megsérti.
- De R lehet úgy BCNF-ben, hogy közben nincs 4NF-ben.

Dekompozíció és 4NF

- Ha $X \twoheadrightarrow Y$ megsérti a 4NF-t, akkor R -t hasonlóan dekomponáljuk, mint a BCNF esetén.
 1. XY az egyik dekomponált reláció.
 2. Az $Y - X$ -be nem tartozó attribútumok a másik.

Példa: 4NF dekompozíció

Alkesz(név, cím, tel, kedveltSörök)

FF: név \rightarrow cím

TÉF-ek: név \rightarrow tel

név \rightarrow kedveltSörök

- Kulcs {név, tel, kedveltSörök}.
- Ezért az összes függőség megsérti 4NF-et.

Példa folytatás

- Dekompozíció **név -> cím** szerint:

1. Alkesh1(név, cím)

- Ez 4NF-beli; az egyetlen függőség **név-> cím**.

2. Alkesh2(név, tel, kedveltSörök)

- Nincs 4NF-ben. A **név ->-> tel** és **név ->-> kedveltSörök** függőségek teljesülnek. A három attribútum együtt kulcs (mivel nincs nemtriviális FF).

Példa: Alkesz2 dekompozíciója

- Bármelyik, $név \rightarrow tel$, vagy a $név \rightarrow kedveltSörök$ TÉF szerinti dekompozíció ugyanazt eredményezi:
 - $Alkesz3(\underline{név}, \underline{tel})$
 - $Alkesz4(\underline{név}, \underline{kedveltSörök})$

TÉF és FF-ek együttes következtetése

- **Probléma:** R relációsémához adott a TÉF-ek és FF-ek egy halmaza, kérdés: egy adott FF vagy TÉF következik-e ezekből R fölött?
- **Megoldás:** használjunk egy táblázatot (tablót), hogy a függőségek hatásait feltárjuk. (A chase mögötti ötletet terjesztjük ki.)

Miért foglalkozunk ilyesmivel egyáltalán?

1. 4NF azon múlik, hogy van-e olyan TÉF, ami sérti a feltételt.
 - Előfordulhat, hogy a megadott FF-ek és TÉF-ek nem sértik a feltételt, de egy belőlük következő függőség igen.
2. Amikor dekomponálunk az FF-eket és TÉF-eket is vetítenünk kell.

Példa: CHASE TÉF-ek és FF-ek esetére

- Az FF-ek esetén ugyanúgy tegyük egyenlővé a szimbólumokat, mint korábban.
- Egy TÉF esetén írjuk be azokat a sorokat, melyek szükségesek ahhoz, hogy az előfordulás ne sértse meg a TÉF-et.
- $X \rightarrow - \rightarrow Y$: ha van két sor a tablóban, amelyek megegyeznek X -en \rightarrow készíthetünk 2 újabb sort, megcserélve Y -on elhelyezkedő komponenseiket

Példa: CHASE TÉF-ek és FF-ek esetére

- A 2 új sornak a relációban szerepelnie kell → a tablóban is
- Ha FF-ekből és TÉF-ekből szeretnénk levezetni egy $X \rightarrow Y$, akkor 2 soros tablóval kezdünk, amelyek X -en megegyeznek a többinél különböznek
- A fentieket alkalmazzuk; ha észrevesszük, hogy az eredeti sorok egyikében az Y attr.-okat kicseréljük egy másik eredeti sorból ugyanazokkal → beláttuk a függőséget

Példa: CHASE TÉF-ek és FF-ek esetére

- Kiindulásként (TÉF): legyen az első sor olyan, hogy nem indexelt betűket tartalmaz **X**-en és **Y**-on, a második pedig ugyanilyeneket **X**-en, és azonkívül a nem **Y**-belieken.
- A két sorban fennmaradó helyeken új, egyszer szereplő szimbólumok legyenek
- Kérdés: előfordul-e az a sor a tablóban, amelynek minden eleme indexeletlen?

A tabló $A \rightarrow C$ bizonyítása

□ **Példa:** ha $A \rightarrow BC$ és $D \rightarrow C$, akkor
 $A \rightarrow C$ is teljesül minden esetben.

Cél: bizonyítani, hogy $c_1 = c_2$.

A	B	C	D
a	b1	c1	d1
a	b2	c1	d2
a	b2	c1	d1
a	b1	c1	d2

$A \rightarrow BC$ használata.

$D \rightarrow C$ -t használjuk.

Példa: tranzitivitás TÉF-ek esetén

- Ha $A \rightarrow B$ és $B \rightarrow C$, akkor $A \rightarrow C$?
 - Ha a séma ABC , akkor a komplementálási szabályból ez valóban következik.
 - A példában feltesszük hogy a séma: $ABCD$, és be fogjuk látni, hogy ott is igaz.

A tábló $A \rightarrow - \rightarrow C$ esetén

Cél: megjelenjen az (a, b, c, d) sor.

A	B	C	D
a	b1	c	d1
a	b	c1	d
a	b	c	d1
a	b1	c1	d
a	b	c1	d1
a	b	c	d
a	b1	c1	d1
a	b1	c	d

$A \rightarrow - \rightarrow B$
használata.

$B \rightarrow - \rightarrow C$
használata.

Következtetés: FF használata

- FF $X \rightarrow Y$ alkalmazásánál keressük meg azon sorpárokat, amelyek megegyeznek X attribútumain. Az Y attribútumain is tegyük őket egyenlővé.
 - Egy változót egy másikra cseréljük.
 - Ha a lecserélt változó a célsorban is megjelenik, ott is cseréljük.

Következtés: TÉF használata

- Egy $X \rightarrow Y$ TÉF használatánál keressünk két sort, amelyek megegyeznek X attribútumain.
 - Adjuk hozzá a táblához azokat a sorokat, amelyeket az Y attribútumaihoz tartozó értékek felcserélésével kapunk.

Következtetés: célok

- Az $U \rightarrow V$ ellenőrzésekor akkor nyertünk, ha a megfelelő változók V -hez tartozó minden oszlopban egyenlőek.
- $U \rightarrow \neg V$ akkor győztünk, ha sikerül egy olyan sort kigenerálni, ami az eredeti két sorból keletkezik V értékeinek felcserélésével.

Következtetés: Végjáték

- Használjuk az összes FF-et és TÉF-et, amíg bármiféle változtatás történhet.
- Ha nyertünk, nyertünk.
- Ha nem, egy ellenpéldát kaptunk.
 - A kapott előfordulás az összes előre megadott függőséget teljesíti.
 - Az eredeti két sor megsérti a kikövetkeztetendő függőséget.

TÉF-ek vetítése

- Le kell tudnunk vetíteni megadott függőségeket 2 reláció sémára
- Legrosszabb eset: ki kell próbálnunk minden lehetséges FF-et és TÉF-et a felbontott relációkra
- Chase teszt alkalmazása; Cél egy TÉF ellenőrzésénél: olyan sor előállítása a tablóban, amely indexeletlen betűket tartalmaz a felbontott reláció oszlopaira

Példa: vetítés

- Példa: adott $R(A,B,C,D,E) \rightarrow$ felbontunk
- Egyik létrejövő reláció: $S(A,B,C)$
- Tfh. $A \twoheadrightarrow CD$ R-ben fennáll
- Cél: bizonyítani, hogy $A \twoheadrightarrow C$ fennáll S-ben

A	B	C	D	E
a	b1	c	d1	e1
a	b	c2	d	e
a	b1	c2	d	e1
a	b	c	d1	e



$A \twoheadrightarrow CD$ használata.