

10.előadás: Adatbázisok

dr. Hajas Csilla (ELTE IK)

<http://sila.hajas.elte.hu/>

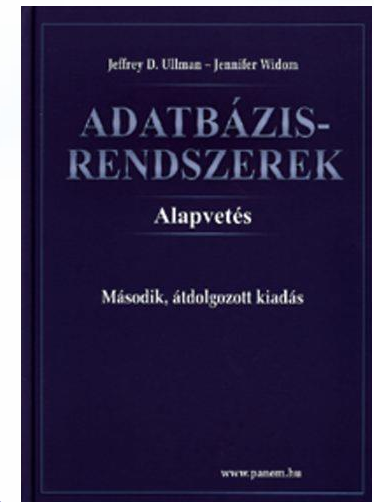
Relációs adatbázis sématervezés

A mai témakörök a Tankönyvben:

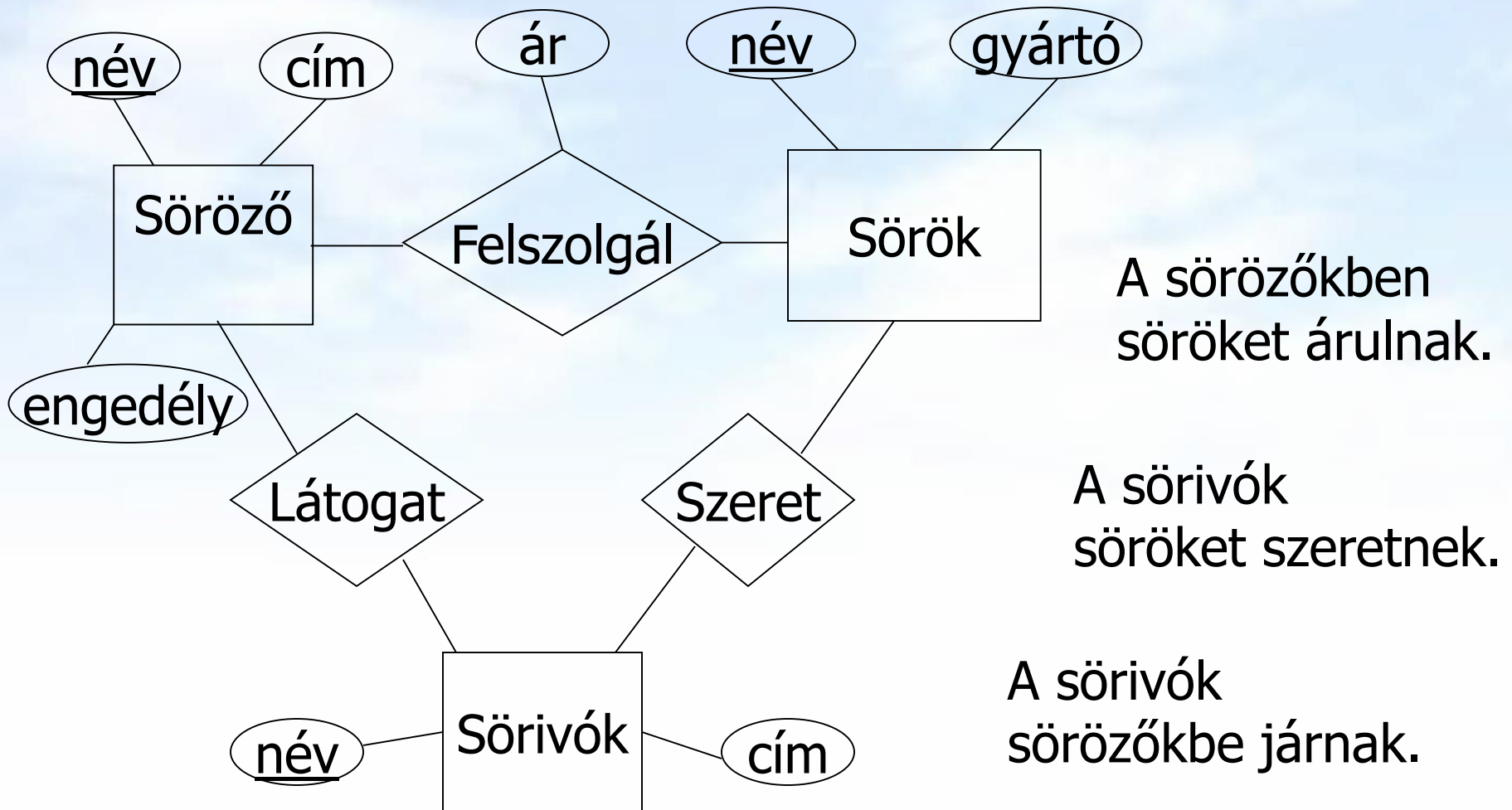
3.3.1. Bevezetés: Relációk felbontása
(motiváció, normalizálás)

3.1. Funkcionális függőségek

3.2. Funkcionális függőségekre vonatkozó szabályok: az Armstrong-axiómák, Attribútumhalmaz lezártjának kiszámítása



Emlékeztető: Az előadás példája (E/K)



Példa--1: E/K diagram átírása relációkká

Az egyedhalmazok és kapcsolatok átírása

- Egyedhalmazok átírása:
(aláhúzás jelöli a kulcs attribútumokat)

Sörök(sör, gyártó)

Sörözők(söröző, város, tulaj, engedély)

Sörivők(név, város, tel)

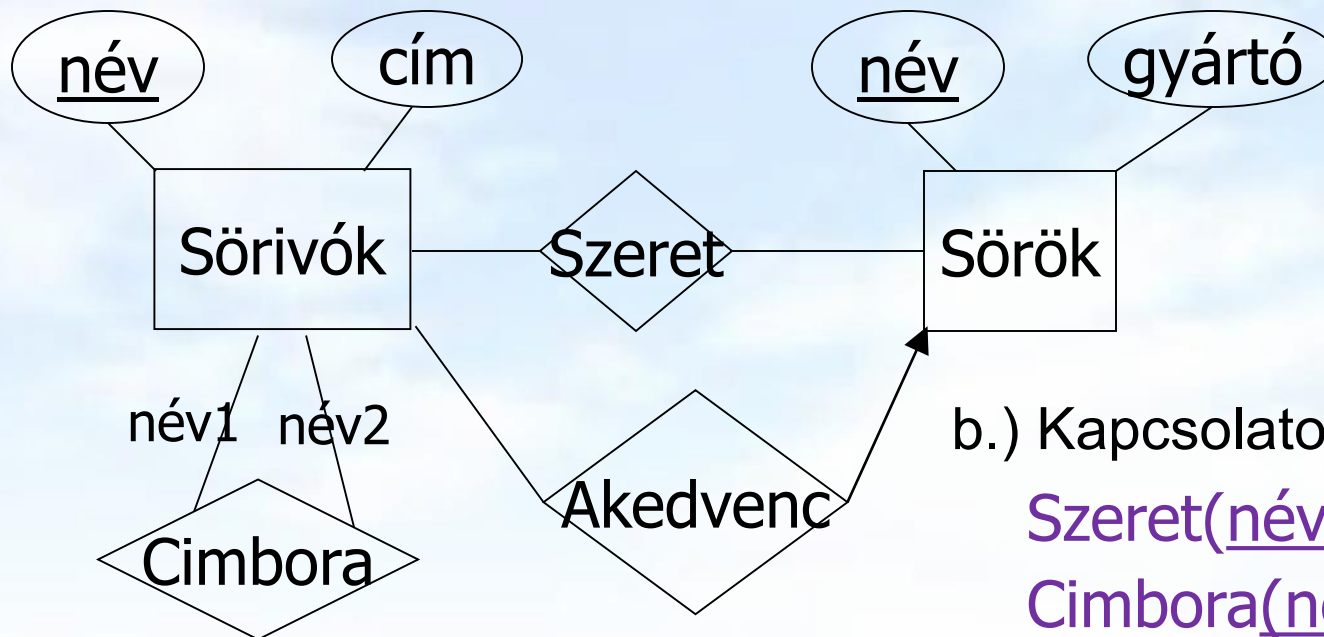
- Kapcsolatok átírása:

Szeret(név, sör)

Felhasznál(söröző, sör, ár)

Látogat(név, söröző)

Példa--2: E/K diagram átírása relációkká sok-egy kapcsolatok összevonása



a.) Egyedhalmazok átírása:

Sörök(sör, gyártó)
* Sörivók(név, cím)

b.) Kapcsolatok átírása:

Szeret(név, sör)
Cimbora(név1, név2)
* Akedvenc(név, sör)

c.) Összevonás: Sörivók + Akedvenc
Sörivók(név, cím, aKedvencSör)

Tk.3.1. Motiváció: Bevezető példa

- Mi lenne, ha a Sörivók és a Sörök, meg a köztük lévő Szeret illetve aKedvencSör kapcsolatokat az eddigi több tábla helyett egy táblába egyesítve tárolnánk?

Sörivó(név, cím, sör, gyártó, aKedvencSör)

név	cím	sör	gyártó	aKedvenc
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	Voyager	WickedAle	Pete's	WickedAle
Spock	Enterprise	Bud	A.B.	Bud

- **Redundancia**, az információk feleslegesen ismétlődnek több sorban, hiszen mindenkinek csak egy lakcíme és egy aKedvencSöre lehet, továbbá a söröknek is csak egy gyártója lehet!

Hibás tervezés problémái

A rosszul tervezettség anomáliákat is eredményez
Sörivó(név, cím, sör, gyártó, aKedvencSör)

név	cím	sör	gyártó	aKedvenc
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	Voyager	WickedAle	Pete's	WickedAle
Spock	Enterprise	Bud	A.B.	Bud

- **Módosítási anomália:** ha az A.B. gyártót B.A.-ra módosítjuk, megteesszük-e ezt minden sornál?
- **Törlési anomália:** Ha senki sem szereti a Bud sört, azzal töröljük azt az infót is, hogy ki gyártotta.
- **Beillesztési anomália:** és felvinni ilyen gyártót?

Relációs sémák tervezése

- Cél: redundancia csökkentése és az anomáliák megszüntetése.
 - **Redundancia**: az adatok felesleges ismétlődése
 - **Módosítási anomália**: egy adat egy előfordulását megváltoztatjuk, más előfordulásait azonban nem.
 - **Törlési anomália**: törléskor olyan adatot is elveszítünk, amit nem szeretnénk.
 - **Beillesztési anomália** : nem tudunk felvinni adatsorokat

Normalizálás (a középiskolákban)

- Sörivő(név, cím, sör, gyártó, aKedvencSör) ez 1NF, de még nem 2NF, a részleges függést kell megszüntetni, hogy csökkentsük a redundanciát.
- Dolgozó(dkod, dnév, fizetés, foglalkozás, belépés, oazon, onév, telephely) ez már 2NF, de nem 3NF, a tranzitív függést kell megszüntetnünk, hogy csökkentsük a redundanciát, és kiküszöböljünk minden anomáliát.

Relációk felbontása (dekomponálás)

A fenti problémáktól dekomponálással (felbontással) tudunk megszabadulni:

➤ **Definíció:**

$d=\{R_1, \dots, R_k\}$ az (R, F) **dekompozíciója**, ha nem marad ki attribútum, azaz $R_1 \cup \dots \cup R_k = R$.
Az adattábla felbontását projekcióval végezzük.

➤ **Példa:**

$R=ABCDE$, $d=\{AD, BCE, ABE\}$

3 tagú dekompozíció, ahol

$R_1=AD$, $R_2=BCE$, $R_3=ABE$,

Felbontásra vonatkozó elvárások

➤ Elvárások

(1) **A vetületek** legyenek jó tulajdonságúak, és a vetületi függőségi rendszere egyszerű legyen (normálformák: BCNF, 3NF, 4NF, később)

(2) **A felbontás** is jó tulajdonságú legyen, vagyis ne legyen információvesztés:

Veszteségmentes legyen a felbontás (VM)

(3) **Függőségek megőrzése** a vetületekben (FŐ)

➤ **Tételek** (ezekre nézünk majd algoritmusokat)

➤ Mindig van **VM BCNF-ra való felbontás**

➤ Mindig van **VM FŐ 3NF-ra való felbontás**

Relációs sématervezés (vázlat)

- **I. Függőségek:** a sématervezésnél használjuk
 - **Funkcionális függőség (10ea)**
 - **Többértékű függőség (12ea)**
- **II. Normalizálás:** „jó” sémákra való felbontás
 - **Funkcionális függ. -> BCNF (11ea)**
 - **Funkcionális függ. -> 3NF (12ea)**
 - **Többértékű függ. -> 4NF (12ea)**
- **III. Felbontás tulajdonságai:** „jó” tulajdonságok
 - **Veszteségmentesség (11ea)**
 - **Függőségörzés (11ea)**

(Tk.3.1.) Funkcionális függőségek

- $X \rightarrow Y$ az R relációra vonatkozó megszorítás, miszerint ha R bármely előfordulásában bármely két sor megegyezik az X összes attribútumán, akkor Y attribútumain is meg kell egyezniük.
- **Jelölés:** X, Y, Z, \dots attribútum halmazokat; A, B, C, \dots attribútumokat jelöl.
- **Jelölés:** $\{A, B, C\}$ attribútum halmaz helyett ABC -t írunk.

Funkcionális függőségek (definíció)

Definíció. Legyen $R(U)$ egy relációséma, továbbá X és Y az U attribútum-halmaz részhalmazai.

X -től funkcionálisan függ Y (jelölésben $X \rightarrow Y$), ha bármely R séma feletti r előfordulás esetén ha két sor megegyezik X -en, akkor megegyezik Y -on is $\forall t1, t2 \in T$ esetén $(t1[X]=t2[X] \Rightarrow t1[Y]=t2[Y])$.

Ez lényegében azt jelenti, hogy az X -beli attribútumok értéke egyértelműen meghatározza az Y -beli attribútumok értékét.

- **Jelölés:** $R \models X \rightarrow Y$ vagyis R kielégíti $X \rightarrow Y$ függőséget

Példa: Funkcionális függőség

Sörivők(név, cím, sör, gyártó, aKedvencSör)

Feltehetjük például, hogy az alábbi FF-ek teljesülnek:

név	cím	sör	gyártó	aKedvencS
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	Voyager	WickedAle	Pete's	WickedAle
Spock	Enterprise	Bud	A.B.	Bud

mert név -> cím

mert név -> aKedvencSör

mert sör -> gyártó

Jobboldalak szétvághatósága és összevonhatósága (FF)

- $X \rightarrow A_1 A_2 \dots A_n$ akkor és csak akkor teljesül R -en, ha $X \rightarrow A_1, X \rightarrow A_2, \dots, X \rightarrow A_n$ is teljesül R -en.
- Példa: $A \rightarrow BC$ ekvivalens $A \rightarrow B$ és $A \rightarrow C$ függőségek kettősével, az előző példában $név \rightarrow \{cím, aKedvencSör\}$ ekvivalens a $név \rightarrow sör$ és $név \rightarrow aKedvencSör$ f.f.-kel
- Megj.: Baloldalak szétvágására nincs szabály, például $\{időpont, terem\} \rightarrow kurzus$ teljesül, de sem az $időpont \rightarrow kurzus$ és sem a $terem \rightarrow kurzus$ nem teljesül.

Kulcs, superkulcs

- Funkcionális függőség $X \rightarrow Y$ speciális esetben, ha $Y = U$, ez a kulcsfüggőség.
- $R(U)$ relációséma esetén az U attribútum-halmaz egy K részhalmaza akkor és csak akkor **superkulcs**, ha a $K \rightarrow U$ FF teljesül.
- A kulcsot tehát a függőség fogalma alapján is lehet definiálni: **olyan K attribútum-halmazt nevezünk kulcsnak, amelytől az összes többi attribútum függ (vagyis superkulcs), de K -ból bármely attribútumot elhagyva ez már nem teljesül (vagyis minimális superkulcs)**

Példa: superkulcs, kulcs

Sörivők(név, cím, sör, gyártó, aKedvencSör)

- {név, sör} **superkulcs**, ez a két attr. meghatározza funkcionálisan a maradék attr-kat.

név -> cím aKedvencSör

sör -> gyártó

- {név, sör} **kulcs**, hiszen sem {név}, sem {sör} nem superkulcs.

név -> gyártó; sör -> cím nem teljesül

- Az előbbi kívül nincs több kulcs, de számos superkulcs megadható (ami ezt tartalmazza)

Példa: több kulcs is lehet

- **Példa:** Autók egyedhalmazt leíró tulajdonságok közül több kulcs is van, és ebből egy elsődleges kulcsot választunk ki, például egyszerű kulcs az alvázszám (ez is egyértelműen azonosítja), de elsődleges kulcs itt legyen egy összetett kulcs: rendszerszám, ország (ez is minimális superkulcs).
- **Másik példa:** Legyen ABC relációs émán def.funkcionális függőségek $AB \rightarrow C$ és $C \rightarrow B$.
 - Példa: A = utca, B = város, C = irányítószám.
 - Itt két kulcs is van: $\{A, B\}$ és $\{A, C\}$.

(Tk.3.2.) Az implikációs probléma

Funkcionális függőségekre vonatkozó szabályok, funkciófüggőségek levezetése

- Legyenek $X_1 \rightarrow A_1, X_2 \rightarrow A_2, \dots, X_n \rightarrow A_n$ adott FF-k, szeretnénk tudni, hogy $Y \rightarrow B$ teljesül-e olyan relációkra, amire az előbbi FF-k teljesülnek.
- Példa: $A \rightarrow B$ és $B \rightarrow C$ teljesülése esetén $A \rightarrow C$ biztosan teljesül.
- Implikációs probléma eldöntése definíció alapján (minden előfordulásra ellenőrizni) lehetetlen, de van egyszerűbb lehetőség: levezetési szabályok (ún. Armstrong-axiómák) segítségével előállítani.

Armstrong-axiómák

Legyen $R(U)$ relációséma és $X, Y \subseteq U$, és jelölje XY az X és Y attribútum-halmazok egyesítését.
 F legyen funkcionális függőségek tetsz. halmaza.

Armstrong axiómák:

- **A1 (reflexivitás):** $Y \subseteq X$ esetén $X \rightarrow Y$ (ezeket hívjuk **triviális funkcionális függőségeknek**, mert minden esetben minden sémán teljesülnek).
- **A2 (bővíthetőség):** $X \rightarrow Y$ és tetszőleges Z esetén
$$XZ \rightarrow YZ$$
- **A3 (tranzitivitás):** $X \rightarrow Y$ és $Y \rightarrow Z$ esetén $X \rightarrow Z$

Levezetés fogalma

- F implikálja $X \rightarrow Y$ -t (F -nek következménye $X \rightarrow Y$), ha minden olyan táblában, amelyben F összes függősége teljesül, azokra $X \rightarrow Y$ is teljesül.

Jelölés: $F \models X \rightarrow Y$, ha F implikálja $X \rightarrow Y$ -et.

- $X \rightarrow Y$ levezethető F -ből, ha van olyan $X_1 \rightarrow Y_1, \dots, X_k \rightarrow Y_k, \dots, X \rightarrow Y$ véges levezetés, hogy $\forall k$ -ra $X_k \rightarrow Y_k \in F$ vagy $X_k \rightarrow Y_k$ az $A1, A2, A3$ Armstrong axiómák alapján kapható a levezetésben előtte szereplő függőségekből.

Jelölés: $F \vdash X \rightarrow Y$, ha $X \rightarrow Y$ levezethető F -ből

További levezethető szabályok:

4. Szétvághatósági (vagy felbontási) szabály

$F \vdash X \rightarrow Y$ és $Z \subseteq Y$ esetén $F \vdash X \rightarrow Z$.

5. Összevonhatósági (vagy unió) szabály

$F \vdash X \rightarrow Y$ és $F \vdash X \rightarrow Z$ esetén $F \vdash X \rightarrow YZ$.

6. Pszeudotranzitivitás

$F \vdash X \rightarrow Y$ és $F \vdash WY \rightarrow Z$ esetén $F \vdash XW \rightarrow Z$.

Bizonyítás (4): Reflexivitási axióma miatt $F \vdash Y \rightarrow Z$, és tranzitivitási axióma miatt $F \vdash X \rightarrow Z$.

Bizonyítás (5): Bővítési axióma miatt $F \vdash XX \rightarrow YX$ és $F \vdash YX \rightarrow YZ$, és $XX = X$, valamint a tranzitivitási axióma miatt $F \vdash X \rightarrow YZ$.

Bizonyítás (6): Bővítési axióma miatt $F \vdash XW \rightarrow YW$, és $YW = WY$, és a tranzitivitási axióma miatt $F \vdash XW \rightarrow Z$.

Szétvághatóság / összevonhatóság

- A szétvághatósági és összevonhatósági szabályok következménye:

$$F \vdash X \rightarrow Y \Leftrightarrow \forall A_i \in Y \text{ esetén } F \vdash X \rightarrow A_i$$

- A következmény miatt feltehető, hogy a függőségek jobb oldalai 1 attribútumból állnak.
- **Fontos!** A függőségeknek csak a jobboldalát lehet szétbontani, a baloldalra ez természetesen nem igaz (például {filmcím, év} → stúdió)

Armstrong-axiómák (tétel)

- **TÉTEL:** Az Armstrong-axiómarendszer helyes és teljes, azaz minden levezethető függőség implikálódik is, illetve azok a függőségek, amelyeket F implikál azok levezethetők F -ből.

$$F \vdash X \rightarrow Y \Leftrightarrow F \models X \rightarrow Y$$

Implikáció eldöntése --- Lezárással

- Implikációs probléma:

Legyenek $X_1 \rightarrow A_1, X_2 \rightarrow A_2, \dots, X_n \rightarrow A_n$ adott FF-k, szeretnénk tudni, hogy $Y \rightarrow B$ teljesül-e minden olyan relációkra, amire az előbbi FF-k teljesülnek. Hogyan tudjuk ellenőrizni, hogy egy előfordulás nem teljesíti $Y \rightarrow B$?

- Mivel az Armstrong axiómarendszer helyes és teljes, elegendő a levezetési szabályokkal levezetni. Még a levezetési szabályoknál is van egyszerűbb út: **kiszámítjuk Y lezártját: Y^{+} -t**
- Attribútum-halmaz lezárására teszt:

Attribútumhalmaz lezártja (definíció)

- Adott R séma és F funkcionális függőségek halmaza mellett, X^+ az összes olyan A attribútum halmaza, amire $X \rightarrow A$ következik F -ből.
- (R, F) séma esetén legyen $X \subseteq R$.
- **Definíció:** $X^{+(F)} := \{A \mid F \vdash X \rightarrow A\}$
az X attribútum-halmaz lezárása F -re nézve.

Attribútumhalmaz lezártja (lemma)

➤ LEMMA: $F \mid\!\!\! \dashv\!\!\! \rightarrow X \rightarrow Y \Leftrightarrow Y \subseteq X^+$.

Bizonyítás:

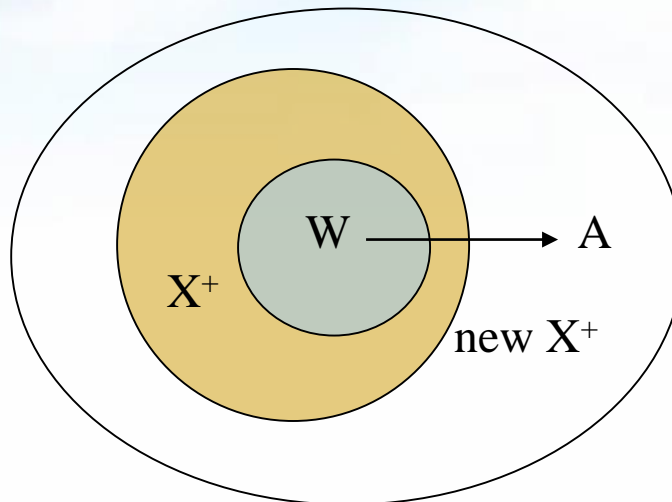
(\Rightarrow) $\forall A \in Y$ esetén a reflexivitás és tranzitivitás miatt $F \mid\!\!\! \dashv\!\!\! \rightarrow X \rightarrow A$, azaz $Y \subseteq X^+$.

(\Leftarrow) $\forall A \in Y \subseteq X^+$ esetén $F \mid\!\!\! \dashv\!\!\! \rightarrow X \rightarrow A$, és az egyesítési szabály miatt $F \mid\!\!\! \dashv\!\!\! \rightarrow X \rightarrow Y$.

➤ Lemma következménye: az implikációs probléma megoldásához elég az X^+ -t hatékonyan kiszámolni.

Lezárás (az algoritmus vázlat)

- **Kiindulás:** $X^+ = X$
- **Indukció:** Olyan FF-ket keresünk, melyeknek a baloldala már benne van X^+ -ban. Ha $W \rightarrow A$ ilyen, A -t hozzáadjuk X^+ -hoz.
- **Output:** ha már nem bővül, ez a halmaz az X^+



Algoritmus X^+ attr.halmaz lezártja

➤ **Kiindulás:** $X^+ = X$

➤ **Algoritmus X^+ kiszámítására:**

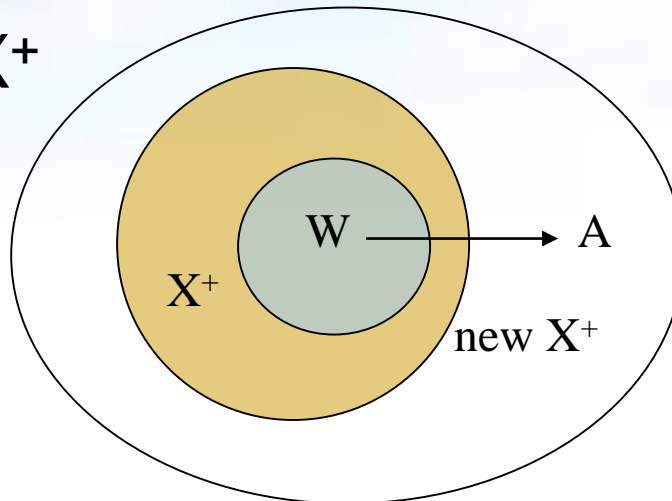
/* Iteráció, amíg $X(n)$ változik */

$X(0) := X$

$X(n+1) := X(n) \cup \{A \mid \underline{W \rightarrow Z} \in F, A \in Z, W \subseteq X(n)\}$

Ha $X(v+1) = X(v)$, akkor Output: $X(v) = X^+$.

➤ **Output: X^+**



Példa: Attribútumhalmaz lezárása

$R=ABCDEFG$, $\{AB \rightarrow C, B \rightarrow G, CD \rightarrow EG, BG \rightarrow E\}$

$X=ABF$, $X^+=?$

$X(0):=ABF$

$X(1):=ABF \cup \{C, G\} = ABCFG$

$X(2):=ABCFG \cup \{C, G, E\} = ABCEFG$

$X(3):=ABCEFG$

$X^+ = ABCEFG$

A lezárást kiszámító algoritmus „helyes”

- Miért működik az X^+ lezárási algoritmus?
(Tankönyv 3.2.5. szakasz, 81-83.oldal)
- Az algoritmus „tényleg” X^+ -t számítja ki. Vagyis:
 1. Ha az A attribútum valamely j -re belekerül X^j -be, akkor A valóban eleme X^+ -nak.
 2. Másfelől, ha $A \in X^+$, akkor létezik olyan j , amire A belekerül X^j -be.
- Az első állítás: Miért csak az igaz funkcionális függőségeket fogadja el a lezárási algoritmus? Ezt könnyen be lehet bizonyítani indukcióval.

A lezárást kiszámító algoritmus „teljes”

- A második állítás: Miért talál meg minden igaz függőséget a lezárási algoritmus?
- Konstruktív bizonyítás: Tegyük fel, hogy $A \in X^+$, és nem olyan j , amire A belekerül X^j -be.

	X^+ elemei	más attribútumok
t	111 ... 111	000 ... 000
s	111 ... 111	111 ... 111

- Ekkor I -re minden F^+ -beli függőség teljesül
- I -re viszont nem teljesül $X \rightarrow A$

Tankönyv 3.5.2. feladata (111.o.)

- **Órarend adatbázis:** Kurzus(K), Oktató(O), Időpont(I), Terem(T), Diák(D), Jegy(J)
- **Feltételek:**
 - Egy kurzust csak egy oktató tarthat: $K \rightarrow O$.
 - Egy helyen egy időben egy kurzus lehet: $IT \rightarrow K$.
 - Egy időben egy tanár csak egy helyen lehet: $IO \rightarrow T$.
 - Egy időben egy diák csak egy helyen lehet: $ID \rightarrow T$.
 - Egy diák egy kurzust egy végső jeggyel zár: $KD \rightarrow J$.
- **R=KOITDJ** $F = \{K \rightarrow O, IT \rightarrow K, IO \rightarrow T, ID \rightarrow T, KD \rightarrow J\}$
- **Feladat:** Határozzuk meg a (R, F) kulcsait az X^+ kiszámítási algoritmus segítségével.

Kérdés / Válasz

- **Köszönöm a figyelmet! Kérdés/Válasz?**
- **Feladatok:** Tankönyv 3.fejezetében az egyes szakaszok után található gyakorló feladatok
- **Mai 10.előadáson:** Feldolgoztuk a Tankönyv 3.1.szakaszát és a 3.2.szakasz nagy részét, de innen két részt később tárgyalunk:
 - 3.2.7.Minimális bázis (3NF-ra felbontásánál)
 - 3.2.8.Függőségek vetülete (BCNF-ra bontásnál)