

Tervezés: Funkcionális függőségek

Tankönyv: Ullman-Widom:
Adatbázisrendszerek Alapvetés
Második, átdolgozott kiadás,
Panem, 2009



-
- 3.1. Funkcionális függőségek,
relációk (szuper)kulcsai
 - 3.2. Funkcionális függőségekre
vonatkozó szabályok,
lezárási algoritmusok,
funkcionális függőségek vetítése

Relációs adatbázisok tervezése

- A tervezésre a BSc-s gyakorlatokon nem jut idő, csak az előadáson, **vizsgán és záróvizsgán** van, ezért itt **célszerű a tankönyvben is átnézni** a tananyagot és a példákat! Tankönyv 3.fejezet.
- **Függőségek:** funkcionális, többértékű, ezeket a tervezésnél használják (az adatbázisrendszerek nem támogatják, ott megszorítások vannak).
- **Normalizálás:** jó sémákra való felbontás, funkcionális függőségek -> (1,2,)3NF, BCNF
többértékű függőségek -> 4NF

Bevezető példa

- Több tábla -> vegyük a táblák összekapcsolását
Melyik séma jobb?

Sörivő(név, cím, kedveltSörök, gyártó, kedvencSör)

név	cím	kedveltSörök	gyártó	kedvencS
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	???	WickedAle	Pete's	???
Spock	Enterprise	Bud	???	Bud

- **Redundáns adat**, a ??? helyén mi szerepel, ha
név -> cím, kedvencSör és kedveltSörök -> gyártó
funkcionális függőségek (def.később) érvényesek

Hibás tervezés problémái

A rosszul tervezettség anomáliákat is eredményez
Sörivő(név, cím, kedveltSörök, gyártó, kedvencSör)

név	cím	kedveltSörök	gyártó	kedvencS
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	???	WickedAle	Pete's	???
Spock	Enterprise	Bud	???	Bud

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

Dekomponálás (felbontás)

- 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).

- Elvárások:

- (1) **Anomáliák kiküszöbölése.** A vetületek legyenek egyszerűek, jó tulajdonságúak (normálformák: BCNF, 3NF)
- (2) **Információ-visszaállíthatóság**, vagyis ne legyen információvesztés, veszteségmentes felbontások
- (3) **Függőségek megőrzése a vetületekben**

Funkcionális függőségek (bevezetés)

- $X \rightarrow Y$ az R relációra vonatkozó megszorítás, miszerint ha két sor megegyezik X összes attribútumán, Y attribútumain is meg kell, hogy egyezzenek.
- **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 attribútumhalmaz részhalmazai.

X -től funkcionálisan függ Y (jelölésben $X \rightarrow Y$), ha bármely R feletti T tábla esetén valahányszor két sor megegyezik X -en, akkor megegyezik Y -on is $\forall t_1, t_2 \in T$ esetén $(t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y])$.

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

Jobboldalak szétvágása (FF)

- $X \rightarrow A_1 A_2 \dots A_n$ akkor és csak akkor teljesül R relációra, 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.
- Baloldalak szétvágására nincs szabály!!!
- Ezért elég nézni, ha a FF-k jobboldalán egyetlen attribútum szerepel

Példa: FF

Sörivők(név, cím, kedveltSörök, gyártó, kedvencSör)

Adott FF-k, amelyek valószínűleg teljesülnek:

1. név -> cím kedvencSör

Ez az FF ugyanaz, mint

név -> cím és név -> kedvencSör

2. kedveltSörök -> gyártó

Példa: egy lehetséges előfordulásra

név	cím	kedveltSörök	gyártó	kedvencSör
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	Voyager	WickedAle	Pete's	WickedAle
Spock	Enterprise	Bud	A.B.	Bud

Mert **név** -> **cím**

Mert **név** -> **kedvencSör**

Mert **kedveltSörök** -> **gyártó**

Relációk kulcsai

- K *szuperkulcs* R relációra, ha K funkcionálisan meghatározza R attribútumait.
- K *kulcs* R -en, ha K szuperkulcs, de egyetlen valódi részhalmaza sem szuperkulcs.

Kulcs, superkulcs

- ▶ Funkcionális függőség $X \rightarrow Y$ speciális esetben, ha $Y = U$, ez a kulcsfüggőség.
- ▶ $R(U)$ relációséma esetén az U attribútumhalmaz 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 attribútumhalmazt nevezünk kulcsnak, amelytől az összes többi attribútum függ, de K -ból bármely attribútumot elhagyva ez már nem teljesül (vagyis minimális superkulcs)**

Példa: superkulcs, kulcs

Sörivók(név, cím, kedveltSörök, gyártó, kedvencSör)

- {név, kedveltSörök} **superkulcs**, ez a két attr. meghatározza funkcionálisan a maradék attr-kat.
 - név -> cím kedvencSör
 - kedveltSörök -> gyártó
- ▶ {név, kedveltSörök} **kulcs**, hiszen sem {név}, sem {kedveltSörök} nem superkulcs.
 - ▶ név -> gyártó; kedveltSörök -> cím **nem telj.**
- ▶ Az előbbin kívül nincs több kulcs, de számos superkulcs megadható még (ami ezt tartalm.)

Hogyan kaphatjuk meg a kulcsokat?

1. Szimplán megadunk egy K kulcsot.
 - Az FF-k $K \rightarrow A$ alakúak, ahol A „végigmegy” az összes attribútumon
2. Vagy: megadjuk az FF-eket, és ezekből következtetjük ki a kulcsokat.

Függőségek implikációja

- 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.
- Ez az adatbázis sémájának megtervezésekor lesz majd fontos.

Implikáció eldöntése --- 1

- Hogyan tudjuk ellenőrizni, hogy egy előfordulás nem teljesíti $Y \rightarrow A$
- $Y \rightarrow B$ teljesülésének ellenőrzésekor vegyünk két sort, amelyek megegyeznek az összes Y -beli attribútumon.

$\leftarrow Y \rightarrow$

0000000...0

00000??...?

Implikáció eldöntése --- 2

- Használjuk a megadott FF-ket annak igazolására, hogy az előbbi két sor más attribútumokon is meg kell, hogy egyezzen.
 - Ha B egy ilyen attribútum, akkor $Y \rightarrow B$ teljesül.
 - Egyébként az előbbi két sor olyan előfordulást ad majd, ami az összes előírt egyenlőséget teljesíti, viszont $Y \rightarrow B$ mégsem teljesül, azaz $Y \rightarrow B$ nem következménye a megadott FF-eknek.

Armstrong-axiómák

Legyen $X, Y \subseteq R$, és

XY jelentse az X és Y attribútumhalmazok egyesítését.

F legyen funkcionális függőségek tetsz. halmaza.

Armstrong axiómák:

- FD1 (reflexivitás): $Y \subseteq X$ esetén $X \rightarrow Y$.
- FD2 (bővíthetőség): $X \rightarrow Y$ és tetszőleges Z esetén $XZ \rightarrow YZ$.
- FD3 (transzitivitás): $X \rightarrow Y$ és $Y \rightarrow Z$ esetén $X \rightarrow Z$.

Levezetés fogalma

- F implikálja $X \rightarrow Y$ -t, ha minden olyan táblában, amelyben F összes függősége teljesül, $X \rightarrow Y$ 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 FD1, FD2, FD3 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. (Összevonhatósági szabály):

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

5. (Pseudotranzitivitás):

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

6. (Szétvághatósági szabály):

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

Bizonyítás (1): 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 (2): Bővítési axióma miatt $F \vdash XW \rightarrow YW$, és $YW = WY$, valamint a tranzitivitási axióma miatt $F \vdash XW \rightarrow Z$.

Bizonyítás (3): Reflexivitási axióma miatt $F \vdash Y \rightarrow Z$, és tranzitivitási axióma miatt $F \vdash X \rightarrow Z$.

Armstrong-axiómák

Következmény:

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

A következmény miatt feltehető, hogy a függőségek jobb oldalai 1 attribútumból állnak.

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

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

Mi is az a lezárt?

- 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 attribútumhalmaz lezárása F -re nézve.

Attribútumhalmaz lezártja

- **LEMMA:** $F \mid\!\!\!\rightarrow X \rightarrow Y \Leftrightarrow Y \subseteq X^+$.

Bizonyítás:

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

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

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

Algoritmus X^+ attr.halmaz lezártja

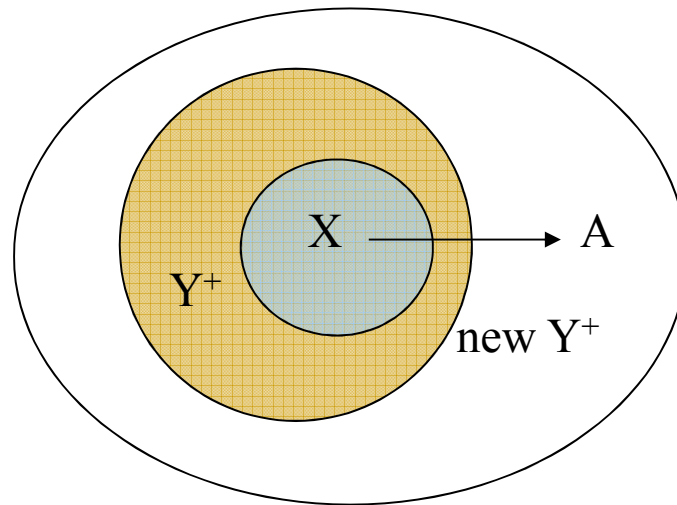
- Input: X attribútumhz., F funk.függőségek hz.
- Output: X^+ (zárás, típusa: attribútumhalmaz)
- Algoritmus X^+ kiszámítására:
/* Iteráció, amíg $X(n)$ változik */
 $X(0) := X$
 $X(n+1) := X(n) \cup \{A \mid \underline{W} \rightarrow Z \in F, A \in Z, W \subseteq X(n)\}$
 Ha $X(v+1) = X(v)$, akkor Output: $X(v) = X^+$.
- Miért működik az X^+ lezárási algoritmus?
(Tankönyv 3.2.5. szakasz, 81-83.oldal)

Implikáció eldöntése --- Lezárással

- 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.
- Hogyan tudjuk ellenőrizni, hogy egy előfordulás nem teljesíti $Y \rightarrow A$
- Van egyszerűbb út, ha kiszámítjuk Y *lezártját*, jelölésben Y^+ .
- **Kiindulás:** $Y^+ = Y$
- **Indukció:** (köv.lapon)

Lezárás (teszt)

- **Kiindulás:** $Y^+ = Y$
- **Indukció:** Olyan FF-ket keresünk, melyeknek a baloldala már benne van Y^+ -ban. Ha $X \rightarrow A$ ilyen, A -t hozzáadjuk Y^+ -hoz.



A lezárást kiszámító algoritmus „helyes” ---1

- 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? Könnyen bizonyítható indukcióval [Tk.3.2.5.]

A lezárást kiszámító algoritmus „helyes” ---2

- A második állítás: Miért talál meg minden igaz függőséget a lezárási algoritmus? [Tk.3.2.5.]
- 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 $X \rightarrow A$

Példa: Attribútumhalmaz lezárása

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

$X = \text{ABF}$, $X^+ = ?$

$X(0) := \text{ABF}$

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

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

$X(3) := \text{ABCEFG}$

$X^+ = \text{ABCEFG}$

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**a segítségével.

FF-i halmazok lezárása

- Adott S függőségi halmazzal ekvivalens függőségi halmaz S **minimális bázisa** B , ha az alábbi három feltétel teljesül.
 1. B összes függőségének jobb oldalán egy attribútum van.
 2. Ha bármelyik B -beli függőséget elhagyjuk, a fennmaradó halmaz már nem bázis
 3. Ha bármelyik B -beli funkcionális függőség bal oldaláról elhagyunk egy vagy több attr-t, akkor az eredmény nem marad bázis.

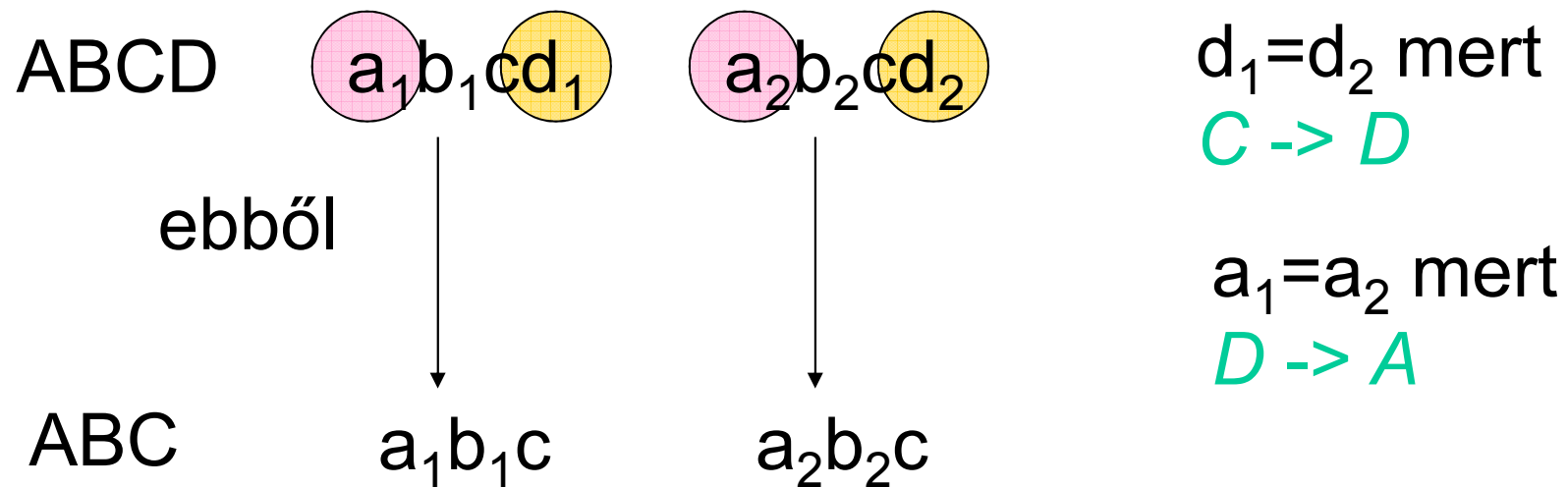
Minimális bázis kiszámolása

1. Jobboldalak szétvágása.
2. Próbáljuk törölni az FF-eket egymás után.
Ha a megmaradó FF-halmaz nem ekvivalens az eredetivel, akkor nem törölhető az épp aktuális FF.
3. Egymás után próbáljuk csökkenteni a baloldalakat, és megnézzük, hogy az eredetivel ekvivalens FF-halmazt kapunk-e.

FF-ek vetítése

- **Motiváció:** „normalizálás”, melynek során egy reláció sémát több sémára bonthatunk szét.
- Példa: $ABCD$ $F=\{AB \rightarrow C, C \rightarrow D, D \rightarrow A\}$
 - Bontsuk fel ABC és AD -re.
Milyen FF-k teljesülnek ABC -n?
 - Nem csak $AB \rightarrow C$, de $C \rightarrow A$ is!

Miért igaz a fenti példa?



Emiatt, ha két vetített sor C-n
megegyezik A-n is, azaz: $C \rightarrow A$.

FF-i halmaz vetületének kiszámítása

1. Induljunk ki a megadott FF-ekből és keressük meg az összes *nem triviális* FF-t, ami a megadott FF-ekből következik.
 - ❑ Nem triviális = a jobboldalt nem tartalmazza a bal.
2. Csak azokkal az FF-kel foglalkozzunk, amelyekben a projektált séma attribútumai szerepelnek.

Algoritmus (FF-i halmaz vetülete)

1. Legyen T az előálló FF-ek halmaza.
Kezdetben T üres
2. Minden X attribútum halmazra számítsuk ki X^+ -t.
3. Adjuk hozzá a függőségeinkhez $X \rightarrow A$ -t minden A -ra $X^+ - X$ -ből.
4. Dobjuk ki $XY \rightarrow A$ -t, ha $X \rightarrow A$ is teljesül.
 - ♦ Mert $XY \rightarrow A$ $X \rightarrow A$ -ból minden esetben következik.
5. Végül csak azokat az FF-eket használjuk, amelyekben csak a projektált attribútumok szerepelnek.

Példa: FF-k projekciója

- ABC , $A \rightarrow B$ és $B \rightarrow C$ FF-kel. Projektáljunk AC -re.
 - $A^+ = ABC$; ebből $A \rightarrow B$, $A \rightarrow C$.
 - Nem kell kiszámítani AB^+ és AC^+ lezárásokat.
 - $B^+ = BC$; ebből $B \rightarrow C$.
 - $C^+ = C$; semmit nem ad.
 - $BC^+ = BC$; semmit nem ad.
- A kapott FF-ek: $A \rightarrow B$, $A \rightarrow C$ és $B \rightarrow C$.
- AC -re projekció: $A \rightarrow C$.