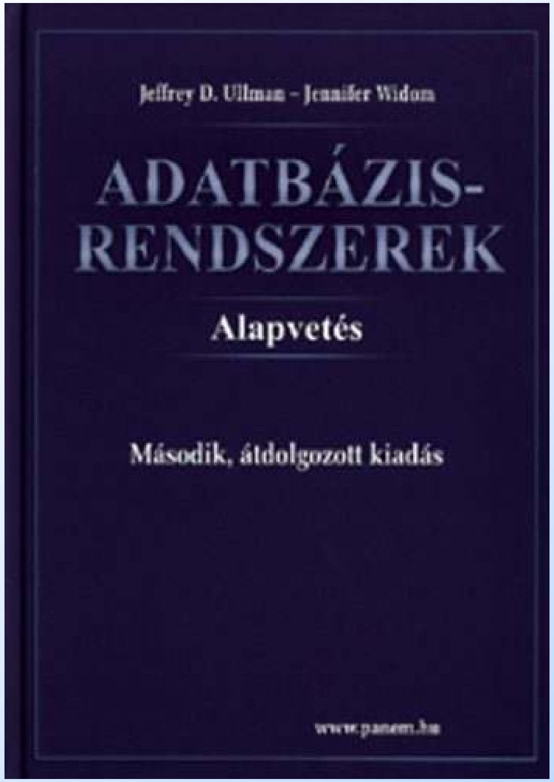
Relációs adatbázisok tervezése
1.rész (funkcionális függőségek)



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

3.1. Funkcionális függőségek
3.2. Funkcionális függőségekre vonatkozó szabályok

(Jeffrey D. Ullman 2007 előadásdiái alapján, Benczúr András, Kiss Attila és Kósa Balázs előadásainak felhasználásával, Hajas Csilla)

Relációs adatbázisok tervezése

- A tervezésre a BSc-s gyakorlatokon nem jut idő, ezért célszerű a tankönyvben is átnézni a tananyagot és a példákat! Tankönyv 3.fejezet.
- Függőségek - funkcionális, többértékű, ezeket a tervezésnél használják/használták, de az AB rendszerek nem támogatják, ott megszorítások
- Jó sémák - normálformák, normalizálás
- Bevezető példa: Filmek, Színészek, SzerepelBenne táblák összekapcsolása. Melyik séma jobb?

Funkcionális függőségek

- X -> 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 ,... attribútum halmazokat; A, B, C,... attribútumokat jelöl.
  - Jelölés: {A,B,C} attribútum halmaz helyett ABC-t írunk.

Funkcionális függőségek

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 $\forall\ t1,t2 \in T$ 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

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, szuperkulcs

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 szuperkulcs, ha a $K \rightarrow U$ funkcionális függőség 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 szuperkulcs).

Jobboldalak szétvágása (ff)

- $X\text{->}A_1A_2\ldots A_n$ akkor és csak akkor teljesül R relációra, ha $X\text{->}A_1$, $X\text{->}A_2$,…, $X\text{->}A_n$ is teljesül R-en.
- Példa: $A\text{->}BC$ ekvivalens $A\text{->}B$ és $A\text{->}C$ függőségek kettősével.
- Baloldalak szétvágására nincs általános szabály.
- Általában FF-k jobboldalán egyetlen attribútum szerepel majd.

Példa: FF

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

- FF-k, amelyek vszleg teljesülnek:
  1. név -> cím kedvencSör
     - Ez az FF ugyanaz, mint név -> cím és név -> kedvencSör.
  2. kedveltSörök -> gyártó.

Példa: egy lehetséges előfordulás

név	cím	kedveltSörök	gyártó	kedvenSör
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

Példa: szuperkulcs

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

- {név, kedveltSörök} szuperkulcs, hiszen a két attribútum meghatározza funkcionálisan a maradék attribútumokat.
  - név -> cím kedvencSör
  - kedveltSörök -> gyártó

Példa: kulcs

- {név, kedveltSörök} kulcs, hiszen sem {név}, sem {kedveltSörök} nem szuperkulcs.
  - név -> gyártól; kedveltSörök -> cím nem teljesülnek.
- Az előbbin kívül nincs több kulcs, de számos szuperkulcs megadható még.
  - Minden olyan halmaz, amit tartalmazza {név, kedveltSörök}-t.

Hogyan kaphatjuk meg a kulcsokat?

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

Egy természetes példa FF-re

- Példa: az „ugyanabban az időben nem lehet két előadás ugyanabban a teremben” lefordítva:

  idő terem -> előadás.

Függőségek implikációja

- Legyenek X_1 -> A_1, X_2 -> A_2,…, X_n -> A_n adott FF-k, szeretnénk tudni, hogy Y -> B teljesül-e olyan relációkra, amire az előbbi FF-k teljesülnek.
  - Példa: A -> B és B -> C teljesülése esetén A -> C biztosan teljesül.
- Ez az adatbázis sémájának megtervezésekor lesz majd fontos.

Implikáció eldöntése

- 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?? . . . ?

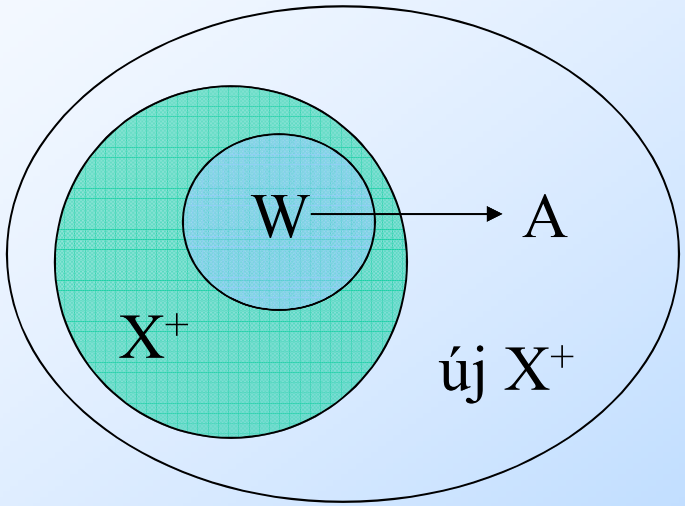
Implikáció eldöntése

- 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 -> B* teljesül.
  - Egyébként az előbbi két sor olyan előfordulást ad majd, ami az összes előírt egyenlőséget teljesíti, viszont *Y -> B* mégsem teljesül, azaz *Y -> B* nem következménye a megadott FF-eknek.

Lezárás

- Egy egyszerűbb út, ha kiszámítjuk X *lezártját*, jelölésben X^+.
- Kiindulás: $X^+ = X$.
- Indukció: Olyan FF-ket keresünk, melyeknek a baloldala már benne van X^+-ban. Ha $W \to A$ ilyen, A-t hozzáadjuk X^+-hoz.

Lezárás



Mi is az a lezárt?

- Adott *R* séma és *F* funkcionális függőségek halmaza mellett, X^+ az összes olyan *A* attribútum halmaza, amire *X->A* következik *F*-ből.

Attribútumhalmaz lezárása

- (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.
- LEMMA: $F \vdash X \rightarrow Y \Leftrightarrow Y \subseteq X^{+}$.

  Bizonyítás:

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

  ($\Leftarrow$) $\forall A \in Y \subseteq X^{+}$ esetén $F \vdash X \rightarrow A$, és az egyesítési szabály miatt $F \vdash X \rightarrow Y$.
- Lemma következménye: az implikációs probléma megoldásához elég az <u>X^{+}-t hatékonyan kiszámolni</u>

Algoritmus X^+ kiszámítására

- 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)

A lezárást kiszámító algoritmus „helyes” I.

- 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.
- Megjegyzés: az első állítás könnyen bizonyítható indukcióval.

A lezárást kiszámító algoritmus „helyes” II.

- (2) 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

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

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

$X(0) := ABF$

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

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

$X(3) := ABCEFG$

$X^+ = ABCEFG$

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 algoritmusa segítségével.

Az összes következmény FF megtalálása

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

Miért?

ABCD	$a_1b_1cd_1$	$a_2b_2cd_2$	$d_1=d_2$ mert *C -> D*
ebből	$\downarrow$	$\downarrow$	$a_1=a_2$ mert *D -> A*
ABC	a_1b_1c	a_2b_2c	

Emiatt, ha két vetített sor C-n megegyezik A-n is, azaz: *C -> A*.

Alapötlet

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.

Exponenciális algoritmus

1. Minden X attribútum halmazra számítsük ki X^+-t.
2. Adjuk hozzá a függőségeinkhez $X \to A$-t minden A-ra $X^+ - X$-ből.
3. Dobjuk ki $XY \to A$-t, ha $X \to A$ is teljesül.
   - Mert $XY \to A$ $X \to A$-ból minden esetben következik.
4. Végül csak azokat az FF-ket használjuk, amelyekben csak a projektált attribútumok szerepelnek.

Példa: FF-k projekciója

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

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 (tranzitivitás): $X \rightarrow Y$ 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
  $X_1 \rightarrow Y_1$, ..., $X_k \rightarrow Y_k$,..., $X \rightarrow Y$ 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):
   $F|\!\!-\!\!- X \to Y$ és $F|\!\!-\!\!- X \to Z$ esetén $F|\!\!-\!\!- X \to YZ$.
5. (Pszeudotranzitivitás):
   $F|\!\!-\!\!- X \to Y$ és $F|\!\!-\!\!- WY \to Z$ esetén $F|\!\!-\!\!- XW \to Z$.
6. (Szétvághatósági szabály):
   $F|\!\!-\!\!- X \to Y$ és $Z \subseteq Y$ esetén $F|\!\!-\!\!- X \to Z$.

Bizonyítás (1): Bővítési axióma miatt $F|\!\!-\!\!- XX \to YX$ és $F|\!\!-\!\!- YX \to YZ$, és $XX=X$, valamint a tranzitivitási axióma miatt $F|\!\!-\!\!- X \to YZ$.

Bizonyítás (2): Bővítési axióma miatt $F|\!\!-\!\!- XW \to YW$, és $YW=WY$, valamint a tranzitivitási axióma miatt $F|\!\!-\!\!- XW \to Z$.

Bizonyítás (3): Reflexivitási axióma miatt $F|\!\!-\!\!- Y \to Z$, és tranzitivitási axióma miatt $F|\!\!-\!\!- X \to Z$.

Armstrong-axiómák

Következmény:

$F \vdash X \rightarrow Y \Leftrightarrow \forall A_i \in Y$ 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.