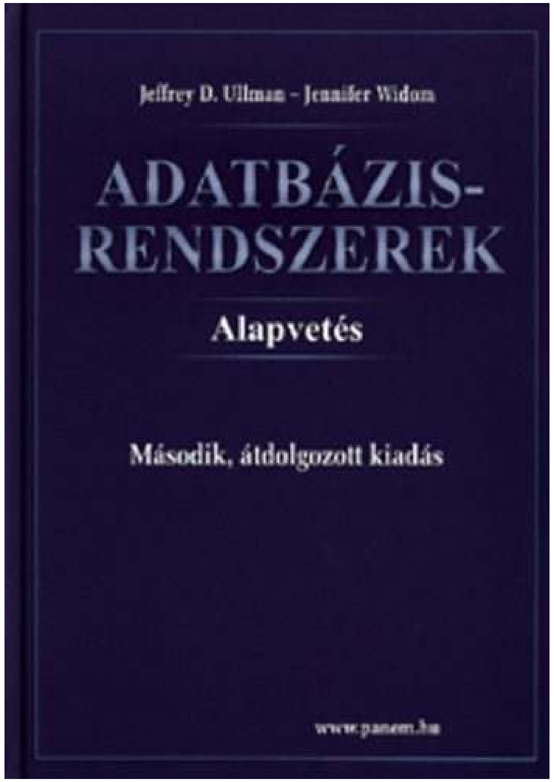
Adatbázisok 1

Funkcionális függőségek

Témakör – Funkcionális függőségek

- Ullman-Widom: Adatbázisrendszerek Alapvetés Második, átdolgozott kiadás, Panem, 2009
- 3.1. Funkcionális függőségek, relációk (szuper)kulcsai
- 3.2. Funkcionális függőségekre vonatkozó szabályok, lezárási algoritmusok, f.f. hz minimális fedése, f.f.-ek 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?

Sörivó(<u>név</u>, cím, <u>kedveltSörök</u>, 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 a név -> cím kedvencSör és kedveltSörök -> gyártó FF-ek felhasználásával tudjuk, mi szerepel.

Hibás tervezés problémái

A rosszul tervezettség anomáliákat is eredményez

Sörivó(<u>név</u>, cím, <u>kedveltSörök</u>, 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 *Karcsira* módoítjuk, megtesszük-e ezt minden sornál? Törlési anomália: Ha senki sem szereti a Bud sört, azt sem tudjuk, ki gyártotta. Beszúrási anomália

Dekomponálás (felbontás)

- A fenti problémáktól dekomponálással (felbontással) tudunk megszabadulni!

Definíció:
$d=\{R_1,...,R_k\}$ az (R,F) dekompozíciója, ha nem marad ki attribútum, azaz $R_1 \cup ... \cup R_k = R$.
(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)

- $X \to Y$ 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)

Definíció. Legyen R(U) egy relációséma, továbbá X és Y az U attribútumhalmaz részhalmazai. X-től *funkcionálisan függ* Y (jelölésben $X \to Y$), 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:

$R \models X \to Y$ vagyis R kielégíti $X \to Y$ függőséget

Jobboldalak szétvágása (FF)

- X->$A_1A_2\ldots A_n$ akkor és csak akkor teljesül R relációra, ha X->A_1, X->A_2,…, X->A_n is teljesül R-en.
- Példa: A->BC ekvivalens A->B és A->C függőségek kettősével.
- Baloldalak szétvágására nincs szabály!!!
- Ezért elég nézni, ha a FF-k jobboldalán egyetlen attribútum szerepel

Példa: FF

Sörivók(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, 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 $K \rightarrow U$ 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 szuperkulcs)

Példa: szuperkulcs, kulcs

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

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

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.

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 --- 1

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

Y

00000 00. . . 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* -> *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.

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 $X \to Y$-t, ha minden olyan táblában, amelyben F összes függősége teljesül, $X \to Y$ is teljesül.
- Jelölés: $F \models X \to Y$, ha F implikálja $X \to Y$ –et.
- $X \to Y$ levezethető F-ből, ha van olyan

  $X_1 \to Y_1, \ldots, X_k \to Y_k, \ldots, X \to Y$ véges levezetés, hogy $\forall k$-ra

  - $X_k \to Y_k \in F$ vagy
  - $X_k \to Y_k$ az FD1,FD2,FD3 axiómák alapján kapható a levezetésben előtte szereplő függőségekből
- Jelölés: $F \vdash X \to Y$, ha $X \to Y$ levezethető F-ből

További levezethető szabályok:

4. (Összevonhatósági szabály):
   $F \vdash X \to Y$ és $F \vdash X \to Z$ esetén $F \vdash X \to YZ$.
5. (Pszeudotranzitivitás):
   $F \vdash X \to Y$ és $F \vdash WY \to Z$ esetén $F \vdash XW \to Z$.
6. (Szétvághatósági szabály):
   $F \vdash X \to Y$ és $Z \subseteq Y$ esetén $F \vdash X \to Z$.

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

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

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

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.

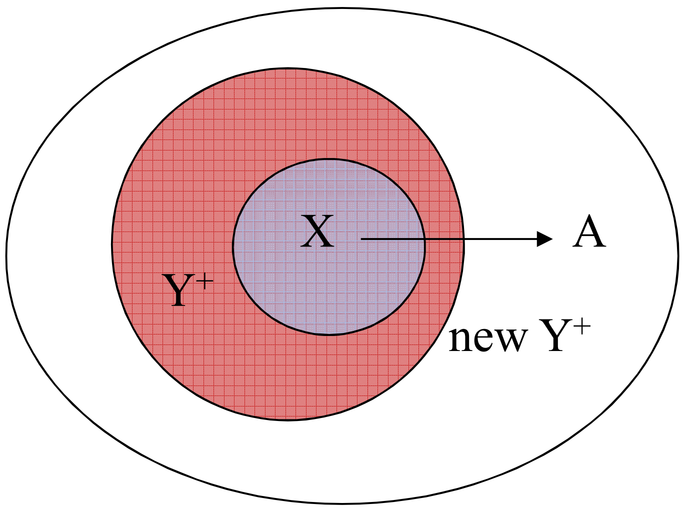
$$F \vdash X \rightarrow Y \Leftrightarrow F \models X \rightarrow Y$$

Implikáció eldöntése --- Lezárással

- 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.
- Hogyan tudjuk ellenőrizni, hogy egy előfordulás nem teljesíti Y -> 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.



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

  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^+ attribútumhalmaz lezártja

- Input: X attribútumhz., F funk.függőségek hz.
- Output: X^+ (zárás, típusa: attribútumhalmaz)
- Algoritmus X^+ kiszámítására:

/* Iteráció, amíg X(n) változik */

$X(0) := X$

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

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” ---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.]
- Konstrukciós 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

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

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

$X(0) := ABF$

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

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

$X(3) := ABCEFG$

$X^+ = ABCEFG$

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: $K \to O$.

Egy helyen egy időben egy kurzus lehet: $IT \to K$.

Egy időben egy tanár csak egy helyen lehet: $IO \to T$.

Egy időben egy diák csak egy helyen lehet: $ID \to T$.

Egy diák egy kurzust egy végső jeggyel zár: $KD \to J$.

- R=KOITDJ F= $\{K \to O,\ IT \to K,\ IO \to T,\ ID \to T,\ KD \to J\}$
- Feladat: Határozzuk meg a (R, F) kulcsait az X^+ kiszámítási algoritmusa 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* ->*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 igaz a fenti példa?

ABCD	$a_1b_1cd_1$	$a_2b_2cd_2$
ebből	↓	↓
ABC	a_1b_1c	a_2b_2c

$d_1=d_2$ mert
C -> D

$a_1=a_2$ mert
D -> A

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

1. Legyen T az előálló FF-ek halmaza. Kezdetben T üres
2. Minden X attribútum halmazra számítsük 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-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: $A \to C$.