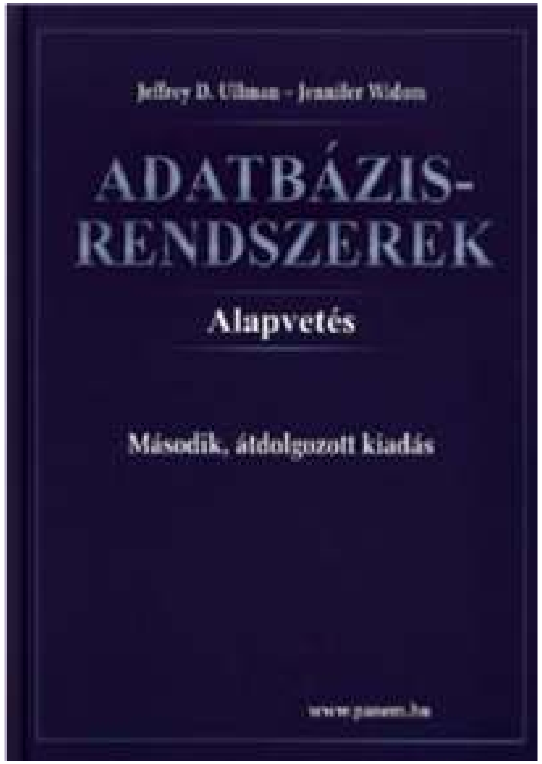
Relációs adatbázisok tervezése
1.rész: funkcionális függőségek

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 (vázlat)

- I. Függőségek: a sématervezésnél használjuk
  - funkcionális függőség, többértékű függőség
- II. Felbontás: „jó” tulajdonságú felbontás
  - Veszteségmentesség
  - Függőségőrző felbontás
- III. Normalizálás: „jó” sémákra való felbontás
  - Funkcionális függőségek -> 1,2,3NF, BCNF
  - Többértékű függőségek -> 4NF
  - Összekapcsolási függőségek -> 5NF

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)

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

Példa: Funkcionális függőség

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

Feltehetjük például, hogy az alábbi FF-ek teljesülnek:

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

Jobboldalak szétvágása (FF)

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

Kulcs, szuperkulcs

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

Másik példa (több kulcs is lehet)

- Legyen *ABC* sémán def.FF-ek: *AB* ->*C* és *C* ->*B*.
  - Példa1: *A* = utca, *B* = város, *C* = irányítószám.
  - Példa2: *A* = oktató, *B* = időpont, *C* = kurzus.
- Itt két kulcs is van: {*A*,*B* } és {*A*,*C* }.

Az implikációs probléma

- 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.
- Implikációs probléma eldöntése definíció alapján túl nehéz, egyszerűbb lehetőség levezetési szabályok segítségével, lásd Armstrong-axiómák:

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:

- 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, 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

  $X_1 \rightarrow Y_1, \ldots, X_k \rightarrow Y_k, \ldots, X \rightarrow Y$ 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. (Pszeudotranzitivitás):
   $F \vdash X \rightarrow Y$ 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 (4): 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 (5): Bővítési axióma miatt $F \vdash XW \rightarrow YW$, és $YW=WY$, valamint a tranzitivitási axióma miatt $F \vdash XW \rightarrow Z$.

Bizonyítás (6): Reflexivitási axióma miatt $F \vdash Y \rightarrow Z$, és tranzitivitási axióma miatt $F \vdash X \rightarrow Z$.

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.

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

- Következmény:

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

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

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 -> B
- A levezetési szabályoknál van egyszerűbb út: kiszámítjuk *Y lezártját*: Y^+-t
- Attribútum-halmaz lezárására teszt: a köv.lapon

Attribútumhalmaz lezártja (definíció)

- 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 attribútum-halmaz lezárása F-re nézve.

Attribútumhalmaz lezártja (lemma)

- LEMMA: $F \vdash X \to Y \Leftrightarrow Y \subseteq X^+$.

  Bizonyítás:

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

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

- 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 */

$X(0) := X$

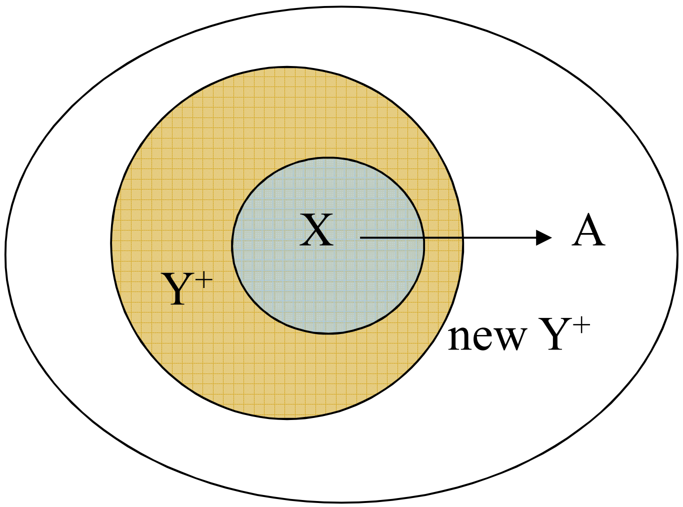
$X(n+1) := X(n) \cup \{A \mid \underline{W \rightarrow Z \in F}, A \in Z, W \subseteq X(n)\}$

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)

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”

- 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 „teljes”

- 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

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

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

$X(0):=ABF$

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

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

$X(3):=ABCEFG$

$X^+= ABCEFG$

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.

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?
  - *ABC* –n nemcsak az *AB* ->*C*, de a *C* ->*A* is teljesül! (lásd a köv.oldalon)

Miért igaz a fenti példa?

ABCD	$a_1b_1cd_1$	$a_2b_2cd_2$
ebből	$\downarrow$	$\downarrow$
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 = vagyis a jobboldalt nem tartalmazza a baloldal, lásd A1 axióma.
2. Csak azokkal az FF-kel foglalkozzunk, amelyekben a projektált séma attribútumai szerepelnek.

Algoritmus (FF-i halmaz vetülete)

- Legyen T az előálló FF-ek halmaza. Kezdetben T üres
- Minden X attribútum halmazra számítsük ki X^+-t.
- Adjuk hozzá a függőségeinkhez $X \to A$-t minden A-ra $X^+ - X$-ből.
- Dobjuk ki $XY \to A$ -t, ha $X \to A$ is teljesül

  mert $XY \to A$ $X \to A$ -ból minden esetben következik
- Végül csak azokat az FF-ket használjuk, amelyekben csak a projektált attribútumok szerepelnek.

Példa: FF-k projekciója

- *ABC, A ->B* és *B ->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: $A \to C$.