Adatbázisok elmélete 14. előadás

Csima Judit
Budapesti Műszaki és Gazdaságtudományi Egyetem
Számítástudományi Tsz.
I. B. 136/b
csima@cs.bme.hu

2003. Április 1.

Relációs sémák tervezése

Van elméleti alap $\Longrightarrow$ érv a relációs technika mellett
(Objektumosnak nincs ilyen.)

Kérdés:

- Mik a jó relációk?
- Milyen relációkat érdemes tárolni?
- Hogyan alakíthatunk tetszőleges relációkat jókká?

Cél: El akarunk kerülni kellemetlen jelenségeket, **anomáliákat**:

- *Módosítási anomália:* pl. ha a Termék(Termelő, Cím, Terméknév, Ár) reláció esetén egy termelő címe több sorban is előfordul, változáskor mindenhol át kell írni. Hiba esetén inkonzisztencia.

- *Beszúrási anomália:* Nem tudunk beszúrni adatot, ha az egyik attribútum hiányzik, mert nem ismerjük (és nem lehet NULL).
- *Törlési anomália:* Csak egész sorok törölhetők, így elveszhetnek hasznos adatok. Pl. ha egy termelő épp nem termel semmit, kitöröljük a címét is.

A relációk, tárolás jósága attól függ, hogy milyen megkötések vannak az adatokon.

Megszorítások két osztálya:

- *Értékfüggő:* Pl. ÁR≥ 0, ÉLETKOR egész ≤ 1000, NÉV karaktersor, CÍM$\neq$NULL, (típusleírások)
- *Értékfüggetlen:* TERMÉKNÉV, TERMELŐ kulcs; $\forall$ TERMELŐ-nek egy címe van, egy TERMELŐ azonos nevű termékéből csak egy árú van

Utóbbi: az attribútumok mennyire függenek egymástól $\Longrightarrow$ **funkcionális függőség**

Funkcionális függőségek

Jelölés: $R(A_1,\ldots,A_n)$ reláció, X attribútum halmaz $\Longrightarrow X \subseteq R$
$X = \{A_{i_1}, A_{i_2}, \ldots, A_{i_k}\}$ helyett $X = A_{i_1}A_{i_2}\ldots A_{i_k}$

Definíció. *$Y \subseteq R$ **funkcionálisan függ** $X \subseteq R$**-től**, (jelölés: $X \to Y$), ha R bármely két sorára igaz, hogy ha ők megegyeznek X-en, akkor Y-on is megegyeznek.*

Pl. $X =$TERMELŐ, TERMÉKNÉV; Y=ÁR $\Longrightarrow X \to Y$

Megjegyzések:

- Azok az érdekes összefüggések, amik minden ilyen attribútumokkal rendelkező táblában fenn kell, hogy álljanak: axiómaszerű feltételek, az adatbázis bármely változása esetén is fennállnak $\Longrightarrow$ *érdemi függés*
  Azok, amik csak véletlenül, csak egy pillanatban állnak fenn $\Longrightarrow$ *eseti függés*, (ezek nem érdekelnek, például lehetséges hogy egy adott pillanatban minden ár csak egyszer szerepel és ekkor úgy tűnik, mintha Ár$\to$ Termék érvényes függés lenne)

- Tehát az érdemi függések megadása modellezési kérdés: a séma megadásakor döntjük el, hogy milyen függéseket akarunk fenntartani mindenáron.
  Ezentúl a relációs sémának része lesz a függőségek halmaza F is $\Longrightarrow (R,F)$
  Vagyis megadjuk, hogy mik a séma attribútumai és mik az érdemi függései.

  Pl.
  TERMELŐ, TERMÉKNÉV $\rightarrow$ TERMELŐ, TERMÉKNÉV, ÁR, CÍM
  TERMELŐ $\rightarrow$ CÍM

  S(NÉV, CÍM, VÁROS, IRÁNYÍTÓSZ, TELEFON)

  CÍM, VÁROS $\rightarrow$ IRÁNYÍTÓSZ
  IRÁNYÍTÓSZ $\rightarrow$ VÁROS
  NÉV, CÍM, VÁROS $\rightarrow$ TELEFON
- Egy adott reláció adott állapotából nem következik semmilyen érdemi függés. Viszont látszódhat olyan, hogy mi nem függhet mitől.
- $X \rightarrow Y$ teljesülhet úgy is, hogy az adott relációban nincs is két olyan sor, amik X-en megegyeznek.

- X-nek és Y-nak nem kell diszjunktaknak lenniük

A séma megadása csak a keretet jelenti, beleértve a függeseket is, ha ezt feltöltjük adatokkal, akkor kapunk egy a sémára illeszkedő relációt. A r reláció akkor illeszkedik az (R,F) sémára ha az attribútumai az R-ben adottak és teljesülnek benne az F függések.

Logikai következmény

Kérdés: ha adott egy F függéshalmaz és egy reláció, amiben F függései igazak, akkor milyen további függések lesznek még biztosan igazak?

Például: ha HALLGATÓ, TÁRGY $\rightarrow$ GYAKORLAT és GYAKORLAT $\rightarrow$ GYAKVEZ, akkor HALLGATÓ, TÁRGY $\rightarrow$ GYAKVEZ.

Azaz általánosabban: ha $XY \rightarrow Z$ és $Z \rightarrow W$, akkor attól függetlenül, hogy mi a reláció és X,Y,Z,W, igaz lesz, hogy $XY \rightarrow W$.
Definíció. *Adott (R,F). Az $X \rightarrow Y$ függés **logikai következménye** (szemantikai következménye) F-nek, ha az $X \rightarrow Y$ minden olyan r relációban teljesül, ahol F függései mind teljesülnek.*
Jelölése: $F \models X \rightarrow Y$

Azaz ez a fogalom azt adja meg, hogy mely függéseknek kell szükségszerűen

teljesülniük minden olyan sémában/relációban, ahol F függései fennállnak.

Hogyan lehetne ezeket meghatározni, illetve eldönteni, hogy egy függés ilyen-e?

Felveszünk axiómákat, és azok segítségével próbáljuk levezetni. Persze ehhez az kell, hogy pontosan azokat lehessen levezetni F-ből, akik logikai következményei neki.
Levezethetőség jele: $F \vdash X \rightarrow Y$

Tehát bevezetünk axiómákat, levezethetőséget és belátjuk, hogy $\models \Longleftrightarrow \vdash$.
(Pl. logikában így van.)
$\models \Rightarrow \vdash$: *Teljességi tétel*, ami igaz az levezethető.
$\vdash \Rightarrow \models$: *Igazság tétel*, csak igaz dolgok vezethetők le.
Definíció. *Egy $X \rightarrow Y$ függőség akkor vezethető le egy adott F függőséghalmazból, ha az axiómák ismételt alkalmazásával F-ből megkapjuk $X \rightarrow Y$-t. Jele: $F \vdash X \rightarrow Y$.*

Armstrong-axiómák

1. Reflexivitás: Ha $X,Y \subseteq R$ és $Y \subseteq X$, akkor $X \rightarrow Y$.
2. Kiegészítési tulajdonság: Ha $X,Y \subseteq R$ és $X \rightarrow Y$, akkor $XW \rightarrow YW$ igaz tetszőleges $W \subseteq R$-re.

3. Tranzitivitás: Ha $X,Y,Z \subseteq R$, $X \to Y$ és $Y \to Z$, akkor $X \to Z$.

Bizonyítás: **(Igazság tétel)**
Azt kell belátni, hogy ha egy függés (esetleg több lépésben) levezethető F-ből a három axióma segítségével, akkor ez a függés logikai következménye is F-nek, azaz minden olyan relációban, ahol F minden függése teljesül, ott teljesül a levezetett függés is. Ehhez elég azt belátni, hogy külön-külön, az egyes axiómák egyszeri használata ilyen függést ad.

1. Azt kell belátni, hogy minden r relációban, minden $Y \subseteq X \subseteq R$ attribútumhalmaz esetén $X \to Y$ igaz, azaz ha r bármely két adott sora megegyezik X-en, akkor megegyeznek Y-on is. De mivel $Y \subseteq X$, ezért nyilván megegyeznek Y-on, ha X-en megegyeztek. $\surd$
2. Az kell, hogy ha egy R-re illeszkedő r relációban $X \to Y$ igaz, akkor $XW \to YW$ is igaz lesz. Vegyünk két sort r-ben, ami megegyezik XW-n. Ekkor ezek megegyeznek X-en és W-n is, külön-külön. Mivel $X \to Y$, így megegyeznek Y-n is, tehát YW-n is. $\surd$

3. Az kell, hogy ha egy R-re illeszkedő r relációban $X \to Y$ és $Y \to Z$ igaz, akkor $X \to Z$ is igaz lesz. Vegyünk két sort, ami megegyezik X-en. Mivel $X \to Y$, megegyeznek Y-n is. De mivel $Y \to Z$, megegyeznek Z-n is. $\surd$

Példa

Állítás. *Ha R(Város, Utca, Irányítószám) és $F = \{VU \to I, I \to V\}$, akkor $F \vdash IU \to VIU$ (és mivel $\vdash \Rightarrow \models$-t már láttuk, ezért $F \models IU \to VIU$).*

Bizonyítás:
- i) $I \to V$: ez F-beli
- ii) $IU \to VU$: kiegészítve U-val
- iii) $IU \to IVU$: kiegészítve I-vel

Néhány további szabály, ami levezethető az axiómákból (és az igazságtétel miatt igazak is).

Állítás (Unió szabály). $\{X \to Y,\ X \to Z\} \vdash X \to YZ$

Bizonyítás:

- i) $X \to Y$: ez F-beli
- ii) $XZ \to YZ$: kiegészítve Z-val
- iii) $X \to Z$: ez F-beli
- iv) $X \to XZ$: kiegészítve X-vel
- v) $X \to YZ$: iv) és ii) + tranzitivitás

Állítás (Áltranzitiv szabály). $\{X \to Y,\ YW \to Z\} \vdash XW \to Z$

Bizonyítás:
- i) $X \to Y$: ez F-beli
- ii) $XW \to YW$: kiegészítve W-val
- iii) $YW \to Z$: ez F-beli
- iv) $XW \to Z$: ii) és iii) + tranzitivitás

Állítás (Felbontási szabály). *Tegyük fel, hogy $Z \subseteq Y$, ekkor $\{X \to Y\} \vdash X \to Z$*

Bizonyítás:
- i) $X \to Y$: ez F-beli
- ii) $Y \to Z$: reflexivitás
- iii) $X \to Z$: i) és ii) + tranzitivitás

Lezárás

Definíció. *Ha F egy függéshalamaz, akkor a **lezártja** F^+ az F-ből levezethető összes függés:*

$$F^+ = \{X \to Y \mid F \vdash X \to Y\}$$

Jó: mert ha majd belátjuk $\models \Longleftrightarrow \vdash$-t, akkor kiderül, hogy ez éppen az F-ből szükségszerűen következő összes függést adja meg.

Gond: nagyon nagy lehet
Pl. $R(A_1,\ldots,A_n,B_1,\ldots B_n)$ és $F = \{A_i \to B_j \mid 1 \le i,j \le n\}$, akkor ez n^2 db függés
F^+-ban benne van minden $A_{i_1}\ldots A_{i_k} \to B_{j_1}\ldots B_{j_l}$, azaz $(2^n-1)(2^n-1) \approx 2^{2n}$ eleme van.

Ezért ehelyett valami mást nézünk, amit könnyebb lesz meghatározni és jól közelíti F^+-t:

Definíció. *Ha X egy attribútum halmaz (R,F)-ben, akkor* ***lezártja***

$$X^+(F) = \{A \in R \mid F \vdash X \to A\},$$

azaz azon attribútumok, amik függnek X-től.

Állítás. $X \subseteq X^+(F) \subseteq R$
Lemma. *(Fontos!!!)* $F \vdash X \to Y \iff Y \subseteq X^+(F)$

Bizonyítás: $\boxed{\Longrightarrow:}$ Tegyük fel, hogy $F \vdash X \to Y$ és legyen $A \in Y$.
$F \vdash X \to A$, hiszen vegyük $X \to Y$ levezetését és alkalmazzuk a felbontási szabályt a végén.
Definíció szerint ekkor $A \in X^+(F)$. Ez minden $A \in Y$-ra igaz. $\surd$

$\boxed{\Longleftarrow:}$ Legyen $Y = A_1 \dots A_k \subseteq X^+(F)$.
Így definíció szerint $\forall A_i \in Y$-ra $F \vdash X \to A_i$.
Ekkor $X \to Y$ levezetése: vesszük az A_i-k levezetését és a végén alkalmazzuk az uniós szabályt $k-1$-szer. $\surd$

Következménye: Ha minden X-re ismerjük/ki tudjuk számítani $X^+(F)$-et, akkor tetszőleges $X \to Y$ függésről eldönthető, hogy F^+-beli-e vagy sem, mert $X \to Y \in F^+$ pontosan akkor teljesül (definíció szerint), ha $F \vdash X \to Y$, de ez meg az előbbi lemma szerint pontosan akkor van, ha $Y \subseteq X^+(F)$

Megjegyzés: Majd látjuk, hogy $X^+(F)$ kiszámolására lesz gyors algoritmus.