Adatbázisok elmélete 6. előadás

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

2003. Március 4.

Relációs adatmodell

Legfontosabb és leggyakoribb a létező adatmodellek közül.

Mit fogunk róla tanulni?

1. elvi keret (alapfogalmak, alapműveletek)
2. konkrét nyelvek (ISBL, QBE, QUELL, SQL, sémadefinícióra, adatmódosításra és lekérdezésre)
3. tervezés (minél jobb séma kialakítása, matematikai elmélet)

Egyetlen alapfogalom (nincs külön egyedhalmaz és kapcsolat): reláció.

A reláció definíciója

1. Gondolhatunk rá úgy, mint egy síkbeli táblázatra:

R_1	A_1	A_2
	1	y
	1	z
	3	y

R_2	A_1	A_2
	2	y
	1	z

Itt R_1 a reláció neve, A_1 és A_2 az attribútumok nevei, a sorok pedig a reláció elemei. Az oszlopokban levő értékek az attribútumokhoz tartozó értékkészletből kerülnek ki.

2. Tekinthetjük egy Descartes-szorzat részhalmazának is a relációt:

$A_1, A_2, \ldots, A_n$ tetszőleges halmazok (attribútumok)

$$R \subseteq A_1 \times \cdots \times A_n$$

$\Longrightarrow$ Minden sor csak egyszer szerepel
$\Longrightarrow$ a sorok sorrendje lényegtelen.

Példa: $A_1 = \{1,2,3\}, A_2 = \{x,y,z\}$
$R_1 = \{\{1,y\},\{1,z\},\{3,z\}\}$
$R_2 = \{\{2,y\},\{1,z\}\}$

De R elemeit tekinthetjük halmazoknak is, nem rendezett n-eseknek.
Ekkor az attribútumok sorrendje is mindegy.

3. Gondolhatunk egy relációra úgy is, mint függvények halmazára:

Definíció. *Egy sor = egy függvény:* $s : \{attribútumok\} \to \{attr.\ értékkészlete\}$
Egy R reláció ilyen függvények halmaza.

Így tényleg nem számít a sorrend, se a sorok között, se az attribútumok között.
Nincs két azonos sor.

Például:
R_1-ben: 1. sor: $A_1 \to 1$; $A_2 \to$ y;

Jelölés:

Definíció. *Relációs séma:* $R(A_1,\ldots,A_n)$, *ahol R a reláció neve, az A_i-k pedig az attribútumok nevei.*

Például: Személy(Vezetéknév, Keresztnév, Neme, Végzettsége)

Gyakorlatban azért mégis rögzítünk egy sorrendet, azt, amelyikben felsoroljuk az attribútumokat.

Relációs modell

Teljes adatmodell: nem csak azt mondja meg hogyan írok le, hanem vannak műveletek is.

Ezeket a műveleteket relációkra alkalmazhatom és így újabb relációkat kapok majd.

A relációs algebra alapműveletei

- Halmazműveletek (bármilyen halmazra mennének)
  - unió: $\cup$
  - különbség: $\setminus$
  - szorzat: $\times$
- Relációs műveletek (ezek már kihasználják, hogy itt relációkról van szó)
  - vetítés, projekció: π
  - kiválasztás, szelekció: σ

Ezek mind tiszta műveletek: reláció $\to$ reláció
$\Longrightarrow$ gond nélkül egymásba ágyazhatók

Műveletek

Unió

- R, S relációk $\Longrightarrow R \cup S =$ sorai vagy R vagy S sorai
  Azonos sorok csak egyszer szerepeljenek. (Gyakorlatban néha lehetnek azonos sorok.)
- csak akkor alkalmazható, ha R és S oszlopszáma egyenlő
- nem feltétlenül örököl típusokat vagy attribútum neveket

- Példa:

R	A	B
	a	a
	a	c
	b	a

S	A	C
	a	a
	a	d
	a	c
	b	b

$R \cup S$	A	$(R \cup S)_2$
	a	a
	a	c
	b	a
	a	d
	b	b

Műveletek

Különbség

- R, S relációk $\Longrightarrow R \setminus S = R$ azon sorai, amelyek S-ben nem szerepelnek
- nincs kompatibilitási követelmény (Ha pl. különböző az oszlopszám, nem szerepelhetnek azonos sorok úgysem. Ekkor $r \setminus S = R$)
- Az eredmény örökli R típusait és attribútum neveit (mert $R \setminus S \subseteq R$)
- Példa:

R	A	B
	a	a
	a	c
	b	a

S	A	C
	a	a
	a	d
	a	c
	b	b

$R \setminus S$	A	B
	b	a

Műveletek

Szorzat (direkt szorzat, Descartes szorzat)

- $R(A_1, \ldots, A_k), S(B_1, \ldots, B_l)$ k ill. l attribútumos relációk $\Longrightarrow R \times S =$ egy $k+l$ attribútumos reláció, R minden sora mögé odatesszük S minden sorát, minden lehetséges módon.
  Ha R-nek n sora van S-nek m sora $\Longrightarrow R \times S$-nek nm sora van
- nincs kompatibilitási követelmény
- Az eredmény lényegében örökli R és S típusait és attribútum neveit, esetleg át kell nevezni.

- Példa:

R	A	B
	a	a
	a	c
	b	a

S	A	C
	a	a
	a	d
	a	c
	b	b

$R \times S$	A	B	A'	C
	a	a	a	a
	a	a	a	d
	a	a	a	c
	a	a	b	b
	a	c	a	a
	a	c	a	d
	a	c	a	c
	a	c	b	b
	b	a	a	a
	b	a	a	d
	b	a	a	c
	b	a	b	b

Az unió és különbség könnyű művelet, a szorzat nehezebb. Vigyázni kell mennyit használjuk.

Műveletek

Vetítés

- $R(A_1, \ldots, A_l)$ alakú reláció $\Longrightarrow \pi_{A_{i_1}, \ldots, A_{i_n}}(R)$
  R vetítése $A_{i_1}, \ldots, A_{i_n}$-re (fontos a sorrend) $\Longrightarrow$
  Veszem az oszlopokat ebben a sorrendben, a többit eldobom és a többszörös sorokat is eldobom.
  Egy oszlop akár többször is szerepelhet. $\Longrightarrow$ átnevezés
- nincs kompatibilitási követelmény (persze amire vetítünk az R-nek attribútuma kell, hogy legyen)
- Az eredmény örökli R típusait és attribútum neveit

- Példa:

R	A	B	C
	a	b	2
	a	c	3
	b	c	4

$\pi_A(R)$	A
	a
	b

$\pi_{C,B,B}(R)$	C	B	B
	2	b	b
	3	c	c
	4	c	c

Műveletek

Kiválasztás, szelekció

- R egy reláció $\Longrightarrow \sigma_F(R) =$ a reláció azon sorai, amelyekre az F formula teljesül.
- Teljesülni fog, hogy $\sigma_F(R) \subseteq R$
- Nincs megszorítás, csak hogy F értelmes legyen.
- Az eredmény örökli R típusait és attribútum neveit
- Példa:

R	A	B	C
	a	b	2
	a	c	3
	b	c	4

$\sigma_{A \neq B \wedge C > 2}(R)$	A	B	C
	a	c	3
	b	c	4

Az F formula:

Atomok: $A\,\theta\,B,\;\; A\,\theta\,c,\;\; c\,\theta\,B,$

ahol A,B attribútumok, c érték (konstans), $\theta \in \{<,>,=,\boxed{\leq,\geq,\neq}\}$

Építkezés: $\wedge, \vee, \neg$ Kvantorok, nincsenek!

- Példa:

  DOLGOZÓ(NÉV,CÍM,FIZE TÉS)

  $\sigma_{\text{CÍM='BP., Várna u.'}\wedge\text{FIZETÉS}>\text{'50000'}}(\text{DOLGOZÓ})$

Relációs algebra

Definíció. *Alapreláció: A bevezetés, tervezés során definiált tábla, ami meg van adva.*
A relációs algebra relációi: amik kifejezhetők az alaprelációkból $\cup, \setminus, \times, \pi, \sigma$ *segítségével.*
Származtatott reláció: nem alapreláció, de kifejezhető.
Definíció. *Egy lekérdező nyelv (igazi vagy modell) relációsan teljes, ha benne megvalósíthatók a relációs algebra alapműveletei:* $\cup, \setminus, \times, \pi, \sigma$

Ez fontos követelmény, általában tudja is mindegyik.

Inkább az a baj, hogy néha túl sokat tudnak, de nincs hatékony implementáció.

Származtatott műveletek

Hasznosak, de mivel kifejezhetők az 5 alapművelettel, ezért lényegében csak rövidítések.

Metszet

- R,S relációk $\implies R \cap S = R \setminus (R \setminus S)$ azok a sorok, amelyek mindkettőben benne vannak.
- nincs kompatibilitási követelmény $\impliedby \setminus$ tulajdonságából
- Az eredmény örökli R típusait és attribútum neveit $\impliedby \setminus$ tulajdonságából

- Példa:

R	A	B
	a	a
	a	c
	b	a

S	A	C
	a	a
	a	d
	a	c
	b	b

$R \cap S$	A	$B \cap C$
	a	a
	a	c

Származtatott műveletek

Természetes illesztés (natural join)

- $R(A_1,\ldots,A_k,B_1,\ldots,B_r), S(A_1,\ldots,A_k,C_1,\ldots,C_s)$ relációk
  $\Longrightarrow R \bowtie S =$
  - Vegyük $R \times S$-t
  - Vesszük azokat a sorokat, ahol $R.A_1 = S.A_1,\ldots,R.A_k = S.A_k$, a többit kidobjuk.
  - $\forall A_i$-ból az egyik példányt eldobjuk, azaz vetítünk $R.A_1,\ldots,R.A_k,R.B_1,\ldots,R.B_r,S.C_1,\ldots,S.C_s$-re
  - Azonos sorokat kidobjuk.
- $$R \bowtie S = \pi_{R.A_1,\ldots}(\sigma_{R.A_1=S.A_1,\ldots}(R \times S))$$

  $R \bowtie S$-nek $k+r+s$ oszlopa lesz.
  Ha nincs közös attribútum. $\Longrightarrow R \bowtie S = R \times S$.

- nincs kompatibilitási követelmény
- Az eredmény örökli R és S típusait és attribútum neveit
- Gyakorlatban ennél hatékonyabban számítjuk ki.
- Az oszlopok sorrendje nem definiált, de általában: R oszlopai, aztán S saját oszlopai.
- Példa:

R	A	B	C
	a	b	2
	a	c	3
	b	a	4

S	D	C
	a	2
	b	3
	x	2

$R \bowtie S$	A	B	C	D
	a	b	2	a
	a	b	2	x
	a	c	3	b

Miért „természetes”?

Példa: TERMELŐ(TNÉV,TERMÉK,ÁR,CÍM)

- TNÉV $\rightarrow$ CÍM
- TNÉV, TERMÉK $\rightarrow$ ÁR

Gond: TERMELŐ címét minden terméknél tároljuk
$\Longrightarrow$ redundancia + veszélyek : cím mindig kell, minden módosításhoz; könnyen sérülhet a fent megadott függés, ha elírom a címet; akkor is kell a cím, ha csak új árut akarok felvenni)

Megoldás: Inkább tároljuk két táblában:

$R = \pi_{\text{TNÉV, CÍM}}(\text{TERMELŐ})$ és
$S = \pi_{\text{TNÉV, TERMÉK, ÁR}}(\text{TERMELŐ})$

$\Longrightarrow$ TERMELŐ$=R \bowtie S$ (ha kell egyben a tábla, vissza lehet állítani)

Jó-e bármilyen felbontás?

$R' = \pi_{\text{TNÉV, CÍM, ÁR}}(\text{TERMELŐ})$ és
$S' = \pi_{\text{TNÉV, TERMÉK}}(\text{TERMELŐ})$

$\Longrightarrow$ minden terméknek ugyanannyi lesz az ára (sok ára lesz)
$\Longrightarrow$ TERMELŐ$\subsetneq R' \bowtie S'$

Az lesz majd a kérdés, hogy mik lesznek a jó felbontások?