Adatbázisok elmélete
Relációs algebra

Katona Gyula Y.

Számítástudományi és Információelméleti Tanszék
Budapesti Műszaki és Gazdaságtudományi Egyetem

Relációs adatmodell

Ahogy már szó volt róla:

Legfontosabb és leggyakoribb a létező adatmodellek közül.

Eddig:

- intuitív kép arról, hogy mi egy reláció
- hogyan kell E/K modellből relációkat előállítani

Mit tanulunk a relációs adatmodellről?

1. elvi keret: alapfogalmak, alapműveletek
2. konkrét nyelv: SQL (sémadefinícióra, adatmódosításra és lekérdezésre)
3. tervezés: minél jobb séma kialakítása, séma átalakítása, matematikai elmélet

Relációs adatmodell

Leginkább úgy gondolunk a relációra, mint egy síkbeli táblázatra:

R_1	A_1	A_2
	1	y
	1	z
	3	y

R_2	A_1	A_3
	2	y
	1	z

Igazából a sorok sorrendje nem számít és az oszlopoké (attribútumoké) sem. Egyelőre feltesszük, hogy egy sor csak egyszer szerepel (a multihalmazos lehetőségről majd az SQL-nél beszélünk).

Relációs séma vs. reláció

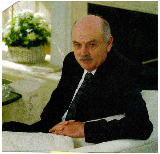
Relációs séma: amikor megadjuk, hogy mi a neve a táblázatnak és mik az oszlopai általánosan: $R(A_1, \ldots, A_n)$, ahol R a reláció neve, az A_i-k pedig az attribútumok nevei.

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

Reláció: a konkrét, kitöltött, a sémára illeszkedő táblázat (a sorok összessége)

A sémának további részei is lesznek még (a függések), a sémára illeszkedő relációnak ezeket is be kell tartania.

Relációs modell



Edgar F. Codd, (1932–2003)

1970-es cikk: *A Relational Model of Data for Large Shared Data Banks*

Teljes adatmodell: nem csak azt mondja meg hogyan írok le az adatot, 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 $\rightarrow$ reláció
$\Longrightarrow$ gond nélkül egymásba ágyazhatók

Unió

- R, S relációk $\Longrightarrow R \cup S$: R és S sorai együtt
  Azonos sorok csak egyszer szerepelnek. (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

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

Szorzat (direkt szorzat, Descartes szorzat)

- $R(A_1,\ldots,A_k), S(B_1,\ldots,B_\ell)$ k ill. ℓ attribútumos relációk $\Longrightarrow R \times S : k+\ell$ 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.
- Az unió és különbség könnyű művelet, a szorzat nehezebb. Vigyázni kell mennyit használjuk.

Szorzat, 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

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

Kiválasztás, szelekció

- R egy reláció $\implies \sigma_F(R) =$ a reláció azon sorai, amelyekre az F formula teljesül.
- Teljesülni fog, hogy $\sigma_F(R) \subseteq R$
- Nincs más megszorítás, csak hogy F értelmes legyen (erről mindjárt).
- 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,FIZETÉS)

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

Relációs algebra, fogalmak

- 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.
- Fontos fogalom: egy lekérdező nyelv (igazi vagy modell) relációsan teljes, ha benne megvalósíthatók a relációs algebra alapműveletei: $\cup, \setminus, \times, \pi, \sigma$

Ez utóbbi 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 öt alapművelettel, ezért lényegében csak rövidítések.

Metszet

- R, S relációk $\Longrightarrow R \cap S : R \setminus (R \setminus S)$ azok a sorok, amelyek mindkettőben benne vannak.
- nincs kompatibilitási követelmény ($\setminus$ tulajdonságai miatt)
- Az eredmény örökli R típusait és attribútum neveit ($\setminus$ tulajdonságai miatt)

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

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.
- $R \bowtie S = \pi_{R.A_1,\ldots}(\sigma_{R.A_1=S.A_1,\ldots}(R \times S))$

Természetes illesztés (natural join)

- $R \bowtie S = \pi_{R.A_1,\ldots}(\sigma_{R.A_1=S.A_1,\ldots}(R \times S))$
- $R \bowtie S$-nek hány oszlopa lesz? $k + r + s$
  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

Példák relációs algebra alkalmazására

ÁRU(ÁRUKÓD, ÁRUNÉV, EGYSÉGÁR)
ELADVA(DÁTUM, ÁRUKÓD, DB)
BEVÉTEL(DÁTUM, ÖSSZEG)
BEFIZ(ÖSSZEG, BEFIZ) BEFIZ=ÖSSZEG$-$4000

A 2017. jan. 1. utáni napok bevételei a dátummal együtt:

$$\sigma_{\text{DÁTUM} > \text{'2017-01-01'}}\left(\text{BEVÉTEL}\right)$$

A 2017. jan. 15-i befizetett összeg és bevétel:

$$\pi_{\text{ÖSSZEG, BEFIZ}}\left(\sigma_{\text{DÁTUM='2017-01-15'}}\left(\text{BEVÉTEL} \bowtie \text{BEFIZ}\right)\right)$$

$$\pi_{\text{ÖSSZEG, BEFIZ}}\left(\sigma_{\text{DÁTUM='2017-01-15'}}\left(\text{BEVÉTEL}\right) \bowtie \text{BEFIZ}\right)$$

Példa még

Hány darabot adtak el 2017. jan. 15-én az A123 kódú áruból, mi a neve és az ára?

$$\pi_{\text{DB, ÁRUNÉV, EGYSÉGÁR}}\left(\sigma_{\text{ÁKÓD='A123'} \wedge \text{D='2002-01-15'}}\left(\text{ELADVA} \bowtie \text{ÁRU}\right)\right)$$

$$\pi_{\text{DB, ÁRUNÉV, EGYSÉGÁR}}\left(\sigma_{\text{ÁKÓD='A123'} \wedge \text{D='2002-01-15'}}(\text{ELADVA}) \bowtie \text{ÁRU}\right)$$

ÁKÓD ÁRUKÓD-ot, D pedig DÁTUM-ot jelenti, csak rövidítenem kellett, különben nem fért volna ki :)

Miért „természetes” illesztés?

Példa: TERMELŐ(TermelőNév,Termék,Ár,Cím)

- TermelőNév $\to$ Cím
- TermelőNév, Termék $\to$ Ár

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

Megoldás: Inkább tároljuk két táblában:
$R = \pi_{\text{TermelőNév, Cím}}(\text{TERMELŐ})$ és
$S = \pi_{\text{TermelőNév, Termék, Ár}}(\text{TERMELŐ})$
$\Longrightarrow$ TERMELŐ=$R \bowtie S$ (ha kell egyben a tábla, vissza lehet állítani természetes módon)

Kitérő: Jó-e bármilyen felbontás?

$R' = \pi_{\text{TermelőNév, Cím, Ár}}(\text{TERMELŐ})$ és
$S' = \pi_{\text{TermelőNév, Termék}}(\text{TERMELŐ})$

$\Longrightarrow$ minden terméknek ugyanolyan árai lesznek (sok ár lesz)

$\Longrightarrow \text{TERMELŐ} \subsetneq R' \bowtie S'$

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

Származtatott művelet még: bal (jobb) félillesztés

- $R(A_1, \ldots, A_k, B_1, \ldots, B_r), S(A_1, \ldots, A_k, C_1, \ldots, C_s)$ relációk
  $\Longrightarrow R \rtimes S =$ R azon sorai, amelyhez vannak passzoló sorok S-ben
  $R \rtimes S = \pi_R(R \bowtie S)$
- $R \rtimes S \subseteq R$
- $R \ltimes S =$ ugyanez jobbról
-

$R \rtimes S$	A	B	C
	a	b	2
	a	c	3

$R \ltimes S$	D	C
	a	2
	b	3
	x	2

Származtatott művelet még: θ-illesztés

- R, S relációk
  $\Longrightarrow R \underset{R.A_i \theta S.B_j}{\bowtie} S = R \times S$ azon sorai, amelyben az adott oszlopok θ relációban vannak
  $R \underset{R.A_i \theta S.B_j}{\bowtie} S = \sigma_{R.A_i \theta S.B_j}(R \times S)$
- Példa:

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

S	D	E
	a	2
	b	3
	x	2

$R \underset{C \leq E}{\bowtie} S$	A	B	C	D	E
	a	b	2	a	2
	a	b	2	b	3
	a	b	2	x	2
	a	c	3	b	3

Példa még

Mely nevű áruk azok, amelyekkel van azonos egységárú másik áru?

Itt az ÁRU reláció két sorát kell összevetni.

Átnevezés:

- Technikai segítség, ha pl. két relációban ugyanolyan attribútumnév van, és direkt szorzatot akarunk. Nem változtatja meg a reláció sorait, csak az attribútumok és a reláció nevét, ezért nem igazi művelet.
- $R(A_1, \ldots, A_n)$ egy reláció
  $\Longrightarrow \rho_{S(B_1,\ldots,B_n)}(R) =$ sorai megegyeznek R soraival, a reláció neve S, attribútumai rendre $B_1, \ldots, B_n$.
- Ha csak a relációt akarjuk átnevezni: $\rho_S(R)$

Megoldás az előbbi kérdésre

$$\text{ÁRU1} = \rho_{\text{ÁRU1(ÁRUKÓD1, ÁRUNÉV1, EGYSÉGÁR1)}}(\text{ÁRU})$$

$$\text{ÁRU2} = \rho_{\text{ÁRU2(ÁRUKÓD2, ÁRUNÉV2, EGYSÉGÁR2)}}(\text{ÁRU})$$

$$\text{ÁRU3} = \text{ÁRU1} \underset{\text{EGYSÉGÁR1 = EGYSÉGÁR2} \,\wedge\, \text{ÁRUKÓD1} \neq \text{ÁRUKÓD2}}{\bowtie} \text{ÁRU2}$$

$$\text{ÁRU4} = \pi_{\text{ÁRUNÉV1}}\left(\text{ÁRU3}\right)$$

További példák

TERMÉK(GYÁRTÓ, MODELL, TÍPUS)

PC(MODELL, SEBESSÉG, MEMÓRIA, MEREVLEMEZ, CD, ÁR)

LAPTOP(MODELL, SEBESSÉG, MEMÓRIA, MEREVLEMEZ, KÉPERNYŐ, ÁR)

NYOMTATÓ(MODELL, SZÍNES, TÍPUS, ÁR)

A relációk jelentése

TERMÉK: az adott nevű gyártó gyártja az adott modellszámú és adott típusú (PC, Laptop vagy nyomtató) terméket

PC: modellszám, sebesség megaHz-ben, memória gigabájtban, merevlemez gigabájtban, a CD sebessége (pl. 4x), az ár

LAPTOP: mint PC-nél, plusz a képernyő mérete hüvelykben

NYOMTATÓ: modellszám, színes-e (i/n), típusa (tintasugaras, lézer, mátrix), ára

A modellszámokról feltesszük, hogy egyediek.

Kérdések

Melyek azok a PC modellek, amelynek sebessége legalább 150?

$$\pi_{\text{MODELL}}\left(\sigma_{\text{SEBESSÉG} >= 150}(\text{PC})\right)$$

Mely gyártók készítenek legalább egy gigás merevlemezű laptopot?

$$\pi_{\text{GYÁRTÓ}}\left(\text{TERMÉK} \bowtie \sigma_{\text{MEREVLEMEZ} >= 1}(\text{LAPTOP})\right)$$

Kérdés még

Adjuk meg a B gyártó által gyártott összes termék modellszámát és árát típustól függetlenül!

$$\pi_{\text{MODELL, ÁR}}\left(\sigma_{\text{GYÁRTÓ='B'} \wedge \text{TÍPUS = 'PC'}}\left(\text{TERMÉK}\right) \bowtie \text{PC}\right) \cup$$
$$\pi_{\text{MODELL, ÁR}}\left(\sigma_{\text{GYÁRTÓ='B'} \wedge \text{TÍPUS = 'LAPTOP'}}\left(\text{TERMÉK}\right) \bowtie \text{LAPTOP}\right) \cup$$
$$\pi_{\text{MODELL, ÁR}}\left(\sigma_{\text{GYÁRTÓ='B'} \wedge \text{TÍPUS = 'NY'}}\left(\text{TERMÉK}\right) \bowtie \text{NYOMTATÓ}\right)$$

Kérdés még

Melyek azok a gyártók, akik laptopot gyártanak, de PC-t nem?

$$\text{TERMÉK1} = \rho_{\text{TERMÉK1}}(\pi_{\text{GYÁRTÓ, TÍPUS}}\left(\text{TERMÉK}\right))$$

$$\pi_{\text{GYÁRTÓ}}\left(\sigma_{\text{TÍPUS='LAPTOP'}}\left(\text{TERMÉK1}\right)\right) \setminus \pi_{\text{GYÁRTÓ}}\left(\sigma_{\text{TÍPUS='PC'}}\left(\text{TERMÉK1}\right)\right)$$

Utolsó kérdés

Melyek azok a gyártók, amelyek gyártanak legalább két, egymástól különböző, legalább 133 Mhz-en működő PC-t vagy Laptopot? (Nincs két azonos modellszám!)

$$R1 = \pi_{\text{MODELL, SEBESSÉG}}\,(\text{PC}) \cup \pi_{\text{MODELL, SEBESSÉG}}\,(\text{LAPTOP})$$

$$R2 = \pi_{\text{GYÁRTÓ, MODELL}}\left(\sigma_{\text{SEBESSÉG>=133}}\,(\text{R1}) \bowtie \text{TERMÉK}\right)$$

$$R3 = \rho_{R3(\text{GYÁRTÓ2, MODELL2})}(R2)$$

$$R4 = R2 \underset{\text{GYÁRTÓ = GYÁRTÓ2} \wedge \text{MODELL <> MODELL2}}{\bowtie} R3$$

$$R5 = \pi_{\text{GYÁRTÓ}}\,(\text{R4})$$

További műveletek

Az eddigi műveletek kifejezhetők voltak a relációs algebra alapműveleteivel. Amiket most mutatok, azok nem, de fontosak (SQL folyton használ ilyeneket, ott majd részletesen is nézzük).

- aggregátumok: MIN, MAX, AVG, SUM, CNT (darabszám)
  Pl. leggyorsabb gép, átlagár, hányféle printer
  eredmény mindig egy szám
- aggregátum csoportosítva: Bizonyos feltételek szerinti partíciókban aggregátumok.
  Pl. átlagos ár tintasugaras nyomtatók között, egy gyártónak hány terméke van
  $\Longrightarrow$ eredmény egy reláció pl. (gyártó, szám) párokból.

További műveletek még: rekurzív lezárás

- hagyományos adatkezelésben ritka, intelligensebb rendszerekben inkább előfordul)
- Pl. reláció: ki főnöke kinek $\Longrightarrow$ lezárás: ki felettese kinek
- Pl. reláció: melyik városból melyikbe van repülő járat $\Longrightarrow$ lezárás: átszállással el lehet-e jutni
- Ezt a relációs algebra nem tudja, csak fix mélységre: pl. max 4 átszállás,

A NULL érték, emlékeztető

Lehet, hogy vannak kitöltetlen mezők, ezt meg akarjuk engedni: NULL érték. 2 alapvető értelmezés (majd SQL-nél lesz, hogy hogyan kell megmondani, hogy melyik van éppen, illetve, hogy lehet-e egyáltalán NULL valahol):

- $\nexists$
- $\exists$, de nem ismerjük.

Attól függően, hogy hogyan értelmezzük a NULL-t:

Mi legyen egy ilyen kérdéssel?:
Pl. $\pi_{\text{CÍM='BP'}}\left(\text{TERMELŐ}\right)$

Ilyenkor belevegyük-e ha a cím NULL?

Külső illesztés (outer join)

R, S relációk $\Longrightarrow R ⟕ S$ **bal külső illesztés**: $R \bowtie S$-hez azokat az R-beli sorokat is hozzávesszük, amihez nem illeszkedik S-beli. Hiányzó helyekre NULL kerül.

Pl. SZEMÉLY(NÉV, KÓD), CÍM(KÓD, CÍM)
SZEMÉLY $\bowtie$ CÍM $\Longrightarrow$ akinek nincs címe nem lesz rajta
SZEMÉLY ⟕ CÍM $\Longrightarrow$ kiderül, kinek nincs meg a címe

Külső illesztés

SQL-ben van, relációs algebrával elvileg nem fejezhető ki (NULL miatt), de elkerülhető.

Lényegében: $(R \bowtie S) \cup (R \setminus (R \ltimes S))$

Van jobb külső illesztés is: $R ⟖ S$

Teljes külső illesztés: $R ⟗ S := (R ⟕ S) \cup (R ⟖ S)$

Példa

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

S	*D*	*C*
	a	2
	b	3
	x	2
	y	1

R ⟕ *S*	*A*	*B*	*C*	*D*
	a	*b*	2	*a*
	a	*b*	2	*x*
	a	*c*	3	*b*
	b	**a**	**4**	NULL

R ⟖ *S*	*A*	*B*	*C*	*D*
	a	*b*	2	*a*
	a	*b*	2	*x*
	a	*c*	3	*b*
	NULL	NULL	**1**	**y**

R ⟗ *S*	*A*	*B*	*C*	*D*
	a	*b*	2	*a*
	a	*b*	2	*x*
	a	*c*	3	*b*
	b	**a**	**4**	NULL
	NULL	NULL	**1**	**y**

Külső unió

Részben kompatibilis relációk egyesítésére:
DIÁK(NÉV, TÉMAVEZ, TSZK)
TANÁR(NÉV, TSZK, BEOSZT)

DIÁK $\cup_k$ TANÁR

	NÉV	TSZK	TÉMAVEZ	BEOSZT
diák				NULL
tanár			NULL	

Multihalmazos szemantika

A relációs algebrában ugyan minden reláció halmaz, ezért nincsenek többszörös sorok, de pl. SQL-nél lesznek. A multihalmazokkal kicsit máshogy vannak a halmazműveletek:

Ha a t sor $m_R(t)$ példányban van meg R-ben és $m_S(t)$ példányban van meg S-ben, akkor

- $m_{(R \cup S)}(t) := m_R(t) + m_S(t)$ példányban lesz meg R és S uniójában
- $m_{(R \cap S)}(t) := \min\{m_R(t), m_S(t)\}$ példányban lesz meg R és S metszetében
- $m_{(R \setminus S)}(t) := \max\{m_R(t) - m_S(t), 0\})$ példányban lesz meg $R \setminus S$-ben