

Relációs tervezés – fogalmak és algoritmusok

Szárnyas Gábor (szarnyas@db.bme.hu)

Kulcsok és szuperkulcsok

Némi félreértésre adhat okot a „szuper” kifejezés alkalmazása. Míg a szupersportautó (*supercar*) egyben autó (*car*) is, a szuperkulcs (*superkey*) **nem** kulcs (*key*).

A szuper (*super*) szó jelentése ebben az esetben tartalmazó halmaz (*superset*), ezzel összhangban a szuperkulcs olyan halmaz, ami tartalmaz kulcsot.

A kulcs minimalitása

Amikor arról beszélünk, hogy a kulcs minimális szuperkulcs, arra utalunk, hogy attribútumok elhagyásával tovább nem csökkenthető. Ez nem jelenti, hogy az elemszámát tekintve is minimális lenne. A kapcsolódó angol kifejezések:

- *Minimal*: tovább nem csökkenthető.
- *Minimum*: legkisebb elemszámú.

Ebben az értelemben *minimal* szuperkulcsokat keresünk. Például:

$$R(A, B, C)$$
$$F = \{AB \rightarrow C, C \rightarrow AB\}$$

A sémának egy-egy kulcsa az AB és a C attribútumhalmaz, mert egyik sem csökkenthető.

A függéshalmaz minimalitása

A kulcshoz hasonlóan, a minimális függéshalmaz sem csökkenthető tovább, de nem feltétlenül a legkisebb elemszámú.

$$F = \{A \rightarrow B, B \rightarrow A, B \rightarrow C, C \rightarrow A, AB \rightarrow C\}$$

Az F függéshalmazból az alábbi minimális függéshalmazok állíthatók elő:

- $F_{\min 1} = \{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$
- $F_{\min 2} = \{A \rightarrow B, A \rightarrow C, B \rightarrow A, C \rightarrow A\}$

Kulcskeresés

Vizsgáljuk meg, hogy egy attribútum hol szerepel a függéshalmazban szereplő függőségekben. A következő lehetőségek vannak:

- **b**: csak függőség baloldalán szerepel,
- **m**: függőség bal- és jobboldalán is szerepel,
- **j**: csak függőség jobboldalán szerepel,
- **e**: nem szerepel függőségben.

Nézzünk egy példát:

$$R(A, B, C, D, E, G, H)$$
$$F = \{AB \rightarrow C, CE \rightarrow DH, D \rightarrow E\}$$

A kitöltött táblázat:

b	m	j	e
<i>A, B</i>	<i>C, D, E</i>	<i>H</i>	<i>G</i>

A csak baloldalon (b) és az egyik oldalon sem szereplő attribútumok (e) minden kulcsnak elemei lesznek, hiszen az ő értékeit semelyik másik attribútum értéke sem határozza meg funkcionális függés szerint:

$$(ABG)^+(F) = ABGC$$

Látható, hogy nem kaptunk szuperkulcsot. Ilyenkor a mindkét oldalon (m) szereplő attribútumokkal kell próbálkoznunk.

$$(ABGC)^+(F) = ABGC$$

$$(ABGD)^+(F) = ABGDCEH$$

$$(ABGE)^+(F) = ABGECDH$$

ABGD és *ABGE* szuperkulcsok. Tudjuk, hogy *ABG* mindenképpen eleme a kulcsnak, viszont önmagában nem alkot kulcsot, ezért *ABGD* és *ABGE* minimális szuperkulcsok, tehát kulcsok is.

Megjegyzés: Ha egyik sem lett volna szuperkulcs, folytattuk volna a próbálkozást a *CD*, *CE* és *DE* attribútumok hozzávételével. Ha így sem kaptunk volna szuperkulcsot, akkor az *CDE* attribútumokkal együtt biztosan szuperkulcsot kapunk, hiszen a kapott attribútumhalmazban már csak olyan attribútumok (*H*) nem szerepelnek, amelyek kizárólag függőség jobb oldalán található meg. A csak jobb oldalon (*j*) szereplő attribútumok pedig egyetlen kulcsnak sem lesznek elemei, hiszen az ő valamely determinánsuk bekerülése után bővíthető velük az attribútumhalmaz lezártkezdeménye.

Normálformák ellenőrzése

Egy séma normálformáját az alábbi módon érdemes végezni?

1. 1NF? Ha nem, készen vagyunk, a séma 0NF.
2. BCNF? Ha igen, készen vagyunk, a séma BCNF.
3. 3NF? Ha igen, készen vagyunk, a séma 3NF.
4. 2NF? Ha igen, készen vagyunk, a séma 2NF.
5. Különben a séma 1NF.

A BCNF és a 3NF tulajdonságok ellenőrzésénél célszerű a szuperkulcsos definícióval dolgozni, amit elegendő a megadott függéshalmaz függéseire ellenőrizni (tehát nem kell a függéshalmaz lezártját vizsgálni).

2NF vizsgálatot az összes kulcs valódi részhalmazainak, mint attribútumhalmazoknak a lezártjával érdemes vizsgálni: ha elérünk másodlagos attribútumot, akkor sérül a 2NF feltétel.