

Adatbázisok

Normálformák

Szárnas Gábor
szarnyas@db.bme.hu

2012. augusztus 5.

1. Definíciók tömören

- 2NF: 1NF és $\forall K$ kulcs $\forall A \in R$ másodlagos: A teljesen függ K -től.
- 3NF/i.: 1NF és $\forall K$ kulcs $\forall A \in R$ másodlagos: A nem függ tranzitívan K -től.
- BCNF/i.: 1NF és $\forall K$ kulcs $\forall A \in R$: A nem függ tranzitívan K -től.
- 3NF/ii.: 1NF és $\forall X \subseteq R \ \forall A \in R, X \rightarrow A$ nemtriviális: X szuperkulcs $\forall A$ elsődleges.
- BCNF/ii.: 1NF és $\forall X \subseteq R \ \forall A \in R, X \rightarrow A$ nemtriviális: X szuperkulcs.

2. Normálformák ellenőrzése

Tipikus feladat, hogy egy adott R sémáról, a sémán értelmezett F függéshalmaz mellett állapítsuk meg, hogy hányadik normálformában van. Feltételezzük, hogy a séma attribútumai atomiak, tehát a séma 1NF. Célszerű a vizsgálatot az erősebb, szigorúbb feltételeket előíró normálformákkal kezdve végezni. Ekkor, ha azt találtuk, hogy a séma egy adott normálformában van, nem kell tovább vizsgálódnunk (pl. ha a séma 3NF, akkor nem kell megvizsgálunk, hogy 2NF-e, mert ezt tétel garantálja), ráadásul először a könnyebb vizsgálatokat kell elvégeznünk.

BCNF Belátható, hogy elég az F -beli nemtriviális függőségek bal oldalának lezártját kiszámítani [5]: ha mindegyik visszaadja a teljes R sémát (azaz szuperkulcs), akkor a séma BCNF, különben nem. Ez a lépés gyorsan végrehajtható, mert nem kell meghatározni a séma kulcsait.

3NF Határozzuk meg a kulcsokat és a kulcsokból az elsődleges attribútumokat. Itt is elég az F -beli, $X \rightarrow A$ ($X \subseteq R, A \in R$) alakúra hozott nemtriviális függőségekre megvizsgálni, hogy a bal oldal szuperkulcs és/vagy a jobboldalon elsődleges attribútum áll: ha mindegyik függőség ilyen, akkor a séma 3NF, különben nem.

2NF Itt már nem elég F vizsgálata.¹ Egy lehetséges, de nagyon költséges megoldás az F^+ -beli nemtriviális függőségek vizsgálata. Ehelyett célszerű minden másodlagos attribútumra megvizsgálni, hogy függ-e bármely kulcs bármely *valódi* részalmazától: ehhez vizsgáljuk meg a kulcsok valódi részalmazainak lezártját. Ha mindig nemleges választ kapunk, a séma 2NF, különben 1NF.²

¹Pl. $R(ABCD), F = \{B \rightarrow C, C \rightarrow B, BC \rightarrow D\}$ esetén a kulcsok halmaza $\{AB, AC\}$, így az elsődleges attribútumok ABC , az egyetlen másodlagos attribútum D . Egyik F -beli függőség sem sérti a 2NF feltételt, pedig a D másodlagos attribútum függ B -től és C -től is, amelyek a kulcs részalmazai.

²A 2NF ellenőrzéséhez átlagosan szükséges lépések számát csökkenthetjük, ha először megvizsgáljuk az F -beli függőségekre a definíció teljesülését: ha találunk olyan függőséget, amely sérti a feltételt (a bal oldalon kulcs részalmazza áll, a jobb oldalon másodlagos attribútum), akkor a séma 1NF. A kulcsok valódi részalmazainak vizsgálatára csak akkor van szükség, ha tudjuk, hogy a séma nem 3NF és egyik F -beli függősége sem sérti a feltételt.

3. Történet

A 2NF és a 3NF/i. definíciót Edgar F. Codd publikálta 1971-ben [1]. A BCNF-et Raymond F. Boyce és Codd alkotta meg a 3NF táblákban előforduló anomáliák megszüntetésére 1974-ben [2] a BCNF/ii. definícióval. Érdekeség, hogy a koncepció már Ian Heath 1971-es cikkében megjelent [3]. A 3NF második definícióját Carlo Zaniolo fogalmazta meg 1982-ben [4]:

A relation R is BCNF if for every nontrivial FD³ of R , $X \rightarrow A$, X is a superkey for R .
[...]

LEMMA 3. A relation R is 3NF iff⁴ for every nontrivial FD of R , $X \rightarrow A$,

- (a) X is a superkey for R , or
- (b) A is a key attribute for R .

[...]

The preceding lemmas provide new and simpler definitions of 3NF. Most important, *they reveal an analogy between the definitions of 3NF and BCNF which was not previously known*: the two definitions are identical except for (b), which waives condition (a) for key attributes. Thus, if we keep condition (a) unchanged but relax (b), we obtain a normal form definition which is stricter than 3NF but weaker than BCNF.

További normálformák: Ronald Fagin 1977-ben megalkotta a 4NF-et a többértékű függőségek (*multivalued dependency*), 1979-ben az 5NF-et az elágazó függőségek (*join dependency*) okozta redundancia megszüntetésére. További normálformák: Fagin, DKNF (Domain/key normal form), 1981; Date–Darwen–Lorentzos, 6NF, 2002.

Hivatkozások

- [1] Codd, E.F. *Further Normalization of the Data Base Relational Model.*, IBM Research Report RJ909 (August 31st, 1971). Republished in Randall J. Rustin (ed.), *Data Base Systems: Courant Computer Science Symposia Series 6*. Prentice–Hall, 1972.
- [2] Codd, E. F. *Recent Investigations into Relational Data Base Systems*. IBM Research Report RJ1385 (April 23rd, 1974). Republished in *Proc. 1974 Congress (Stockholm, Sweden, 1974)*. New York, N.Y.: North–Holland (1974).
- [3] Heath, I. *Unacceptable File Operations in a Relational Database*. *Proc. 1971 ACM SIGFIDET Workshop on Data Description, Access, and Control*, San Diego, Calif. (November 11th–12th, 1971).
- [4] Zaniolo, Carlo. *A New Normal Form for the Design of Relational Database Schemata*. *ACM Transactions on Database Systems* 7(3), September 1982.
- [5] *Informatikai algoritmusok 1.*, Demetrovics János, Sali Attila, *Relációs adatmodellek tervezése*, ELTE Informatikai Kara, 2005., <http://compalg.inf.elte.hu/~tony/Elektronikus/Informatikai/>

³functional dependency

⁴if and only if