

Abimaterjal: funktsionaalsed sõltuvused

1. Miks on andmebaasi loomisel vaja funktsionaalsete sõltuvuste aparatuuri?

Probleem: andmebaasi disainerid saavad kokku panna halbu relatsioone. Halb on selline relatsioon, mis tekitab andmeliiasust (üks reaalelu fakt on andmebaasis korduvalt), lisamisanomaaliaid (kui me ei saa andmebaasi lisada andmeid mingi objekti kohta kuni me ei tea mingeid muid andmeid), eemaldamisanomaaliaid (kui me mingi olemieksemplari andmebaasist eemaldamise tõttu kaotame ka mingi muu informatsiooni). Ja andmeliiasus on potentsiaalne vastuoluliste andmebaasiseisundite allikas, kui andmeid parandades jäävad osa andmeid parandamata ja üks reaalmaailma fakt on baasi erinevate osades erinevate väärtustega. Näiteks sobib eksamitulemuste relatsioon, mille kirjeldus olgu selline:

Eksamitulemused (Tudengi matr.nr, Tudengi eesnimi, Tudengi perenimi, Aine nimetus, Aine kood, Aine maht, Eksami kuupäev, Hinne). (*)

See on väga halb relatsioon. Fakt, et antud matrikli numbriga isikul on antud ees- ja perenimi, on siin iga eksamitulemuse juures. Kui näiteks kellegi perenimi muutub (abiellumine), siis tuleb see parandada kõigi eksamite juures. Tudengite andmeid (ees- ja perekonnanimesid) ei saa siia baasi lisada enne, kui nad pole mingit eksamit teinud (lisamisanomaalia). Kui miskipärast tudengil, kes on juba 2 ainet sooritanud, eksamitulemused tuleb annuleerida ja baasist eemaldada, kaotame ka selle tudengi andmed (matr.nr, ees- ja perenime) – eemaldamisanomaalia.

Kogenud andmebaasi disainer näeb kohe, et sellise relatsiooniga tekivad probleemid. Küsimus on aga selles, kas me suudame luua teooria, mille abil saaks algoritmiliselt määrata relatsiooni headust ja ka halbade relatsioonide puhul saaks algoritm ette anda, kuidas olukorda parandada.

Selline teooria ongi loodud. Aga siiani on meil ainuke teave relatsiooni kohta tema kirjeldus. Sellest ei piisarelatsiooni headuse üle otsustamiseks. Vajame lisainformatsiooni, mis esitaks reaalsuse teatud kitsendusi. Selliseks vajalikuks lisainformatsiooniks on funktsionaalsete sõltuvuste kirjeldamine.

2. Funktsionaalsed sõltuvused

Olgu antud relatsioon $R(A_1, A_2, \dots, A_n)$ ja olgu $X, Y \subseteq \{A_1, A_2, \dots, A_n\}$ mingid atribuutide hulgad selles relatsioonis.

Def. Funktsionaalne sõltuvus. Me ütleme, et atribuutide hulk X relatsioonis R määrab funktsionaalselt atribuutide hulga Y , kirjutame $X \rightarrow Y$, kui relatsiooni R mistahes seisundi korral **ei leidu** kaht sellist rida (korteeži), mis langeksid kokku atribuutide X väärtustel, aga ei lange kokku Y väärtustel. (**)

Funktsionaalsed sõltuvused tuleb leida reaalsuse püsiomadustena, nad defineeritakse kui relatsiooni mistahes seisundi omadused. Kui relatsioon aja jooksul muutub, siis need funktsionaalsed sõltuvused peaks jääma alles. Vaatame eksamitulemuste näidet punktis 1 – ta on tekstis tähistatud (*).

On selge, et igale matriklinumbrile vastab ainult üks võimalik ees- ja perekonnanimi. See tähendab, et tunnuse matrikli number määrab tunnused eesnimi ja perenimi. Kirjutame selle nii:

Tudengi matr.nr → Tudengi eesnimi, Tudengi perenimi

Vaatame definitsiooni (**). Näites on X rollis tunnus Tudengi matr.nr ja Y rollis kaks tunnust: Tudengi eesnimi ja Tudengi perenimi. Ükskõik kui palju meie relatsiooni seisundis ka tudengeid ei oleks, samale matriklinumbrile peab alati vastama sama ees- ja perenimi.

Relatsioonis (*) kehtib veel mitu funktsionaalset sõltuvust. Näiteks:

Aine kood → Aine nimetus, Aine maht

Ka siin on X osas üksainus tunnus ja Y osas 2 tunnust. See funktsionaalne sõltuvus tähendab, et sama koodiga aine on alati sama nimega (ükskõik kelle eksamitulemuste hulgas ta siis on) ja ka tema maht on alati sama. Ei leidu kahte sellist rida, kus Aine kood oleks sama, aga Aine maht oleks erinev (vt. definitsiooni (**)).

Lisaks neile seni vaadatud funktsionaalsetele sõltuvustele kehtib näite (*) mistahes seisundil veel selline funktsionaalne sõltuvus:

Tudengi matr.nr, Aine kood, Eksami kuupäev → Hinne

See tähendab, et sama tudeng, samas aines samal kuupäeval ei või saada kahte erinevat hinnet (vt. definitsioon (**)). Nagu me teame, siis kaks hinnet saab tudengil olla ühes aines vaid siis, kui esimene kord antud aines saab ta hindeks F. Aga see funktsionaalne sõltuvus ütleb, et samal päeval ta järeleksamit teha ja tulemust parandada ei saa.

3. Funktsionaalsete sõltuvuste tähendus

Juba punktis 2. Vaatasime mitmeid funktsionaalseid sõltuvusi ja vastavalt definitsioonile (**) väitsid nad mingeid kitsendusi reaalsusele. Kuigi relatsiooni definitsioonis võis olla relatsiooni seisundis mistahes alamhulk tunnustele vastavate väärtushulkade otsekorrutisest, siis osa nendest alamhulkadest ei esinda reaalsus võimalikke olekuid. Selline rida relatsioonis, kus samale matriklinumbrile vastab kord üks, kord teine ees- ja perekonnanimi, ei ole reaalsuses võimalik. Samuti ei ole võimalik, et tudeng teeb samal kuupäeval kaks korda ühes aines eksamit, saab esimesel korral F, aga teisel korral mingi muu tulemuse. Funktsionaalsed sõltuvused kirjeldavad reaalsuse kitsendusi ja andmete lisamisel andmebaasi aitavad vältida vigu, kus nende kitsenduste vastu eksitakse.

Et harjutada funktsionaalsete sõltuvuste tähendustest arusaamist, vaatame mõnda funktsionaalset sõltuvust, mis ei kehti ja uurime, mida nad võiksid tähendada.

Näide mittekehtivast funktsionaalsest sõltuvusest.

Oletame, et meil peaks kehtima funktsionaalne sõltuvus: Aine maht → Aine kood. See tähendaks, et meil ei saa olla kahte sama mahuga (ainepunktide mõttes) ainet. Igale ainemahule peaks vastama ainult üks võimalik aine kood. Et kui 6 EAP on aine Andmebaasid, siis muid 6-punktiseid aineid olla ei tohiks. Loomulikult selline kitsendus ei vasta tegelikkusele.

Teise mittekehtiv funktsionaalse sõltuvuse näide: Eesnimi → Tudengi matr.nr. Selline sõltuvus tähendab, et iga eesnimi tohib meie andmebaasis esineda tudengite seas vaid üks kord. Kui meil on juba üks tudeng nimega „Mati“, siis teist Matit enam meie andmebaasis olla ei tohi, sest eesnimele pidi kogu baasis vastama üks ja ainult üks matrikli number, kahel erineval Matil on aga ju erinevad matriklinumbrid.

Kokkuvõttes: funktsionaalsete sõltuvustega tuleb ettevaatlikult ümber käia, sest nad väljendavad kitsendusi reaalsusele. Ja me tahame, et need kitsendused vastaksid tegelikkusele ja aitaksid meil andmebaase disainida.