

Normalizacija iki trečios normalinės formos

Normalizacija yra duomenų transformacija į natūraliai sugrupuotą ir gerai suformuotą formą taip, kad vienas faktas būtų vienoje vietoje ir ryšiai tarp faktų būtų tvarkingi.

Normalizacijos procesas yra įvairių taisyklių pritaikymas reliaciniam modeliui, siekiant supaprastinti sąryšius ir sudaryti galimybę netrukdomai juos modifikuoti.

Priimta normalizuoti sąryšius iki trečios normalinės formos. (Skiriama ir tolesnė normalizacija, bet tai išeina už šio kurso ribų).

Pirma normalinė forma

Reikia užtikrinti, kad visi atributai būtų atominiai (t.y. mažiausios įmanomos komponentės). Kitaip tariant, kiekvieno domeno atribute yra galima tik viena reikšmė, o ne reikšmių rinkinys. Tai dažnai išreiškiama reikalavimu, kad sąryšis negali turėti pasikartojančių grupių.

Pailiustruokime tai pavyzdžiu. Sakykime, turime 2 kursus: B11 ir M15.

- Kursas B11 yra bakalauro studijos (Bs), jo pavadinimas „Informatika“ ir jame dėstomi 5 moduliai: B141, B142, B143, B144, B145. Moduliai B141, B142, B143 yra baziniai ir verinami 8 kreditais, o moduliai B143, B144 yra specializacijos ir vertinami 11 kreditų.
- Kursas M15 yra magistro studijos (Ms), jo pavadinimas „Statistika“ ir jame dėstomi 3 moduliai: M215, M216, B141. Moduliai M215 ir M216 yra magistratūros studijų ir vertinami 15 kreditų, o modulis B141, kaip jau buvo minėta, yra bazinis (8 kreditai).

Aprašyta informacija gali būti patalpinta žemiau pateiktoje lentelėje.

Lentelė 1. Pradinė lentelė (iki pirmos normalinės formos)

KURSAS	K_PAVAD	LYGIS	MODULIS	M_PAVADINIMAS	STATUSAS	KREDITAI
B11	Informatika	Bs	B141	Programavimas 1	bazinis	8
			B142	Duomenų bazės		
			B143	Matematinė analizė		
			B144	C++ pagrindai	specializacijos	11
			B145	Tinklų administravimas		
M15	Statistika	Ms	M215	Statistikos teorija	magistratūros	15
			M216	Ekonomika		
			B141	Programavimas 1	bazinis	8

Pagrindinis tokio duomenų saugojimo būdo trūkumas yra tai, kad tokioje lentelėje įrašų eiliškumas turi prasmę. T.y. mes negalime keisti vietomis šitos lentelės eilučių, nes tai gali iškraipyti informaciją. Pavyzdžiui, jeigu priešpaskutinę lentelės 1 eilutę perkelsimė į antros eilutės poziciją, gausis, kad modulis M216 skaitomas kursui B11.

Reliaciniame duomenų modelyje apibrėžta, kad sąryšio kortežų tvarka negali įtakoti duomenų prasmės. Tai yra, kortežų (eilučių) tvarka gali būti bet kokia.

Normalizuoti lentelę 1 iki pirmos normalinės formos yra labai paprasta, tereikia įrašyti trūkstamą informaciją į neužpildytus atributus, kaip parodyta žemiau pateiktoje lentelėje.

Lentelė 2. Sąryšis, atitinkantis pirmos normalinės formos reikalavimus

KURSAS	K_PAVAD	LYGIS	MODULIS	M_PAVADINIMAS	STATUSAS	KREDITAI
B11	Informatika	Bs	B141	Programavimas 1	bazinis	8
B11	Informatika	Bs	B142	Duomenų bazės	bazinis	8
B11	Informatika	Bs	B143	Matematinė analizė	bazinis	8
B11	Informatika	Bs	B144	C++ pagrindai	specializacijos	11
B11	Informatika	Bs	B145	Tinklų administravimas	specializacijos	11
M15	Statistika	Ms	M215	Statistikos teorija	magistratūros	15
M15	Statistika	Ms	M216	Ekonomika	magistratūros	15
M15	Statistika	Ms	B141	Programavimas 1	bazinis	8

Tokioje lentelėje jau galime keisti eilutes vietomis. Šitos lentelės unikalūs raktas gali būti laukų KURSAS ir MODULIS kombinacija, nes kiekvienam įrašui ši kombinacija yra unikali.

Antra normalinė forma

Reikia užtikrinti, kad kiekvienas neraktinis atributas būtų funkcionaliai priklausomas nuo visų raktų. Jei taip nėra, reikia atskirti į skirtingus sąryšius tuos atributus, kurie priklauso tik nuo rakto dalies.

Kaip matome iš apibrėžimo, lentelė 2 netenkina antros normalinės formos reikalavimų, nes atributai M_PAVADINIMAS, STATUSAS ir KREDITAI priklauso tik nuo atributo MODULIS, o tai yra tik rakto dalis. Todėl turime skaidyti sąryšį į du: KURSU_MODULIAI ir MODULIAI

KURSU_MODULIAI

KURSAS	K_PAVAD	LYGIS	MODULIS
B11	Informatika	Bs	B141
B11	Informatika	Bs	B142
B11	Informatika	Bs	B143
B11	Informatika	Bs	B144
B11	Informatika	Bs	B145
M15	Statistika	Ms	M215
M15	Statistika	Ms	M216
M15	Statistika	Ms	B141

MODULIAI

MODULIS	M_PAVADINIMAS	STATUSAS	KREDITAI
B141	Programavimas 1	bazinis	8
B142	Duomenų bazės	bazinis	8
B143	Matematinė analizė	bazinis	8
B144	C++ pagrindai	specializacijos	11
B145	Tinklų administravimas	specializacijos	11
M215	Statistikos teorija	magistratūros	15
M216	Ekonomika	magistratūros	15

Dabar sąryšis MODULIAI yra antroje normalinėje formoje, bet sąryšis KURSU_MODULIAI – dar ne, nes šio sąryšio laukai K_PAVAD ir LYGIS priklauso tik nuo rakto dalies KURSAS. Todėl sąryšį KURSU_MODULIAI turime išskaidyti į du atskirus sąryšius KURSU_MODULIAI ir KURSAI:

KURSU_MODULIAI

KURSAS	MODULIS
B11	B141
B11	B142
B11	B143
B11	B144
B11	B145
M15	M215
M15	M216
M15	B141

KURSAI

KURSAS	K_PAVAD	LYGIS
B11	Informatika	Bs
M15	Statistika	Ms

Paskutinės aprašytos lentelių MODULIAI, KURSU_MODULIAI ir KURSAI versijos yra antroje normalinėje formoje.

Trečia normalinė forma (TNF)

Reikia užtikrinti, kad visi neraktiniai atributai būtų vienas nuo kito nepriklausomi. Jei taip nėra, reikia sukurti naują sąryšį, kuris neturėtų jokios neraktinių atributų tarpusavio priklausomybės.

Šito apibrėžimo netenkina sąryšis MODULIAI, nes jo atributas KREDITAI priklauso kito atributo STATUSAS. Todėl šį sąryšį turime išskaidyti į du:

MODULIAI

MODULIS	M_PAVADINIMAS	STATUSAS
B141	Programavimas 1	bazinis
B142	Duomenų bazės	bazinis
B143	Matematinė analizė	bazinis
B144	C++ pagrindai	specializacijos
B145	Tinklų administravimas	specializacijos
M215	Statistikos teorija	magistratūros
M216	Ekonomika	magistratūros

STATUSAI

STATUSAS	KREDITAI
bazinis	8
specializacijos	11
magistratūros	15

Taigi, normalizuojant pradinį sąryšį (lentelė 1) iki trečios normalinės formos, gavome 4 sąryšius:

KURSU_MODULIAI

KURSAS	MODULIS
B11	B141
B11	B142
B11	B143
B11	B144
B11	B145

M15	M215
M15	M216
M15	B141

KURSAI

KURSAS	K_PAVAD	LYGIS
B11	Informatika	Bs
M15	Statistika	Ms

MODULIAI

MODULIS	M_PAVADINIMAS	STATUSAS
B141	Programavimas I	bazinis
B142	Duomenų bazės	bazinis
B143	Matematinė analizė	bazinis
B144	C++ pagrindai	specializacijos
B145	Tinklų administravimas	specializacijos
M215	Statistikos teorija	magistratūros
M216	Ekonomika	magistratūros

STATUSAI

STATUSAS	KREDITAI
bazinis	8
specializacijos	11
magistratūros	15

Supaprastintas normalizacijos iki TNF apibrėžimas: sąryšio atributai turi priklausyti nuo rakto, tik nuo viso rakto ir nuo nieko kito.