

0619 ინფორმაციისა და კომუნიკაციის ტექნოლოგიები
INFORMATION AND COMMUNICATION TECHNOLOGIES (ICTS)

რელაციურ მონაცემთა ბაზის ნორმალიზაციის ზოგიერთი საკითხი

ქეთევან ჭელიძე

აკაკი წერეთლის სახელმწიფო უნივერსიტეტი

ქუთაისი, საქართველო

E-mail: ketevan.chelidze@atsu.edu.ge;

E-mail: ketiche@gmail.com

რეზიუმე

რელაციური მონაცემთა ბაზის ლოგიკური მოდელის აგება გულისხმობს: რეალური ან წარმოსახვითი არსებისა და მათი ატრიბუტების გამოყოფას, თანადობების შექმნასა და მათ შორის ფუნქციონალური დამოკიდებულებების განსაზღვრას, თანადობებში მონაცემთა ჩასმის, ცვლილებისა და წაშლის დროს არასასურველი ანომალიების გამოვლენას. ნორმალიზაციის პროცესით ხდება თანადობის გარკვეულ კანონიკურ (ნორმალურ) ფორმად მიყვანა. ყოველი მაღალი რიგის ნორმალური ფორმა აკმაყოფილებს უფრო დაბალი რიგის ნორმალური ფორმის თვისებებსა და ასევე აქვს მასზე უკეთესი თვისებები. ნორმალიზაცია უზრუნველყოფს: მონაცემთა სიჭარბის შემცირებას, არასასურველი დამოკიდებულების გაუქმებასა და მონაცემთა მთლიანობას. განიხილავენ: პირველ, მეორე, მესამე, ბიოს-კოდის, მეოთხე, მეხუთე, მეექვსე ნორმალურ ფორმებს. ნორმალიზაციის პროცესი არასავალდებულოა. მონაცემთა ბაზის სპეციალისტი ინტუიციურად აგებს უკვე ნორმალიზებულ თანადობებს, მაგრამ პროცესის განხილვა საშუალებას გვაძლევს გავიგოთ, თუ რა პრობლემებთან გვექნება საქმე მონაცემთა ბაზასთან მუშაობის დროს.

საკვანძო სიტყვები: რელაციური მონაცემთა ბაზა, ფუნქციონალური დამოკიდებულება, ნორმალური ფორმა, ნორმალიზაცია, 1NF, 2NF, 3NF, BCNF, 4NF, 5NF.

შესავალი

სტატიაში განხილულია თანადობათა ნორმალურ ფორმათა მაგალითები. გამოყოფილია თანადობათა საგასაღებო ატრიბუტები და ატრიბუტებს შორის სხვადასხვა ფუნქციონალური დამოკიდებულებები. განხილულია რელაციური ალგებრის თანადობათა პროექციისა და ბუნებრივი შეერთების ოპერაციები რელაციური მონაცემთა ბაზის ნორმალიზაციისათვის. განხილულ საკითხებს აქვს როგორც თეორიული, ასევე სასწავლო-მეთოდური და პრაქტიკული მნიშვნელობა. გამოყენებულია კვლევის თეორიული, მონაცემთა ანალიზის, შედარებისა და დედუქციის მეთოდები.

ძირითადი ნაწილი

რაიმე საგნობრივი არის მონაცემთა ბაზის რელაციური მოდელის აგება (პროექტირება) გულისხმობს რეალური ან წარმოსახვითი ობიექტების (არსები) გამოყოფასა და მათი ატრიბუტების (მახასიათებლების) განსაზღვრას. ყოველი ატრიბუტის მნიშვნელობები აიღება გარკვეული სიმრავლიდან (დომენი). ატრიბუტს შეიძლება ჰქონდეს ერთი ან რამდენიმე მნიშვნელობა ან არ ჰქონდეს ან არ იყოს ცნობილი (NULL-მნიშვნელობა). დომენების დეკარტული ნამრავლის რაიმე ქვესიმრავლე ქმნის თანადობას (იგივეა, რაც დამოკიდებულება, შესაბამისობა, ინგლისურად კი Relation). რადგან თანადობა სიმრავლეა, მასში არ უნდა იყოს ორი ერთნაირი ელემენტი (კორტეჟი), ამიტომ ატრიბუტების რაიმე ქვესიმრავლის მნიშვნელობები უნდა იყოს უნიკალური. კორტეჟი არის ეგზემპლარია.

მონაცემთა რელაციური მოდელი გულისხმობს სამი ძირითადი: სტრუქტურული, მონაცემებზე მანიპულირებისა და მთლიანობის ნაწილის არსებობას. სტრუქტურული ნაწილი ნიშნავს, რომ მონაცემთა ბაზა შედგება ნორმალიზებული n-არული თანადობებისგან. ნორმალური ფორმა არის თანადობის რაიმე კანონიკური ფორმა, რომელიც უნდა აკმაყოფილებდეს გარკვეულ თვისებებს (მოთხოვნებს). მონაცემებზე მანიპულირების (ამორჩევა/ჩასმა/ცვლილება/წაშლა) ნაწილი ეყრდნობა რელაციურ ალგებრასა და რელაციურ აღრიცხვას. მონაცემთა მთლიანობის უზრუნველყოფა გულისხმობს არსთა და კავშირთა მთლიანობას. მონაცემთა რელაციური მოდელი არის ის ლოგიკური

მოდელი, რომელიც აიგება მონაცემების გარეშე. თანადობის ფიზიკური მოდელი არის ცხრილი, კორტეჟის კი ჩანაწერი. მათ შორის არსებობს მსგავსება და განსხვავება.

თანადობისა და ატრიბუტის სახელი არის არსებითი სახელი მხოლოდით რიცხვში. თანადობის სახელს თან მოსდევს მძიმეებით გამოყოფილი ატრიბუტების სახელები. მაგ., $R(A, B, C, D, \dots)$.

ვთქვათ, A და B არის R თანადობის ატრიბუტები. თუ A ატრიბუტის ერთსა და იგივე მნიშვნელობას შეესაბამება B ატრიბუტის ერთი და იგივე მნიშვნელობა, მაშინ B ატრიბუტი A -ზე ფუნქციონალურად დამოკიდებულია (აღინიშნება $A \rightarrow B$). მასთან, A არის განმსაზღვრელი (დეტერმინანტი) და B კი დამოკიდებული ატრიბუტი. A და B შეიძლება შედგენილი (რამდენიმე ატრიბუტი) იყოს.

სუპერგასაღებია თანადობის ატრიბუტთა ქვესიმრავლე, რომელთა მნიშვნელობა კორტეჟებში უნიკალურია. თანადობას ყოველთვის აქვს თუნდაც ერთი სუპერგასაღები – თანადობის ყველა ატრიბუტი ერთად. სავარაუდო (პოტენციური) გასაღები არის სუპერგასაღების ქვესიმრავლე, რომელიც მინიმალურ რაოდენობა ატრიბუტს შეიცავს. მისგან აღარ უნდა გამოიყოს ატრიბუტთა ქვესიმრავლე, რომელის იქნება სუპერგასაღები. თანადობას შეიძლება ჰქონდეს რამდენიმე სავარაუდო გასაღები. აქედან ერთი შეიძლება ავიღოთ როგორც პირველადი გასაღები, დანარჩენი კი ალტერნატიული. პირველადი გასაღების მნიშვნელობა არ შეიძლება იყოს NULL.

თანადობაში მონაცემთა სიჭარბე (რამდენიმე ერთი და იგივე მონაცემი) იწვევს მათი ცვლილების დროს რამდენიმე განახლების შესრულების აუცილებლობას. თანადობებში მონაცემებზე მანიპულირების დროს ადგილი შეიძლება ექნეს ე. წ. ანომალიებს. რაც ნიშნავს: არასწორი მონაცემების გაჩენას, არსის შესახებ მონაცემთა დაკარგვას, NULL-მნიშვნელობიანი ატრიბუტის გაჩენას, პირველადი გასაღების ნაწილისათვის NULL-მნიშვნელობის მინიჭებას, არსთა და კავშირთა მთლიანობის დარღვევას და სხვა.

ნორმალიზაციის პროცესით ხდება თანადობის ნორმალურ ფორმამდე მიყვანა. განასხვავებენ 1NF, 2NF, 3NF, BCNF, 4NF, 5NF, 6NF (1NF=First Normal Form – პირველი ნორმალური ფორმა, 2NF=Second Normal Form – მეორე ნორმალური ფორმა და ა.შ.) ნორმალურ ფორმებს. ყოველი მაღალი რიგის ფორმა აკმაყოფილებს უფრო დაბალი რიგის ნორმალური ფორმის თვისებებსა (იმყოფება უფრო დაბალი რიგის ნორმალურ ფორმამდე) და აქვს დამატებითი მასზე უკეთესი თვისებებიც.

თუ თანადობის რომელიმე ატრიბუტის მნიშვნელობა თანადობაა, მაშინ შეიძლება ვთქვათ, რომ ის იმყოფება ნულოვან ნორმალურ ფორმაში. მაგ., „თანადობა“ **პროგრამა_სტუდენტი** (პროგ_დას, წელი, ხელმძღვანელი („ხელმძღვანელი“ არის დეპარტამენტის თანამშრომელი (პროფესორი) და ვიგულისხმობთ, რომ ის არ იცვლება პროგრამის სააკრედიტაციო პერიოდში), ხელფასი, დეპ_ტელ, დეპარტამენტი, სტუდენტი (სტუდ_პნ, გვარი_სახელი, მიღების_წელი, ტელეფონი)) შეიცავს თანადობას **სტუდენტი**.

ნორმალიზაციის პროცესი სასარგებლოა, მაგრამ არასაკმარისია მონაცემთა ბაზის აგებისთვის. ადამიანი „ბუნებრივად“ აგებს უკვე ნორმალიზებულ მონაცემთა ბაზასა და ეტაპების გავლა საჭირო არაა. ნორმალიზაციის პროცესი საშუალებას გვაძლევს გამოვავლინოთ ანომალიები და აღმოვფხვრათ ისინი. თანადობის დეკომპოზიცია (დაყოფა) თანადობებად ხდება შემდეგი წესის მიხედვით $A \rightarrow BC \implies (A \rightarrow B) \wedge (A \rightarrow C)$, სადაც A, B, C თანადობის ატრიბუტების ქვესიმრავლეებია.

თანადობა იმყოფება 1NF მაშინ და მხოლოდ მაშინ, თუ თანადობის ატრიბუტების მნიშვნელობების ტიპი ატომარულია (მარტივია). ატრიბუტის ტიპი შეიძლება არ იყოს მარტივი. მაგ., თარიღი, სტრიქონი. ის მაინც განიხილება როგორც ერთი მთლიანი, მარტივი ტიპი. მის კომპონენტებთან წვდომისათვის მოთხოვნათა სტრუქტურულ ენაში (SQL) განსაზღვრულია ფუნქციები.

პროგრამა_სტუდენტი თანადობაში **სტუდენტი** ატრიბუტი ობიექტს (თანადობა) წარმოადგენს და არაატომარულია. შევქმნათ თანადობა **პროგრამა_სტუდენტი** შემდეგნაირად: **პროგრამა_სტუდენტი** (პროგ_დას, წელი, ხელმძღვანელი, ხელფასი, დეპ_ტელ, დეპარტამენტი, სტუდ_პნ, გვარი_სახელი, მიღების_წელი, ტელეფონი (სტუდენტს შეიძლება ჰქონდეს რამდენიმე ტელეფონის ნომერი, ამიტომ ატრიბუტი „ტელეფონი“ მრავალმნიშვნელობიანი ატრიბუტია. აქ განვიხილოთ როგორც ატომარული ტიპის სიდიდე)). მიღებული თანადობის ყველა ატრიბუტის ტიპი არის ან ტექსტური ან რიცხვითი ან თარიღი. ე.ი ატომარული. თანადობა **პროგრამა_სტუდენტი** იმყოფება 1NF-ში.

პროგრამა_სტუდენტი თანადობაში ადგილი აქვს მონაცემთა სიჭარბეს: გამეორებულია პროგრამის დასახელება, ხელმძღვანელი, დეპარტამენტი და სხვა. ეს გამოწვეულია იმით, რომ რამდენიმე არსი: „პროგრამა“, „სტუდენტი“ ერთ თანადობაშია გაერთიანებული.

ვთქვათ, სტუდენტი სწავლობს ერთ ან რამდენიმე პროგრამაზე. **პროგრამა_სტუდენტი** თანადობის სავარაუდო გასაღებია **{პროგ_დას, წელი, სტუდ_პნ}**. არსებობს ფუნქციონალური დამოკიდებულება სავარაუდო გასაღებსა და არაგასაღებ ატრიბუტებს შორის. ამავე დროს აქ არსებობს ფუნქციონალური დამოკიდებულებები: **სტუდ_პნ_გვარი_სახელი, სტუდ_პნ_მიღების_წელი** და სხვა. ამრიგად, არაგასაღები ატრიბუტები დამოკიდებულია შედგენილი გასაღები ატრიბუტის ნაწილზე, ამიტომ ერთი ობიექტის წაშლა სავარაუდო გასაღების ნაწილს გახდის NULL-მნიშვნელობის ტოლად, რაც დაარღვევს მონაცემთა მთლიანობას.

მონაცემთა სიჭარბის აღმოსაფხვრელად რელაციური ალგებრის **პროექციის** ოპერატორით ახდენენ თანადობის დეკომპოზიციას თანადობებად.

თანადობა იმყოფება 2NF-ში მაშინ და მხოლოდ მაშინ, თუ მისი არაგასაღები ატრიბუტი სრულადაა დამოკიდებული სავარაუდო გასაღებზე. თუ თანადობის პოტენციური გასაღები არის მარტივი (ერთი ატრიბუტია), მაშინ ნებისმიერი ფუნქციონალური დამოკიდებულება არის სრული და თანადობა უკვე 2NF-შია.

ჩავატაროთ ორი **პროექციის** მოქმედება **პროგრამა_სტუდენტი** თანადობაზე:

პროგრამა $\leftarrow \pi_{\text{პროგ}_{\text{დას}}, \text{წელი}, \text{ხელმძღვანელი}, \text{ხელფასი}, \text{დეპ}_\text{ტელ}, \text{დეპარტამენტი}}(\text{პროგრამა_სტუდენტი})$

სტუდენტი $\leftarrow \pi_{\text{პროგ}_{\text{დას}}, \text{წელი}, \text{სტუდ}_\text{პნ}, \text{გვარი}_\text{სახელი}, \text{მიღების}_\text{წელი}, \text{ტელეფონი}}(\text{პროგრამა_სტუდენტი})$

და მივიღოთ ორი თანადობა: **პროგრამა**(**პროგ_დას, წელი, ხელმძღვანელი, ხელფასი, დეპ_ტელ, დეპარტამენტი**) და **სტუდენტი**(**პროგ_დას, წელი, სტუდ_პნ, გვარი_სახელი, მიღების_წელი, ტელეფონი**). თავდაპირველი **პროგრამა_სტუდენტი** თანადობის მიღება შეიძლება რელაციური ალგებრის **ბუნებრივი შეერთების** ოპერაციით საერთო **{პროგ_დას, წელი}** ატრიბუტების მიხედვით:

პროგრამა_სტუდენტი $\leftarrow \text{პროგრამა} \bowtie_{\text{პროგ}_{\text{დას}}, \text{წელი}} \text{სტუდენტი}$

პროგრამა. **{პროგ_დას, წელი}** არის მშობელი **პროგრამა** თანადობის პირველადი გასაღები და მასთან კავშირისთვის შეიღობილ **სტუდენტ** თანადობაში არის გარე გასაღები **სტუდენტი**.**{პროგ_დას, წელი}**. მონაცემთა ბაზაში კავშირის მთლიანობის დაცვა ნიშნავს, რომ გარე გასაღების მნიშვნელობა არ უნდა იყოს პირველადი გასაღების არარსებული მნიშვნელობა. პირველადი და გარე გასაღებებისთვის პრაქტიკაში იყენებენ სურაგატ (შემცვლელ) ატრიბუტებს, რომელთა ტიპი უმეტესად მთელია ან/და ავტომატურად მთვლელი.

თუ განვიხილავთ **პროგრამა** თანადობას, აქ არსებობს ტრანზიტული ფუნქციონალური დამოკიდებულება: **პროგ_დას_დეპარტამენტი, დეპარტამენტი_დეპ_ტელ** და აქედან კი **პროგ_დას_დეპ_ტელ**.

თანადობა იმყოფება 3NF-ში მაშინ და მხოლოდ მაშინ, თუ მასში არაა ტრანზიტული ფუნქციონალური დამოკიდებულება.

პროგრამა თანადობის დეკომპოზიციისთვის გამოვიყენოთ **პროექციის** ოპერატორი:

პროგრამა $\leftarrow \pi_{\text{პროგ}_{\text{დას}}, \text{წელი}, \text{ხელმძღვანელი}, \text{ხელფასი}, \text{დეპარტამენტი}}(\text{პროგრამა})$

თანამშრომელი_დეპარტამენტი $\leftarrow \pi_{\text{გვარი}_\text{სახელი}, \text{დეპ}_\text{ტელ}, \text{დეპარტამენტი}}(\text{პროგრამა})$

მიიღება ორი თანადობა: **პროგრამა** (**პროგ_დას, წელი, ხელმძღვანელი**) და **თანამშრომელი_დეპარტამენტი**(**გვარი_სახელი, დეპ_ტელ, დეპარტამენტი**). თავდაპირველი **პროგრამა** თანადობა მიიღება **ბუნებრივი შეერთების** ოპერატორის გამოყენებით:

$$\text{პროგრამა} \leftarrow \text{პროგრამა} \times_{\text{ხელმძღვანელი=გვარი}_{\text{სახელი}}} (\text{თანამშრომელი}_{\text{დეპარტამენტი}})$$

ვთქვათ, თანამშრომელი მუშაობს მხოლოდ ერთ დეპარტამენტში, მაშინ **თანამშრომელი_დეპარტამენტი** თანადობის პოტენციური გასაღებია **გვარი_სახელი**¹. ამავე დროს აქ არის ტრანზიტული ფუნქციონალური დამოკიდებულება:

გვარი_სახელი **დეპარტამენტი**, **დეპარტამენტი** **დეპ_ტელ**, **გვარი_სახელი** **დეპ_ტელ**.
თანამშრომელი_დეპარტამენტი თანადობისგან დეკომპოზიციით მივიღოთ **თანამშრომელი** და **დეპარტამენტი** 3NF თანადობები:

$$\text{თანამშრომელი} \leftarrow \pi_{\text{გვარი}_{\text{სახელი}}, \text{დეპარტამენტი}} (\text{თანამშრომელი}_{\text{დეპარტამენტი}})$$

$$\text{დეპარტამენტი} \leftarrow \pi_{\text{დეპ}_\text{ტელ}, \text{დეპარტამენტი}} (\text{თანამშრომელი}_{\text{დეპარტამენტი}})$$

თანამშრომელი_დეპარტამენტი თანადობა მიიღება **ბუნებრივი შეერთების** ოპერაციით საერთო „დეპარტამენტი“ ატრიბუტით:

$$\text{თანამშრომელი}_{\text{დეპარტამენტი}} \leftarrow \text{თანამშრომელი} \bowtie_{\text{დეპარტამენტი=დეპარტამენტი}} (\text{დეპარტამენტი})$$

თანადობა იმყოფება ბიოს-კოდის² ნორმალურ ფორმაში (BCNF) მაშინ და მხოლოდ მაშინ, როცა მასში ყოველი არატრივიალური ფუნქციონალური დამოკიდებულება განისაზღვრება პოტენციური გასაღებით. ფუნქციონალური დამოკიდებულება არის ტრივიალური მაშინ და მხოლოდ მაშინ, თუ დამოკიდებული ნაწილი დეტერმინანტის ქვესიმრავლეა: $(B \subseteq A) \implies (A \rightarrow B)$. BCNF არის 3NF-ის დაზუსტება.

ვთქვათ, პროფესორი ასწავლის საგნებს (სასწავლო კურსებს) და ყოველ საგანს შეიძლება ასწავლიდეს რამდენიმე პროფესორი. საგნისთვის შეიძლება არსებობდეს რამდენიმე კონსპექტი, ხოლო ყოველი კონსპექტი ეკუთვნის მხოლოდ ერთ საგანს. განვიხილოთ თანადობა **პროფესორი_საგანი_კონსპექტი** (**პროფესორი**, **საგანი**, **კონსპექტი**). ამ თანადობაში **საგნებისა** და **კონსპექტების** ჩანაწერები დუბლირებულია, რაც იწვევს ცვლილების ანომალიებს. მასში ვერ ჩავწერთ **პროფესორს**, თუ არ გვეცოდინება **საგანი** და **კონსპექტი**. თუ წავშლით **საგანს**, წაიშლება **პროფესორიც** და **კონსპექტიც**. თანადობაში არის ერთმანეთთან დაუკავშირებელი არსები: **პროფესორი** და **კონსპექტი**. თანადობის პოტენციური გასაღებია **[საგანი, კონსპექტი]**. ასევე არსებობს დამოკიდებულება **პროფესორი/საგანი, საგანი/კონსპექტი**. რადგან ყველა არაგასაღები ატრიბუტი დამოკიდებული უნდა იყოს მხოლოდ სავარაუდო გასაღებზე, ამიტომ თანადობა არაა BCNF-ში.

განვიხილოთ რაიმე $R(A, B, C)$ თანადობა. მრვალმნიშვნელობიანი დამოკიდებულება (აღინიშნება $A \twoheadrightarrow B$) განისაზღვრება შემდეგნაირად – თუ C-გან დამოუკიდებლად A-ის ყოველი მნიშვნელობისთვის არსებობს B-ის მრავალი მნიშვნელობა, და პირიქით.

პროფესორი_საგანი_კონსპექტი თანადობაში მრვალმნიშვნელობიანი დამოკიდებულების დროს ერთ ატრიბუტთან დაკავშირებულია ორი ატრიბუტი, რომლებიც ერთმანეთისგან დამოუკიდებელია. დავყოთ თანადობა **პროფესორი_საგანი_კონსპექტი** ორ თანადობად: **პროფესორი_კონსპექტი** (**პროფესორი**, **კონსპექტი**) და **კონსპექტი_საგანი** (**კონსპექტი**, **საგანი**):

$$\text{პროფესორი}_{\text{კონსპექტი}} \leftarrow \pi_{\text{პროფესორი}, \text{კონსპექტი}} (\text{პროფესორი})$$

$$\text{კონსპექტი}_{\text{საგანი}} \leftarrow \pi_{\text{კონსპექტი}, \text{საგანი}} (\text{პროფესორი})$$

1 ვიგულისხმობთ, რომ არაა ორი ერთნაირი სახელისა და გვარის თანამშრომელი ან გამოვიყენოთ პირადი ნომერი ან სუროგატი ატრიბუტი.

2 რაიმონდ ბოისი (1947-1974 წწ.) და ედგარ ფრენკ კოდი (1923-2003 წწ.) - ამერიკელი მეცნიერები.

თავდაპირველ პროფესორი_საგანი_კონსპექტი თანადობას მივიღებთ მათი ბუნებრივი შეერთებით საერთო „კონსპექტი“ ატრიბუტით:

$$\text{პროფესორი} \leftarrow \text{პროფესორი}_{\text{კონსპექტი}} \bowtie_{\text{კონსპექტი}=\text{კონსპექტი}} (\text{კონსპექტი}_{\text{საგანი}})$$

სტუდენტმა შეიძლება იცოდეს რამდენიმე უცხო ენა და ჰქონდეს სხვადასხვა ჰობი (სპორტული სახეობა, ხელსაქმე, ხატვა, მუსიკა და სხვა). განვიხილოთ თანადობა **სტუდენტი_უცხოენა_ჰობი** (**სტუდენტი**, **უცხო ენა**, **ჰობი**). მრავალმნიშვნელობიანი ატრიბუტები იწვევს მონაცემთა სიჭარბეს: სტუდენტის ყოველი „**უცხო ენისთვის**“ არსებობს მისი „**ჰობის**“ მრავალი მნიშვნელობა და პირიქით.

თანადობა იმყოფება 4NF-ში მაშინ და მხოლოდ მაშინ, თუ მასში არაა მრავალმნიშვნელობიანი ფუნქციონალური დამოკიდებულებები.

იმისათვის, რომ **სტუდენტი_უცხოენა_ჰობი** თანადობისგან მივიღოთ 4NF-თანადობები, შევასრულოთ პროექციის ორი ოპერაცია მისი დეკომპოზიციისთვის:

$$\text{სტუდენტი}_{\text{უცხოენა}} \leftarrow \pi_{\text{სტუდენტი, უცხოენა}} (\text{სტუდენტი}_{\text{უცხოენა } \text{ჰობი}})$$

$$\text{სტუდენტი}_{\text{ჰობი}} \leftarrow \pi_{\text{სტუდენტი, ჰობი}} (\text{სტუდენტი}_{\text{უცხოენა } \text{ჰობი}})$$

მიღებულ თანადობებში აღარაა მრავალმნიშვნელობიანი ატრიბუტები. თავდაპირველ **სტუდენტი_უცხოენა_ჰობი** თანადობას მივიღებთ მათი ბუნებრივი შეერთებით საერთო „სტუდენტი“ ატრიბუტით:

$$\text{სტუდენტი}_{\text{უცხოენა } \text{ჰობი}} \leftarrow \text{სტუდენტი}_{\text{უცხოენა}} \bowtie_{\text{სტუდენტი}=\text{სტუდენტი}} (\text{სტუდენტი}_{\text{ჰობი}})$$

R თანადობაში არსებობს შეერთების დამოკიდებულება მაშინ, როცა ის შეიძლება აიგოს მისგან პროექციით მიღებული თანადობების ბუნებრივი შეერთებით: $[R_1, R_2, \dots, R_n] \longrightarrow R$, სადაც, $R_1, R_2, \dots, R_n$ თანადობები მიღებულია R -ის დეკომპოზიციით და $R \leftarrow R_1 \bowtie R_2 \bowtie \dots \bowtie R_n$. შეერთების დამოკიდებულება არატრივიალურია, თუ $\forall R_i \neq R$.

თანადობა იმყოფება 5NF-ში მაშინ და მხოლოდ მაშინ, თუ მასში არაა არატრივიალური შეერთების დამოკიდებულება.

მაგ., განვიხილოთ **სტუდენტი_გეგმა_საგანი** (**სტუდენტი**, **გეგმა**, **საგანი**) თანადობა. ყოველი სტუდენტი შეიძლება სწავლობდეს რამდენიმე სასწავლო გეგმაზე (ძირითადი პროგრამა, მაინორ პროგრამა), ყოველ გეგმაში ბევრი საგანია, ყოველი საგანი ბევრ გეგმაშია. თანადობაში მონაცემთა სიჭარბეა. პროექციით დავყოთ სამ თანადობად: **სტუდენტი_გეგმა** (**სტუდენტი**, **გეგმა**), **სტუდენტი_საგანი** (**სტუდენტი**, **საგანი**) და **გეგმა_საგანი** (**გეგმა**, **საგანი**):

$$\text{სტუდენტი}_{\text{გეგმა}} \leftarrow \pi_{\text{სტუდენტი, გეგმა}} (\text{სტუდენტი})$$

$$\text{სტუდენტი}_{\text{საგანი}} \leftarrow \pi_{\text{სტუდენტი, საგანი}} (\text{სტუდენტი})$$

$$\text{გეგმა}_{\text{საგანი}} \leftarrow \pi_{\text{გეგმა, საგანი}} (\text{სტუდენტი})$$

მიღებული თანადობები იმყოფება 5NF-ში და მათი ბუნებრივი შეერთებით მიიღება თავდაპირველი **სტუდენტი_გეგმა_საგანი** თანადობა:

$$\text{სტუდენტი} \leftarrow \text{სტუდ.} \bowtie_{\text{სტუდ.}=\text{სტუდ.}} \text{სტუდ.}_{\text{საგანი}} \bowtie_{\text{საგანი, სტუდ.}=\text{საგანი, სტუდ.}} (\text{გეგმა}_{\text{საგანი}})$$

მეექვსე ნორმალური ფორმა გამოიყენება ქრონოლოგიური რელაციურ მონაცემთა ბაზების ნორმალიზაციითვის, სადაც ატრიბუტების მნიშვნელობები შეზღუდულია დროითი ინტერვალის მონაცემებით.

ნორმალიზაციის პროცესი არასავალდებულოა. მონაცემთა ბაზის სპეციალისტი ინტუიციურად აგებს უკვე ნორმალიზებულ თანადობებს, მაგრამ პროცესის განხილვა საშუალებას გვაძლევს გავიგოთ, თუ რა პრობლემებთან გვექნება საქმე მონაცემთა ბაზასთან მუშაობის დროს. პრაქტიკაში რელაციური მონაცემთა ბაზის პროექტირებისთვის მისი თანადობების მესამე ან/და ბოის-კოდის ნორმალურ ფორმამდე მიყვანა სავსებით საკმარისია.

ლიტერატურა

1. Ramez Elmasri, Shamkant B. Navathe (2016). Fundamentals of Database systems, 7th edition.
2. Abraham Silberschatz, Henry F. Korth, S. Sudarshan. (2019). Database System Concepts, 6th edition

Some Issues of Normalization of Relational Database

Ketevan Tchelidze

Abstract

Building a logical model of a relational database involves: defining real or imaginary entities and their attributes, revealing functional dependencies between attributes, creating relationships among the relations, detecting undesirable anomalies during data insertion, modification and deletion. The process of normalization is used to bring relations to certain canonical (normal) forms. Each higher normal form satisfies the properties of the lower normal form and also has better properties than it. Normalization ensures the reduction of data redundancy, the removing of undesirable functional dependencies and the integrity of data. There are considered: First, Second, Third, Boyce–Codd, Fourth, Fifth, Sixth normal forms. The normalization process is optional; a database specialist intuitively constructs already normalized relations, but discussing the process allows us to understand what problems we will encounter when working with the database. The article discusses examples of normal form of relations. Key attributes of relations and various functional dependencies between attributes are defined. The projection and natural join operations of relational algebra are discussed for relational database normalization. The issues discussed in the article have theoretical, educational-methodological and practical significance. The theoretical research, data comparison and analysis, induction, deduction methods are used.

Key words: Relational database, functional dependency, Normal Form, Normalization, 1NF, 2NF, 3NF, BCNF, 4NF, 5NF.