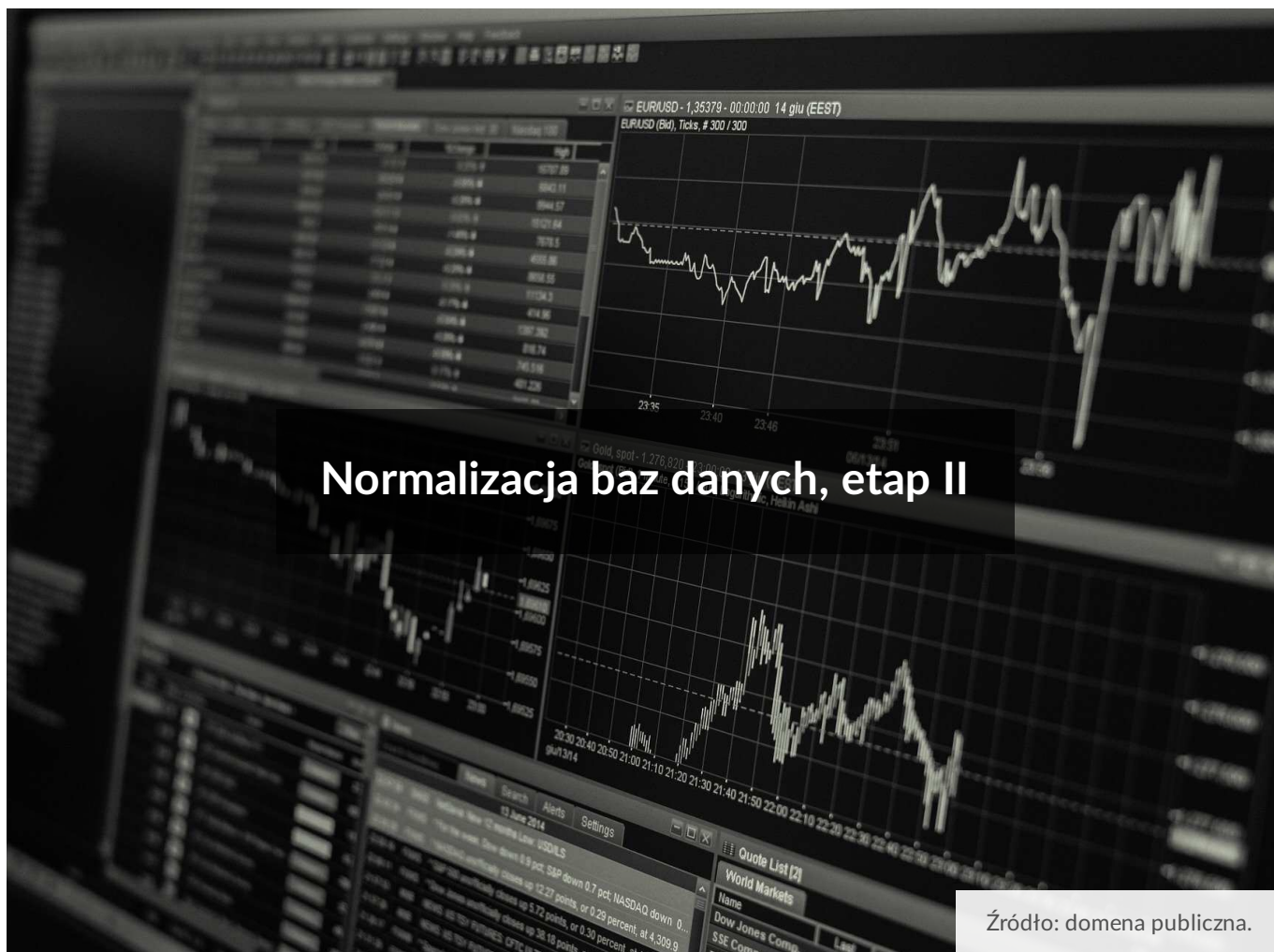




Normalizacja baz danych, etap II

- Wprowadzenie
- Przeczytaj
- Infografika
- Sprawdź się
- Dla nauczyciela



Druga postać normalna bezpośrednio opiera się na pierwszej, ale zwiększa jej zasięg o kolejną zasadę. Pozwala ona – w jeszcze większym stopniu – zapewnić bezpieczeństwo spójności bazy oraz stanowi drogę do kolejnych postaci normalnych.

W tym celu wykorzystuje ona klucze złożone oraz zależności funkcyjne – pojęcia przydające się w wielu innych działach baz danych. Dzięki nim można jeszcze bardziej zmniejszyć redundancję danych.

Więcej informacji o normalizacji baz danych znajdziesz w e-materiałach:

- [Normalizacja baz danych, etap I,](#)
- [Normalizacja baz danych, etap III.](#)

Twoje cele

- Wyjaśnisz znaczenie klucza potencjalnego.
- Zidentyfikujesz zależności funkcyjne w relacji.
- Zapoznasz się z drugą postacią normalną oraz sposobami sprowadzania do niej przykładowych tabel.

Przeczytaj

Już wiesz

Wiedza na temat pierwszej postaci normalnej (i jej implementacji) jest konieczna, aby zrozumieć drugą postać normalną. Jest ona szczególnym przypadkiem pierwszej postaci.

Dla przypomnienia – reguły **pierwszej** postaci normalnej:

- jedna tabela opisuje jeden obiekt;
- wartości atrybutów są atomowe;
- tabela nie zawiera powtarzających się informacji;
- znaczenie kolumn nie zależy od ich kolejności.

Druga postać normalna opiera się na pierwszej postaci normalnej, dodając kilka dodatkowych zasad. Jednak aby je zrozumieć, konieczne jest wprowadzenie paru nowych pojęć.

Klucz potencjalny

Czasem nazywa się go kluczem kandydującym. To minimalny zbiór atrybutów, których zestaw wartości jest unikalny dla każdego rekordu.

Prościej: jedna kolumna lub więcej kolumn, w których nie występują dwa rekordy, mające wszystkie wartości identyczne. Dodatkowo jest to minimalna konieczna liczba kolumn o tej własności.

Klucz potencjalny służy jako wielokolumnowy identyfikator dla rekordów.

Spróbujmy znaleźć klucz potencjalny dla tabeli:

Autor	Tytuł	Rodzaj okładki	Cena
Mark Twain	Przygody Tomka Sawyera	Miękka	19,00 zł
Mark Twain	Przygody Tomka Sawyera	Twarda	39,00 zł
Henryk Sienkiewicz	Krzyżacy Część I	Twarda	39,00 zł
Mark Twain	Przygody Hucka	Miękka	24,00 zł

W każdej kolumnie znajduje się jakieś powtórzenie, a więc pojedyncza kolumna nie może być tu kluczem kandydującym. Wybranie pojedynczej kolumny nie spełniałoby w tym przypadku warunku unikalności.

Teraz należy się zastanowić, jaka para kolumn nie posiada powtórzeń.

Nie są nimi *Autor* i *Tytuł*, bo dwa wiersze mogą mieć wówczas identyczne wartości, a różniłyby się rodzajem okładki. Czyli nie zostałaby spełniona reguła unikalności.

Kluczem kandydującym jest zatem *Tytuł* i *Rodzaj okładki*, ponieważ dwa wiersze nie mogą mieć tych dwóch wartości identycznych.

Dla zainteresowanych

Skąd pewność, że nie mogą?

Założmy, że **istnieją** dwa wiersze o tych samych polach *Tytuł* i *Rodzaj okładki*. To, że mają ten sam *Tytuł*, oznacza, że mają tego samego *Autora*. Podczas gdy ten sam *Tytuł* i *Rodzaj okładki* oznaczałby tę samą cenę.

Po wzięciu pod uwagę wszystkich parametrów, okazuje się, że te wiersze musiałyby być identyczne, na co większość baz danych nie pozwala. Można więc bezpiecznie założyć, że żadne dwa wiersze nie mogą mieć tych samych *Tytułów* i *Rodzajów okładki*.

Zależność funkcyjna

Szeroko rozumiana zależność pomiędzy atrybutami. Zachodzi, gdy wartość jednego atrybutu jest jednoznacznie zdefiniowana przez jeden lub więcej innych atrybutów.

Przykładowa zależność funkcyjna w powyższej tabeli to zależność *Tytuł* → *Autor*. Oznacza ona, że jeśli znamy wartość *Tytuł* w danym wierszu, to z góry wiadomo, jaki będzie *Autor*.

Jeśli np. wiemy, że *Tytuł* to „Krzyżacy”, to z góry wiadomo, że *Autorem* będzie Henryk Sienkiewicz.

Innym przykładem jest zależność *Autor, Tytuł, Rodzaj okładki* → *Cena*. Ponieważ te pola jednoznacznie definiują cenę. Jest to także przykład zależności jednego atrybutu od wielu innych atrybutów.

Ważne!

Zależność funkcyjna nie wynika z danych w tabelach, ale z ich znaczenia w rzeczywistości. Założmy, że w tabeli cena „Przygód Hucka” wynosi 19,00 zł. Można się wtedy pokusić o stwierdzenie, że istnieje zależność *Rodzaj okładki* → *Cena*, jednak byłby to błąd, ponieważ taka zasada nie obowiązuje w tej bazie danych. Jakiś nowy wiersz może w przyszłości nie spełniać tej reguły.

Zależności funkcyjne trzeba przeanalizować logicznie i dojść do wniosku, czy zawsze muszą być spełnione. Zależność funkcyjna jest prawdziwa tylko wtedy, kiedy nie istnieją od niej wyjątki.

Druga postać normalna

Teraz, gdy znane są nam pojęcia klucza potencjalnego oraz zależności funkcyjnej, możemy przejść do właściwej definicji drugiej postaci normalnej.

Definicja: Druga postać normalna

Relacja (tabela) jest w drugiej postaci normalnej wyłącznie wtedy, gdy spełnia warunki pierwszej postaci normalnej oraz gdy żadna kolumna nienależąca do klucza potencjalnego nie zależy funkcyjnie od jego fragmentu.

Prościej – aby być w drugiej postaci normalnej, tabela jednocześnie musi:

- być w pierwszej postaci normalnej;
- nie posiadać kolumny, której wartości można wywnioskować z niektórych wartości klucza kandydującego (mogą być wnioskowane z całego klucza).

Można zauważyć, że druga reguła ma na celu zwiększenie bezpieczeństwa **spójności** – tak samo jak w przypadku pierwszej postaci normalnej.

Przeanalizujmy wcześniejszy przykład tabeli zawierającej informacje o książkach w pewnym wydawnictwie oraz **niespełniającej** wymagań drugiej postaci normalnej.

Autor	Tytuł	Rodzaj okładki	Cena
Mark Twain	Przygody Tomka Sawyera	Miękka	19,00 zł
Mark Twain	Przygody Tomka Sawyera	Twarda	39,00 zł
Henryk Sienkiewicz	Krzyżacy Część I	Twarda	39,00 zł
Mark Twain	Przygody Hucka	Miękka	24,00 zł

Wiemy, że jedyny klucz potencjalny to *Tytuł*, *Rodzaj okładki*. Wiemy także, że istnieje zależność *Tytuł* → *Autor*. Wynika z tego, że pole *Autor* zależy funkcyjnie tylko od fragmentu klucza, a nie od całości. Z tego powodu ta relacja nie spełnia wymagań drugiej postaci normalnej.

Sprowadzenie do drugiej postaci normalnej

Polega ono na stworzeniu dodatkowej relacji, w której zostanie przechowana część danych z modyfikowanej relacji.

W tym przypadku chcemy, aby znalazły się w niej kolumny *Autor* oraz *Tytuł*. Dzięki temu będzie można usunąć *Autora* z tabeli i doprowadzić ją do drugiej postaci normalnej.

Tabele po modyfikacji:

Tytuł	Rodzaj okładki	Cena
Przygody Tomka Sawyera	Miękka	19,00 zł
Przygody Tomka Sawyera	Twarda	39,00 zł
Krzyżacy Część I	Twarda	39,00 zł
Przygody Hucka	Miękka	24,00 zł

Autor	Tytuł
Mark Twain	Przygody Tomka Sawyera
Mark Twain	Przygody Hucka
Henryk Sienkiewicz	Krzyżacy Część I

Teraz zastanówmy się, czy nowa tabela także spełnia wymagania drugiej postaci normalnej. Jedynym kluczem potencjalnym jest *Tytuł*, ponieważ *Autorowi* może odpowiadać wiele *Tytułów*, ale jednemu *Tytułowi* tylko jeden *Autor*. Oznacza to, że pola *Tytuł* nie mogą się powtarzać.

Dodatkowo istnieje zależność funkcyjna $Tytuł \rightarrow Autor$. Taka zależność spełnia wymagania trzeciej postaci normalnej, bo pole *Autor* zależy funkcyjnie od całego klucza potencjalnego w tej relacji.

Ważne!

Tego typu rozwiązanie można uogólnić. Aby doprowadzić tabelę z pierwszej do drugiej postaci normalnej należy:

1. Zidentyfikować zależności funkcyjne, których lewa strona to tylko fragment klucza potencjalnego.
2. Utworzyć nową relację o atrybutach będących fragmentem klucza z poprzedniego punktu.
3. Dodać kolejnych atrybutów, będących prawą stroną zależności funkcyjnej (atrybuty za strzałką w zapisie zależności).
4. Usunąć atrybuty z prawej strony zależności z oryginalnej relacji.

W tym przypadku realizowaliśmy te kroki poprzez:

1. Zidentyfikowanie zależności $Tytuł \rightarrow Autor$, podczas gdy klucz potencjalny to *Tytuł*, *Rodzaj okładki*.
2. Utworzenie relacji o atrybucie *Tytuł*.
3. Dodanie do niej atrybutu *Autor*.
4. Usunięcie *Autora* z oryginalnej tabeli.

Słownik

spójność

brak sprzecznych informacji w bazie danych; może zostać naruszona np., gdy w bazie znajduje się wiele kopii danej informacji, a w wyniku jakiejś kwerendy lub procedury zmieniona zostanie tylko jedna instancja tej informacji

Infografika

Polecenie 1

Przeanalizuj infografikę, a następnie wykonaj ćwiczenie.

Źródło: Contentplus.pl Sp. z o.o., licencja: CC BY-SA 3.0.

Ćwiczenie 1

Doprowadź relację do drugiej postaci normalnej, w razie potrzeby rozłóż ją na pomniejsze relacje. Pamiętaj, że dwa zwierzęta mogą mieć te same imiona, o ile są różnych gatunków (na przykład zebra i pingwin współdzielą imię). Wewnątrz gatunku imiona zawsze są unikalne.

Gatunek	Rodzaj	Wiek zwierząt	Imiona zwierząt	Ilość zwierząt
Lew afrykański	Lądowe	3,6	Nala, Kion	2
Zebra sawannowa	Lądowe	2,5	Tama, Sarabi	2
Foka plamista	Wodne/Lądowe	1	Fuli	1
Pingwin Cesarski	Wodne/Lądowe	2,3	Tama Zuri	2
Orka oceaniczna	Wodne	9	Chumvi	1

Sprawdź się

Pokaż ćwiczenia:   

Ćwiczenie 1



Ćwiczenie 2



Ćwiczenie 3



Ćwiczenie 4



Ćwiczenie 5



Ćwiczenie 6



Ćwiczenie 7



Ćwiczenie 8



Dla nauczyciela

Autor: Maurycy Gast

Przedmiot: Informatyka

Temat: Normalizacja baz danych, etap II

Grupa docelowa:

Szkoła ponadpodstawowa, liceum ogólnokształcące, technikum, zakres rozszerzony

Podstawa programowa:

Cele kształcenia – wymagania ogólne

II. Programowanie i rozwiązywanie problemów z wykorzystaniem komputera oraz innych urządzeń cyfrowych: układanie i programowanie algorytmów, organizowanie, wyszukiwanie i udostępnianie informacji, posługiwanie się aplikacjami komputerowymi.

Treści nauczania – wymagania szczegółowe

II. Programowanie i rozwiązywanie problemów z wykorzystaniem komputera i innych urządzeń cyfrowych.

Zakres rozszerzony. Uczeń spełnia wymagania określone dla zakresu podstawowego, a ponadto:

4) przygotowując opracowania rozwiązań złożonych problemów, posługuje się wybranymi aplikacjami w stopniu zaawansowanym:

d) projektuje i tworzy relacyjną bazę złożoną z wielu tabel oraz sieciową aplikację bazodanową dla danych związanych z rozwiązywanym problemem, formułuje kwerendy, tworzy i modyfikuje formularze oraz raporty, stosuje język SQL do wyszukiwania informacji w bazie i do jej modyfikacji, uwzględnia kwestie integralności danych, bezpieczeństwa i ochrony danych w bazie,

Kształtowane kompetencje kluczowe:

- kompetencje cyfrowe;
- kompetencje osobiste, społeczne i w zakresie umiejętności uczenia się;
- kompetencje matematyczne oraz kompetencje w zakresie nauk przyrodniczych, technologii i inżynierii.

Cele operacyjne (językiem ucznia):

- Wyjaśnisz znaczenie klucza potencjalnego.
- Zidentyfikujesz zależności funkcyjne w relacji.
- Zapoznasz się z drugą postacią normalną oraz sposobami sprowadzania do niej przykładowych tabel.

Strategie nauczania:

- konstruktywizm;
- konektywizm.

Metody i techniki nauczania:

- dyskusja;
- rozmowa nauczająca z wykorzystaniem multimediu i ćwiczeń interaktywnych.

Formy pracy:

- praca indywidualna;
- praca w parach;
- praca w grupach;
- praca całego zespołu klasowego.

Środki dydaktyczne:

- komputery z głośnikami, słuchawkami i dostępem do internetu;
- zasoby multimedialne zawarte w e-materiale;
- tablica interaktywna/tablica, pisak/kreda.

Przebieg lekcji

Przed lekcją:

1. **Przygotowanie do zajęć.** Nauczyciel loguje się na platformie i udostępnia e-materiał: „Normalizacja baz danych, etap II”. Uczniowie zapoznają się z treściami w sekcji „Przeczytaj”.

Faza wstępna:

1. Nauczyciel wyświetla temat i cele zajęć. Prosi uczniów, by na podstawie wiadomości zdobytych przed lekcją zaproponowali kryteria sukcesu.
2. **Rozpoznanie wiedzy uczniów.** Uczniowie tworzą pytania dotyczące tematu zajęć, na które odpowiedzą w trakcie lekcji.

Faza realizacyjna:

1. **Praca z tekstem.** Jeżeli przygotowanie uczniów do lekcji jest niewystarczające, nauczyciel prosi o indywidualne zapoznanie się z treścią zawartą w sekcji „Przeczytaj”.

Każdy uczestnik zajęć podczas cichego czytania wynotowuje najważniejsze kwestie poruszane w tekście.

2. **Praca z multimedium.** Nauczyciel wyświetla zawartość sekcji „Infografika”. Uczniowie wspólnie zapoznają się z treścią zawartego w niej multimedium.
3. **Ćwiczenie umiejętności.** Uczniowie wykonują indywidualnie ćwiczenie nr 1-8, a następnie porównują swoje odpowiedzi z kolegą lub koleżanką.

Faza podsumowująca:

1. Nauczyciel ponownie wyświetla na tablicy temat i cele lekcji zawarte w sekcji „Wprowadzenie”. W kontekście ich realizacji następuje omówienie ewentualnych problemów z rozwiązaniem ćwiczeń z sekcji „Sprawdź się”.

Praca domowa:

1. Uczniowie wykonują Polecenie nr 1 z sekcji „Infografika”: „Doprowadź poniższą relację do drugiej postaci normalnej, w razie potrzeby rozłóż ją na pomniejsze relacje. Pamiętaj, że dwa zwierzęta mogą mieć te same imiona, o ile są różnych gatunków (na przykład zebra i pingwin współdzielą imię). Wewnątrz gatunku imiona zawsze są unikalne.”

Wskazówki metodyczne:

- Treści w sekcji „Infografika” można wykorzystać jako materiał służący powtórzeniu materiału.