

# **Normalizacja schematów baz danych**

Dr hab. inż. Krzysztof Pieczarka

Email: [krzysztof.pieczarka@gmail.com](mailto:krzysztof.pieczarka@gmail.com)

Normalizacja relacji ma na celu takie jej przekształcenie, by nie posiadała cech niepożądanych. Normalizacja polega na dekompozycji/podziale relacji na mniejsze schematy relacji.

Proces normalizacji musi posiadać następujące własności:

- żaden atrybut (składowa definicyjna encji) nie może ulec zagubieniu w trakcie tego procesu,
- dekompozycja relacji nie może prowadzić do utraty informacji – wszystkie (pożądane) zależności funkcyjne (głównie semantyczne) są reprezentowane w otrzymanych mniejszych schematach relacji.

Weźmy dla przykładu relację opisującą lekcje odbywające się w szkole w danym półroczu. Relacja ta może zawierać następujące atrybuty:

- nazwa przedmiotu,
- imię,
- nazwisko,
- adres.

Przykładowa krotka takiej relacji mogłaby mieć postać:

(język polski, Janina, Nowak, ul. Krótka 30/3 52-129 Wrocław)

Jednak z tak zaprojektowaną relacją związane są następujące anomalie:

- adres składający się z kilku części nie został podzielony w związku z tym nie jest możliwe sprawdzenie ile osób mieszka we Wrocławiu i nie potrzebuje hotelu;
- jeden nauczyciel może mieć zajęcia z kilku przedmiotów, w związku z tym występuje redundancja danych;
- zmiana jednej z informacji o nauczycielu (np. adresu) powoduje konieczność zmiany wszystkich krotek zawierających te dane w celu zachowania integralności;
- nie jest możliwe wprowadzenie informacji o nauczycielu, który w aktualnym półroczu nie ma żadnych zajęć;
- usunięcie przedmiotu może spowodować również usunięcie wszelkich informacji o osobie, która go prowadziła.

Utrzymanie integralności takiej bazy nie jest więc proste.

Jednak opisaną relację można zamienić na dwie inne, które nie będą posiadały tych wad:

- relacja Nauczyciel:

(id\_nauczyciel (PK), imie, nazwisko, ulica, nr\_domu, nr\_mieszkania, kod\_pocztowy, poczta)

- relacja Przedmiot

(id\_przedmiot (PK), nazwa, id\_nauczyciel (FK))

Te dwie relacje nie posiadają opisanych wcześniej cech niepożądanych ponieważ:

- adres jest zdekomponowany na części składowe, w związku z czym możliwe jest wyszukiwanie danych np. według miejscowości zamieszkania prowadzącego, ulicy, numeru domu;
- zmiana informacji o nauczycielu (np. adresu) nie powoduje konieczności zmian danych w relacji Zajęcia. Zmiana ta odbywa się tylko w jednym miejscu;
- możliwe jest wprowadzenie informacji o osobie, która nie ma zajęć w aktualnym półroczu, ale być może będzie je miała w półroczu następnym;
- usunięcie przedmiotu nie powoduje usunięcia informacji o osobie, która go prowadziła.

Jednak taka reprezentacja danych posiada wady podobne do opisanych wcześniej, ale dotyczące przedmiotów.

Dlatego w dobrze zaprojektowanej bazie danych konieczne jest wydzielenie trzeciej tabeli, która będzie zawierała przypisanie przedmiotów nauczycielom.

Postaci normalne - określają stopień dekompozycji bazy danych.

Model relacyjny jest powszechnie uznawany za jedno z największych osiągnięć technicznych XX wieku, zrewolucjonizował on sposób postrzegania baz danych. Teoretyczne założenia modelu relacyjnego sformułował Edgar Frank Codd (1923-2003).

W swojej pracy na temat relacyjnego modelu danych E.F.Codd sformułował reguły projektowania relacyjnych baz danych (1970). Reguły te zostały pierwotnie nazwane postaciami normalnymi. Codd opisał trzy postacie normalne oznaczane często symbolami 1NF, 2NF, 3NF. Proces kolejnego przekształcania projektu bazy danych przez te trzy postacie normalne jest znany jako normalizacja bazy danych.

W połowie lat siedemdziesiątych spostrzeżono pewne niedostatki w trzeciej postaci normalnej Codd'a i zdefiniowano mocniejszą postać normalną, znaną jako postać normalna Boyce'a-Codda BCNF. Później Fagin przedstawił czwartą postać normalną, i piątą postać normalną.



Jest pięć (ale jedna z nich dodatkowo ma dwa warianty!) postaci normalnych:

(I, II, III, BCNF, IV, V).

Zwykle doprowadzenie bazy danych do trzeciej postaci normalnej wystarcza do uznania schematów relacji za wystarczające do implementacji.

## **Pierwsza postać normalna 1NF**

Relacja jest w pierwszej postaci normalnej, jeśli wartości atrybutów są elementarne tzn. są to pojedyncze wartości określonego typu, a nie zbiory wartości.

Pierwsza postać normalna jest konieczna aby, tabelę można było nazwać relacją.

Większość systemów baz danych nie ma możliwości zbudowania tabel nie będących w pierwszej postaci normalnej.

Przekształcenie z postaci nieznormalizowanej do pierwszej postaci normalnej ilustruje rysunek:

Id_zamowienie	Id_wydawca	Nazwa_wydawca	Adres_wydawca	Id_ksiazka	Tytul	Ilosc
1	2	PWE	ul. Podwale 17	5	Marketing	15
1	2	PWE		8	Ekonomia	20
2	3	PIW	ul. Foksal 17	4	Wiersze zebrane	10
2	4	PWRiL	ul. Malownicza 14	7	Hodowla lasu	50
3	4	PWRiL		3	Polubić pszczoły	40



Id_zamowienie	Id_wydawca	Nazwa_wydawca	Adres_wydawca	Id_ksiazka	Tytul	Ilosc
1	2	PWE	ul. Podwale 17	5	Marketing	15
1	2	PWE	ul. Podwale 17	8	Ekonomia	20
2	3	PIW	ul. Foksal 17	4	Wiersze zebrane	10
2	4	PWRiL	ul. Malownicza 14	7	Hodowla lasu	50
3	4	PWRiL	ul. Malownicza 14	3	Polubić pszczoły	40

Id_zamowienie	Id_wydawca	Nazwa_wydawca	Adres_wydawca	Id_ksiazka	Tytul	Ilosc
1	2	PWE	ul. Podwale 17	5	Marketing	15
1	2	PWE	ul. Podwale 17	8	Ekonomia	20
2	3	PIW	ul. Foksal 17	4	Wiersze zebrane	10
2	4	PWRiL	ul. Malownicza 14	7	Hodowla lasu	50
3	4	PWRiL	ul. Malownicza 14	3	Polubić pszczoły	40



Id_zamowienie	Id_wydawca	Nazwa_wydawca	Ulica	Nr	Id_ksiazka	Tytul	Ilosc
1	2	PWE	Podwale	17	5	Marketing	15
1	2	PWE	Podwale	17	8	Ekonomia	20
2	3	PIW	Foksal	17	4	Wiersze zebrane	10
2	4	PWRiL	Malownicza	14	7	Hodowla lasu	50
3	4	PWRiL	Malownicza	14	3	Polubić pszczoły	40

## **Druga postać normalna 2NF**

Relacja jest w drugiej postaci normalnej, jeżeli każdy atrybut wtórny (tzn. nie wchodzący w skład żadnego klucza potencjalnego) tej relacji jest w pełni funkcjonalnie zależny od wszystkich kluczy potencjalnych tej relacji.

Jeżeli w tabeli występują klucze złożone, to atrybuty niekluczowe muszą być w pełni funkcjonalnie zależne od całego klucza a nie jego atrybutu składowego.

Można zauważyć, że relacja **Zamówienie** nie jest w drugiej postaci normalnej, ponieważ atrybuty **Id\_wydawca**, **Nazwa\_wydawca**, **Ulica**, **Nr** i **Tytuł** nie są w pełni funkcjonalnie zależne od jedyne go klucza potencjalnego - pary (**Id\_zamowienie** oraz **Id\_ksiazka**).

FK ← PK → FK

Id_zamowienie	Id_wydawca	Nazwa_wydawca	Ulica	Nr	Id_ksiazka	Tytuł	Ilosc
1	2	PWE	Podwale	17	5	Marketing	15
1	2	PWE	Podwale	17	8	Ekonomia	20
2	3	PIW	Foksal	17	4	Wiersze zebrane	10
2	4	PWRiL	Malownicza	14	7	Hodowla lasu	50
3	4	PWRiL	Malownicza	14	3	Polubić pszczoły	40

W celu sprowadzenia relacji do drugiej postaci normalnej, należy podzielić ją na takie relacje, których wszystkie atrybuty będą w pełni funkcjonalnie zależne od jedyne go potencjalnego klucza głównego tej relacji. W tym celu przykładową relację **Zamówi ene** należy podzielić na trzy relacje: **Wydawca na zamówieniu**, **Zamówione książki** oraz **Książka** w następujący sposób:

#### Wydawca na zamówieniu

PK

Id_zamowienie	Id_wydawca	Nazwa_wydawca	Ulica	Nr
1	2	PWE	Podwale	17
2	3	PIW	Foksal	17
2	4	PWRiL	Malownicza	14
3	4	PWRiL	Malownicza	14

#### Zamówione książki

Id_zamowienie	Id_ksiazka	Ilosc
1	5	15
1	8	20
2	4	10
2	7	50
3	3	40

#### Książka

Id_ksiazka	Tytul
5	Marketing
8	Ekonomia
4	Wiersze zebrane
7	Hodowla lasu
3	Polubić pszczoły

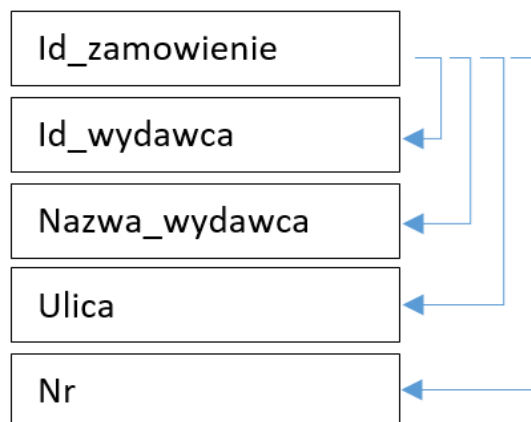


Jak widać wszystkie te trzy relacje są w drugiej postaci normalnej, ponieważ klucze relacji **Wydawca na zamówieniu** oraz **Książka** są kluczami prostymi, natomiast atrybut **Ilosc** w relacji **Zamówione książki** jest w pełni funkcjonalnie zależny od klucza złożonego (**Id\_zamowienie** oraz **Id\_ksiazka**).

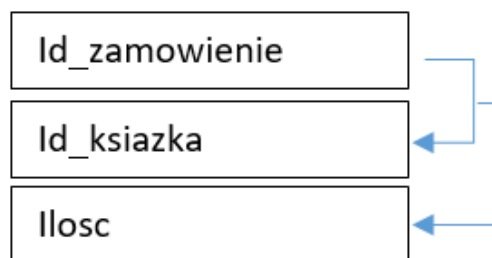
Należy zauważyć, że relacja będąca w pierwszej postaci normalnej, jest równocześnie w drugiej postaci normalnej, jeśli wszystkie jej klucze potencjalne są kluczami prostymi a tabela przechowuje informacje na ściśle określony temat – związane z jedną określoną encją.

Po przekształceniu relacji **Zamówienie** do drugiej postaci normalnej otrzymujemy następujące zależności funkcjonalne:

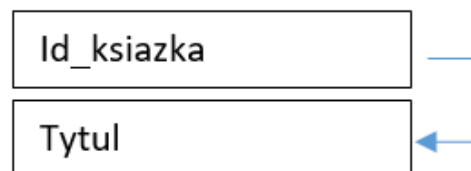
Wydawca na zamówieniu



Zamówione książki



Książki



## **Trzecia postać normalna 3NF**

Dana relacja jest w trzeciej postaci normalnej, jeśli jest ona w drugiej postaci normalnej i każdy jej atrybut nie wchodzący w skład żadnego klucza potencjalnego nie jest przechodnio funkcjonalnie zależny od żadnego klucza potencjalnego tej relacji.

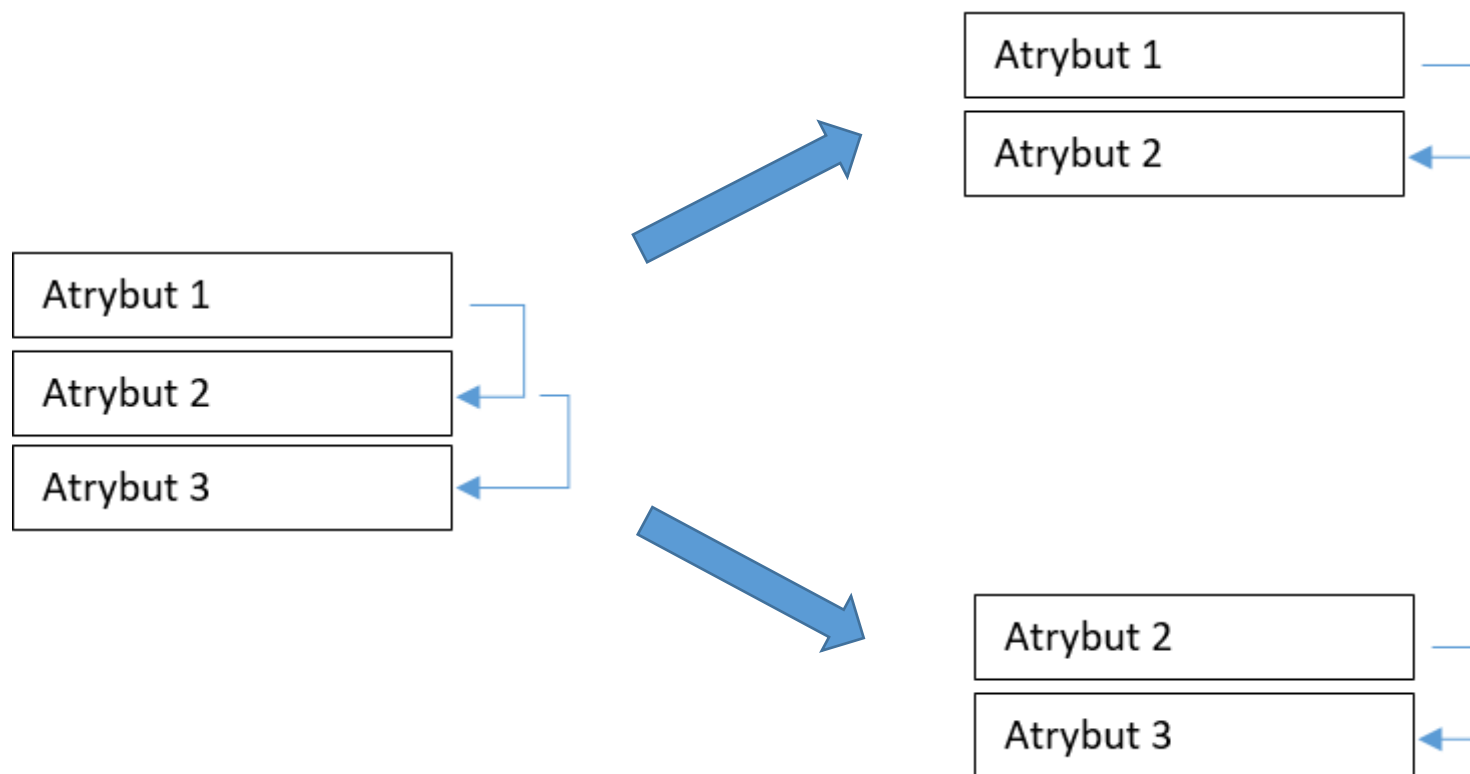
Aby doprowadzić relację, której atrybuty pozostają w przechodniej zależności funkcjonalnej do trzeciej postaci normalnej, należy podzielić ją na relacje zawierające tylko zależność funkcjonalną.

## **BCNF**

Każdy atrybut tabeli, od którego w pełni funkcyjnie zależy inny atrybut, musi być kluczem kandydującym

Doprowadzenie relacji w której występują przechodnie zależności funkcjonalne do trzeciej postaci normalnej wymaga podziału tak by te przechodnie zależności funkcjonalne wyeliminować.

Podział relacji przedstawia rysunek:



W opisywanym przykładzie przechodnia zależność funkcjonalna występuje pomiędzy atrybutami **Nazwa\_wydawca**, **Ulica** oraz **Nr** a atrybutem **Id\_zamowienie** w relacji **Wydawca na zamówieniu**. W związku z tym konieczne jest dokonanie podziału relacji **Wydawca na zamówieniu** na dwie relacje: **Zamówienie** oraz **Wydawca** w następujący sposób:

**Zamówienie**

Id_zamowienie	Id_wydawca
1	2
1	2
2	3
2	4
3	4

**Wydawca**

Id_wydawca	Nazwa_wydawca	Ulica	Nr
2	PWE	Podwale	17
2	PWE	Podwale	17
3	PIW	Foksal	17
4	PWRiL	Malownicza	14
4	PWRiL	Malownicza	14

## **Podsumowanie**

Przekształcenie relacji do kolejnych postaci normalnych wiąże się najczęściej ze zmniejszeniem ilości pamięci potrzebnej do przechowania informacji.

Proces normalizacji ma na celu takie przekształcenie relacji, by uniknąć dublowania informacji. Unikanie powtórzeń pozwala na łatwiejsze i w wielu przypadkach szybsze posługiwanie się bazą danych.