

Normalizacja

1NF, 2NF, 3NF, BCNF, 4NF, 5NF

Plan rozdziału

- Wzorce projektowe
- Zależności funkcyjne
- Postaci normalne
 - pierwsza postać normalna
 - druga postać normalna
 - trzecia postać normalna
 - postać normalna Boyce'a-Codda
- Dekompozycja relacji
- Zależności wielowartościowe
 - czwarta postać normalna
 - piąta postać normalna

Ocena poprawności schematu

- Poziom logiczny
 - czy użytkownicy rozumieją znaczenie relacji i atrybutów?
- Poziom fizyczny
 - jak składowane i modyfikowane są krotki w plikach dyskowych?
- Metodyki projektowania bazy danych
 - **bottom-up**: identyfikacja związków między poszczególnymi atrybutami i łączenie atrybutów w relacje (projektowanie przez syntezę)
 - **top-down**: utworzenie relacji na podstawie schematu pojęciowego i projektowanie przez analizę (dekompozycja relacji)

Wzorce projektowe

semantyka atrybutów relacji

wskazówka 1

Schemat relacji powinien być prosty, relacja powinna zawierać atrybuty jednej encji lub jednego związku, w przeciwnym wypadku semantyka atrybutów staje się niejasna.

PRACOWNICY	id_prac	nazwisko	etat	data_ur	id_zesp	nazwa	id_kierownika
	100	Biały	dyrektor	12 05 1950	10	Marketing	100
	110	Niebieski	sprzedawca	28 04 1972	10	Marketing	100
	120	Zielony	sprzedawca	03 07 1978	20	Sprzedaż	230

powyższy schemat relacji **PRACOWNICY** jest błędny, ponieważ zawiera atrybuty różnych encji, takie schematy podlegają wielu anomaliom.

Wzorce projektowe

redundancja informacji

PRACOWNICY

ID_PRAC	NAZWISKO	ETAT	ID_SZEFA	ZATRUDNI	PLACA_POD	PLACA_DOD	ID_ZESP
100	WEGLARZ	DYREKTOR		68/01/01	1730	420,5	10
110	BLAZEWICZ	PROFESOR	100	73/05/01	1350	210	40
120	SLOWINSKI	PROFESOR	100	77/09/01	1070		30
130	BRZEZINSKI	PROFESOR	100	68/07/01	960		20
140	MORZY	PROFESOR	130	75/09/15	830	105	20
150	KROLIKOWSKI	ADIUNKT	130	77/09/01	645,5		20
160	KOSZLAJDA	ADIUNKT	130	85/03/01	590		20
170	JEZERSKI	ASYSTENT	130	92/10/01	439,7	80,5	20
190	MATYSIAK	ASYSTENT	140	93/09/01	371		20
180	MAREK	SEKRETARKA	100	85/02/20	410,2		10
200	ZAKRZEWICZ	STAZYSTA	140	94/07/15	208		30
210	BIALY	STAZYSTA	130	93/10/15	250	170,6	30
220	KONOPKA	ASYSTENT	110	93/10/01	480		20
230	HAPKE	ASYSTENT	120	92/09/01	480	90	30

ZESPOLY

ID_ZESP	NAZWA	ADRES
10	ADMINISTRACJA	PIOTROWO 3A
20	SYSTEMY ROZP	PIOTROWO 3A
30	SYSTEMY EKSP	STRZELECKA 14
40	ALGORYTMY	WLODKOWICA 16
50	BADANIA OPERA	MIELZYNSKIEGO 30

informacje o nazwie i adresie każdego zespołu są redundantne i powtarzane dla każdego pracownika pracującego w danym zespole

PRACOWNICY2

ID_ZESP	ID_PRAC	NAZWISKO	ETAT	ID_SZEFA	ZATRUDNI	PLACA_POD	PLACA_DOD	NAZWA	ADRES
10	100	WEGLARZ	DYREKTOR		68/01/01	1730	420,5	ADMINISTRACJA	PIOTROWO 3A
40	110	BLAZEWICZ	PROFESOR	100	73/05/01	1350	210	ALGORYTMY	WLODKOWICA 16
30	120	SLOWINSKI	PROFESOR	100	77/09/01	1070		SYSTEMY EKSPERCKIE	STRZELECKA 14
20	130	BRZEZINSKI	PROFESOR	100	68/07/01	960		SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	140	MORZY	PROFESOR	130	75/09/15	830	105	SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	150	KROLIKOWSKI	ADIUNKT	130	77/09/01	645,5		SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	160	KOSZLAJDA	ADIUNKT	130	85/03/01	590		SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	170	JEZERSKI	ASYSTENT	130	92/10/01	439,7	80,5	SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	190	MATYSIAK	ASYSTENT	140	93/09/01	371		SYSTEMY ROZPROSZONE	PIOTROWO 3A
10	180	MAREK	SEKRETARKA	100	85/02/20	410,2		ADMINISTRACJA	PIOTROWO 3A
30	200	ZAKRZEWICZ	STAZYSTA	140	94/07/15	208		SYSTEMY EKSPERCKIE	STRZELECKA 14
30	210	BIALY	STAZYSTA	130	93/10/15	250	170,6	SYSTEMY EKSPERCKIE	STRZELECKA 14
20	220	KONOPKA	ASYSTENT	110	93/10/01	480		SYSTEMY ROZPROSZONE	PIOTROWO 3A
30	230	HAPKE	ASYSTENT	120	92/09/01	480	90	SYSTEMY EKSPERCKIE	STRZELECKA 14

Wzorce projektowe

anomalie modyfikowania

- Anomalie wstawiania

- wstawienie nowego pracownika do relacji PRACOWNICY2 wymaga podania **poprawnych** i **spójnych** danych o zespole, w którym pracuje pracownik (lub wartości pustych)
- wstawienie zespołu, w którym jeszcze nikt nie pracuje, wymaga wstawienia pustych wartości do wszystkich atrybutów opisujących pracownika; co gorsza, późniejsze dodanie pierwszego pracownika w tym zespole wymaga modyfikacji istniejącej krotki (a nie wstawienia!)

PRACOWNICY2

ID_ZESP	ID_PRAC	NAZWISKO	ETAT	ID_SZEFA	ZATRUDNI	PLACA_POD	PLACA_DOD	NAZWA	ADRES
10	100	WEGLARZ	DYREKTOR		68/01/01	1730	420,5	ADMINISTRACJA	PIOTROWO 3A
40	110	BLAZEWICZ	PROFESOR	100	73/05/01	1350	210	ALGORYTMY	WLODKOWICA 16
30	120	SLOWINSKI	PROFESOR	100	77/09/01	1070		SYSTEMY EKSPERCKIE	STRZELECKA 14
20	130	BRZEZINSKI	PROFESOR	100	68/07/01	960		SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	140	MORZY	PROFESOR	130	75/09/15	830	105	SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	150	KROLIKOWSKI	ADIUNKT	130	77/09/01	645,5		SYSTEMY ROZPROSZONE	PIOTROWO 3A

20	160	KOSZLAJDA	ADIUNKT	130	85/03/01	590		SYSTEMY ROZPROSZONE	STRZELECKA 14
----	-----	-----------	---------	-----	----------	-----	--	---------------------	---------------

null	null	null	null	null	null	null	null	BAZY DANYCH	BERDYCHOWO 1
------	------	------	------	------	------	------	------	-------------	--------------

Wzorce projektowe

anomalie modyfikowania

- Anomalia usuwania
 - usunięcie ostatniego pracownika z zespołu usuwa wszystkie informacje o tym zespole

PRACOWNICY_I_ZESPOLY

ID_ZESP	ID_PRAC	NAZWISKO	ETAT	ID_SZEFA	ZATRUDNI	PLACA_POD	PLACA_DOD	NAZWA	ADRES
10	100	WEGLARZ	DYREKTOR		68/01/01	1730	420,5	ADMINISTRACJA	PIOTROWO 3A
40	440	BLAZEWICZ	PROFESOR	100	73/05/04	4350	240	ALGORYTMY	WLODKOWICA 16
30	120	SLOWINSKI	PROFESOR	100	77/09/01	1070		SYSTEMY EKSPERCKIE	STRZELECKA 14
20	130	BRZEZINSKI	PROFESOR	100	68/07/01	960		SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	140	MORZY	PROFESOR	130	75/09/15	830	105	SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	150	KROLIKOWSKI	ADIUNKT	130	77/09/01	645,5		SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	160	KOSZLAJDA	ADIUNKT	130	85/03/01	590		SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	170	JEZERSKI	ASYSTENT	130	92/10/01	439,7	80,5	SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	190	MATYSIAK	ASYSTENT	140	93/09/01	371		SYSTEMY ROZPROSZONE	PIOTROWO 3A
10	180	MAREK	SEKRETARKA	100	85/02/20	410,2		ADMINISTRACJA	PIOTROWO 3A
30	200	ZAKRZEWICZ	STAZYSTA	140	94/07/15	208		SYSTEMY EKSPERCKIE	STRZELECKA 14
30	210	BIALY	STAZYSTA	130	93/10/15	250	170,6	SYSTEMY EKSPERCKIE	STRZELECKA 14
20	220	KONOPKA	ASYSTENT	110	93/10/01	480		SYSTEMY ROZPROSZONE	PIOTROWO 3A
30	230	HAPKE	ASYSTENT	120	92/09/01	480	90	SYSTEMY EKSPERCKIE	STRZELECKA 14

Wzorce projektowe

anomalie modyfikowania

- Anomalia modyfikacji
 - zmiana wartości atrybutu opisującego zespół w jednej krotce musi być propagowana do wszystkich krotek opisujących dany zespół

PRACOWNICY_I_ZESPOLY

ID_ZESP	ID_PRAC	NAZWISKO	ETAT	ID_SZEFA	ZATRUDNI	PLACA_POD	PLACA_DOD	NAZWA	ADRES
10	100	WEGLARZ	DYREKTOR		68/01/01	1730	420,5	ADMINISTRACJA	PIOTROWO 2
40	110	BLAZEWICZ	PROFESOR	100	73/05/01	1350	210	ALGORYTMY	WLODKOWICA 16
30	120	SLOWINSKI	PROFESOR	100	77/09/01	1070		SYSTEMY EKSPERCKIE	STRZELECKA 14
20	130	BRZEZINSKI	PROFESOR	100	68/07/01	960		SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	140	MORZY	PROFESOR	130	75/09/15	830	105	SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	150	KROLIKOWSKI	ADIUNKT	130	77/09/01	645,5		SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	160	KOSZLAJDA	ADIUNKT	130	85/03/01	590		SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	170	JEZERSKI	ASYSTENT	130	92/10/01	439,7	80,5	SYSTEMY ROZPROSZONE	PIOTROWO 3A
20	190	MATYSIAK	ASYSTENT	140	93/09/01	371		SYSTEMY ROZPROSZONE	PIOTROWO 3A
10	180	MAREK	SEKRETARKA	100	85/02/20	410,2		ADMINISTRACJA	PIOTROWO 3A
30	200	ZAKRZEWICZ	STAZYSTA	140	94/07/15	208		SYSTEMY EKSPERCKIE	STRZELECKA 14
30	210	BIALY	STAZYSTA	130	93/10/15	250	170,6	SYSTEMY EKSPERCKIE	STRZELECKA 14
20	220	KONOPKA	ASYSTENT	110	93/10/01	480		SYSTEMY ROZPROSZONE	PIOTROWO 3A
30	230	HAPKE	ASYSTENT	120	92/09/01	480	90	SYSTEMY EKSPERCKIE	STRZELECKA 14

wskazówka 2

Schemat relacji powinien eliminować anomalie wstawiania, modyfikowania i usuwania. Jeśli anomalie występują w schemacie relacji, muszą być udokumentowane i zakomunikowane użytkownikom.

Wzorce projektowe wartości puste

- Relacje zawierające wiele atrybutów dopuszczających wartości puste
 - marnują przestrzeń dyskową,
 - komplikują semantykę atrybutów,
 - komplikują wyliczenia za pomocą funkcji grupowych,
 - powodują błędne wyniki połączeń.
- Atrybuty z dużą liczbą wartości pustych powinny być umieszczane w osobnych relacjach

wskazówka 3

W schemacie relacji należy unikać atrybutów, które mogą mieć wartość pustą dla większości krotek. Jeśli wartości puste są nieuniknione, powinny dotyczyć tylko wyjątkowych sytuacji.

Nadmiarowe krotki

AKTORZY_TEATRY_SEZONY

ID_AKTORA	NAZWISKO	TEATR	MIASTO	SEZON
10	ŁOMNICKI	NOWY	POZNAŃ	1990
20	SENIUK	NARODOWY	WARSZAWA	2000
30	JANDA	NARODOWY	WARSZAWA	2001
40	GAJOS	STARY	KRAKÓW	1998
10	ŁOMNICKI	DRAMATYCZNY	KRAKÓW	1989
30	JANDA	STARY	KRAKÓW	2002

AKTORZY_MIASTA

NAZWISKO	MIASTO
ŁOMNICKI	POZNAŃ
SENIUK	WARSZAWA
JANDA	WARSZAWA
GAJOS	KRAKÓW
ŁOMNICKI	KRAKÓW
JANDA	KRAKÓW

AKTORZY_TEATRY

ID_AKTORA	TEATR	MIASTO	SEZON
10	NOWY	POZNAŃ	1990
20	NARODOWY	WARSZAWA	2000
30	NARODOWY	WARSZAWA	2001
40	STARY	KRAKÓW	1998
10	DRAMATYCZNY	KRAKÓW	1989
30	STARY	KRAKÓW	2002

ID_AKTORA	NAZWISKO	TEATR	MIASTO	SEZON
10	ŁOMNICKI	NOWY	POZNAŃ	1990
20	SENIUK	NARODOWY	WARSZAWA	2000
30	JANDA	NARODOWY	WARSZAWA	2001
40	GAJOS	STARY	KRAKÓW	1998
10	ŁOMNICKI	DRAMATYCZNY	KRAKÓW	1989
30	JANDA	STARY	KRAKÓW	2002
30	SENIUK	NARODOWY	WARSZAWA	2001
20	JANDA	NARODOWY	WARSZAWA	2000
10	GAJOS	DRAMATYCZNY	KRAKÓW	1989
30	GAJOS	STARY	KRAKÓW	2002
40	ŁOMNICKI	STARY	KRAKÓW	1998
30	ŁOMNICKI	STARY	KRAKÓW	2002
40	JANDA	STARY	KRAKÓW	1998
10	JANDA	DRAMATYCZNY	KRAKÓW	2002

wskazówka 4

Relacje muszą się łączyć przez połączenie równościowe na atrybutach będących kluczami podstawowymi lub obcymi

Zależność funkcyjna

Zależność funkcyjna (ang. *functional dependency*) $X \rightarrow Y$ między zbiorami atrybutów X i Y relacji R nakłada ograniczenie na wartości atrybutów w dowolnym stanie r relacji R . Zależność funkcyjna $X \rightarrow Y$ występuje gdy $\forall t_1, t_2 \in r(R) : t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y]$

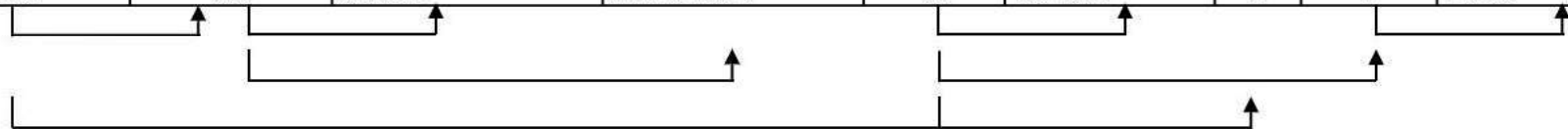
Atrybut Y relacji R jest zależny funkcyjnie od atrybutu X relacji R jeśli w dowolnym stanie relacji $r(R)$ każdej wartości x atrybutu X odpowiada nie więcej niż jedna wartość y atrybutu Y . Definicję można uogólnić na zbiory atrybutów.

- jeśli występuje $X \rightarrow Y$ to mówimy, że X **wyznacza funkcyjnie** (ang. *functionally determines*) Y , lub że Y **zależy funkcyjnie** (ang. *functionally dependent*) od X
- istnienie zależności funkcyjnej $X \rightarrow Y$ jest niezależne od stanu relacji $r(R)$
- jeśli X jest kluczem schematu R to dla każdego podzbioru atrybutów Y relacji R występuje $X \rightarrow Y$
- zależność funkcyjna $X \rightarrow Y$ jest cechą semantyki (znaczenia) atrybutów X i Y

Zależności funkcyjne - przykład

ZAMOWIENIA

NR_ZAMOWIENIA	ID_DOSTAWCY	NAZWA_DOSTAWCY	ADRES_DOSTAWCY	ID_CZESCI	NAZWA_CZESCI	ILOSC	MAGAZYN	ADRES_MAGAZYNU
1001	300	IBM	WARSZAWA	53	RDBMS	2	5	WARSZAWA
1001	300	IBM	WARSZAWA	55	T52	10	3	WARSZAWA
1001	300	IBM	WARSZAWA	57	R52	15	3	WARSZAWA
1002	400	ORACLE	POZNAN	34	iAS	5	2	SWADZIM
1002	400	ORACLE	POZNAN	36	JDEV	30	2	SWADZIM
1003	500	EMAX	KRAKOW	88	MYSZKA	100	4	KRAKOW
1004	600	ORACLE	WARSZAWA	53	RDBMS	5	5	WARSZAWA
1004	600	ORACLE	WARSZAWA	59	TOPLINK	2	8	PLOCK



NR_ZAMOWIENIA → ID_DOSTAWCY
 ID_DOSTAWCY → NAZWA_DOSTAWCY
 ID_DOSTAWCY → ADRES_DOSTAWCY
 ID_CZESCI → NAZWA_CZESCI
 ID_CZESCI → MAGAZYN
 MAGAZYN → ADRES_MAGAZYNU
 NR_ZAMOWIENIA, ID_CZESCI → ILOSC

zamówienie do jednego dostawcy
 jednoznaczna identyfikacja dostawcy
 jednoznaczna identyfikacja dostawcy
 jednoznaczna identyfikacja części
 każda część tylko w jednym magazynie
 jednoznaczna identyfikacja magazynu
 każda część tylko raz na zamówieniu

Reguły wywodzenia zależności funkcyjnych

Zależność funkcyjna $X \rightarrow Y$ jest wywiedziona (ang. *inferred*) ze zbioru zależności funkcyjnych F , co oznaczamy przez $F \models X \rightarrow Y$, jeśli zależność $X \rightarrow Y$ jest spełniona w każdym poprawnym stanie $r(R)$, tzn. zawsze gdy $r(R)$ spełnia wszystkie zależności funkcyjne ze zbioru F , zależność $X \rightarrow Y$ także jest spełniona.

Domknięciem F^+ (ang. *closure*) zbioru zależności F jest zbiór wszystkich zależności funkcyjnych które można wywieść z F .

- Reguły wywodzenia zależności
 - IR_1 (reguła zwrotna): jeśli $X \supseteq Y$ to $X \rightarrow Y$
 - IR_2 (reguła zwiększania): $X \rightarrow Y \models XZ \rightarrow YZ$
 - IR_3 (reguła przechodnia): $X \rightarrow Y, Y \rightarrow Z \models X \rightarrow Z$
 - IR_4 (reguła dekompozycji): $X \rightarrow YZ \models X \rightarrow Y$
 - IR_5 (reguła addytywna): $X \rightarrow Y, X \rightarrow Z \models X \rightarrow YZ$
 - IR_6 (reguła pseudo-tranzytywna): $X \rightarrow Y, WY \rightarrow Z \models WX \rightarrow Z$

Zależności funkcyjne - dalsze definicje

Zbiór zależności funkcyjnych F pokrywa zbiór zależności funkcyjnych E jeśli każda zależność funkcyjna $e \in E$ należy do F^+

Zbiór zależności funkcyjnych E jest równoważny (ekwiwalentny) zbiorowi zależności funkcyjnych F jeśli $E^+ = F^+$

Zbiór zależności funkcyjnych F jest minimalny jeśli

- każda zależność funkcyjna $f \in F$ ma po prawej stronie tylko jeden atrybut
- nie można zastąpić żadnej zależności $X \rightarrow A$ w F przez zależność $Y \rightarrow A$, gdzie $Y \subset X$ i zachować równoważność z F
- nie można usunąć żadnej zależności $X \rightarrow Y$ z F i zachować równoważność z F

Normalizacja

Normalizacja to proces dekompozycji schematów relacji posiadających niepożądane cechy do zbiorów mniejszych schematów charakteryzujących się pożądanymi własnościami. Proces normalizacji jest rozumiany jako proces ulepszania schematu bazy danych przez eliminację anomalii.

- Proces normalizacji musi posiadać następujące własności:
 - żaden atrybut nie zostanie zgubiony w trakcie normalizacji,
 - dekompozycja relacji nie prowadzi do utraty informacji,
 - wszystkie zależności funkcyjne są reprezentowane w pojedynczych schematach relacji.

Repetitorium

- Każdy podzbiór SK atrybutów relacji R taki, że dla dowolnych dwóch krotek jest spełnione $t_1[SK] \neq t_2[SK]$ dla dowolnego stanu relacji $r(R)$ nazywamy **nadkluczem** relacji R.
- Każda relacja posiada co najmniej jeden nadklucz, jest nim zbiór wszystkich atrybutów relacji.
- **Kluczem** K schematu relacji R nazywamy minimalny nadklucz schematu relacji R, tzn. taki nadklucz, dla którego nie istnieje nadklucz K' schematu relacji R taki, że $K' \subset K$.
- Jeśli relacja R zawiera wiele kluczy, to nazywane są one **kluczami kandydującymi**. Najczęściej jeden klucz jest wybierany jako **klucz główny**, a pozostałe stają się **kluczami wtórnymi**.
- Atrybut wchodzący w skład dowolnego klucza kandydującego nazywa się **atrybutem podstawowym**. Atrybut nie będący atrybutem podstawowym nazywa się **atrybutem wtórnym**.

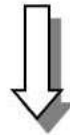
Pierwsza postać normalna (1NF)

Relacja R jest w pierwszej postaci normalnej jeśli wartości wszystkich atrybutów relacji R są atomowe (niepodzielne).

ID_PRAC	NAZWISKO	JEZYKI
100	BIALY	{ANGIELSKI, FRANCUSKI}
101	NIEBIESKI	{ANGIELSKI, NIEMIECKI, ROSYJSKI}
102	ZIELONY	{FRANCUSKI, HISPANSKI}



ID_PRAC	NAZWISKO	JEZYK
100	BIALY	ANGIELSKI
100	BIALY	FRANCUSKI
101	NIEBIESKI	ANGIELSKI
101	NIEBIESKI	NIEMIECKI
101	NIEBIESKI	ROSYJSKI
102	ZIELONY	FRANCUSKI
102	ZIELONY	HISPANSKI



ID_PRAC	NAZWISKO	JEZYK1	JEZYK2	JEZYK3
100	BIALY	ANGIELSKI	FRANCUSKI	
101	NIEBIESKI	ANGIELSKI	NIEMIECKI	ROSYJSKI
102	ZIELONY	FRANCUSKI	HISPANSKI	



ID_PRAC	NAZWISKO
100	BIALY
101	NIEBIESKI
102	ZIELONY

ID_PRAC	JEZYK
100	ANGIELSKI
100	FRANCUSKI
101	ANGIELSKI
101	NIEMIECKI
101	ROSYJSKI
102	FRANCUSKI
102	HISPANSKI

Pełna zależność funkcyjna

Atrybut A w relacji R jest w pełni zależny funkcyjnie od zbioru atrybutów X , jeśli $X \rightarrow A$ i A nie jest funkcyjnie zależny od żadnego podzbioru właściwego X .

Zbiór atrybutów Y w relacji R jest w pełni zależny funkcyjnie od zbioru atrybutów X , jeśli $X \rightarrow Y$ i nie istnieje podzbiór właściwy $X' \subset X$ taki, że $X' \rightarrow Y$



POWIAT	NUMER	WOJEWÓDZTWO	MARKA	KOLOR
PSZ	9834	wielkopolskie	audi	czarny
POB	7873G	wielkopolskie	saab	srebrny
DW	089GY	dolnośląskie	ford	biały
DJ	6752W	dolnośląskie	bmw	czarny
DJA	02390	dolnośląskie	audi	niebieski

Druga postać normalna (2NF)

Relacja R jest w drugiej postaci normalnej jeśli każdy atrybut wtórny A w schemacie relacji R jest w pełni funkcyjnie zależny od klucza podstawowego relacji R.

Relacja R jest w drugiej postaci normalnej jeśli żaden atrybut wtórny A w schemacie relacji R nie jest częściowo funkcyjnie zależny od żadnego klucza relacji R.

INDEKS	NAZWISKO	IMIE	KIERUNEK_	WYDZIAŁ	ADRES_DZIEKANATU
1001	BIALY	ADAM	INFORMATYKA	ELEKTRYCZNY	PIOTROWO 3A
1001	BIALY	ADAM	ELEKTRONIKA	ELEKTRYCZNY	PIOTROWO 3A
1002	POMARANCZOWY	APOLONIUSZ	INFORMATYKA	ELEKTRYCZNY	PIOTROWO 3A
1002	POMARANCZOWY	APOLONIUSZ	URBANISTYKA	ARCHITEKTURA	STRZELECKA 12
1003	ZIELONY	ALFONS	BUDOWA DRÓG	ARCHITEKTURA	STRZELECKA 12

Transformacja do 2NF

<u>INDEKS</u>	<u>NA ZWISKO</u>	<u>IMIE</u>	<u>KIERUNEK</u>	<u>WYDZIAŁ</u>	<u>ADRES_DZIEKANATU</u>
1001	BIALY	ADAM	INFORMATYKA	ELEKTRYCZNY	PIOTROWO 3A
1001	BIALY	ADAM	ELEKTRONIKA	ELEKTRYCZNY	PIOTROWO 3A
1002	POMARANCZOWY	APOLONIUSZ	INFORMATYKA	ELEKTRYCZNY	PIOTROWO 3A
1002	POMARANCZOWY	APOLONIUSZ	URBANISTYKA	ARCHITEKTURA	STRZELECKA 12
1003	ZIELONY	ALFONS	BUDOWA DRÓG	ARCHITEKTURA	STRZELECKA 12

<u>INDEKS</u>	<u>NA ZWISKO</u>	<u>IMIE</u>	<u>KIERUNEK</u>
1001	BIALY	ADAM	INFORMATYKA
1001	BIALY	ADAM	ELEKTRONIKA
1002	POMARANCZOWY	APOLONIUSZ	INFORMATYKA
1002	POMARANCZOWY	APOLONIUSZ	URBANISTYKA
1003	ZIELONY	ALFONS	BUDOWA DRÓG

<u>KIERUNEK</u>	<u>WYDZIAŁ</u>	<u>ADRES_DZIEKANATU</u>
INFORMATYKA	ELEKTRYCZNY	PIOTROWO 3A
ELEKTRONIKA	ELEKTRYCZNY	PIOTROWO 3A
URBANISTYKA	ARCHITEKTURA	STRZELECKA 12
BUDOWA DRÓG	ARCHITEKTURA	STRZELECKA 12

<u>INDEKS</u>	<u>NA ZWISKO</u>	<u>IMIE</u>
1001	BIALY	ADAM
1002	POMARANCZOWY	APOLONIUSZ
1003	ZIELONY	ALFONS

<u>INDEKS</u>	<u>KIERUNEK</u>
1001	INFORMATYKA
1001	ELEKTRONIKA
1002	INFORMATYKA
1002	URBANISTYKA
1003	BUDOWA DRÓG

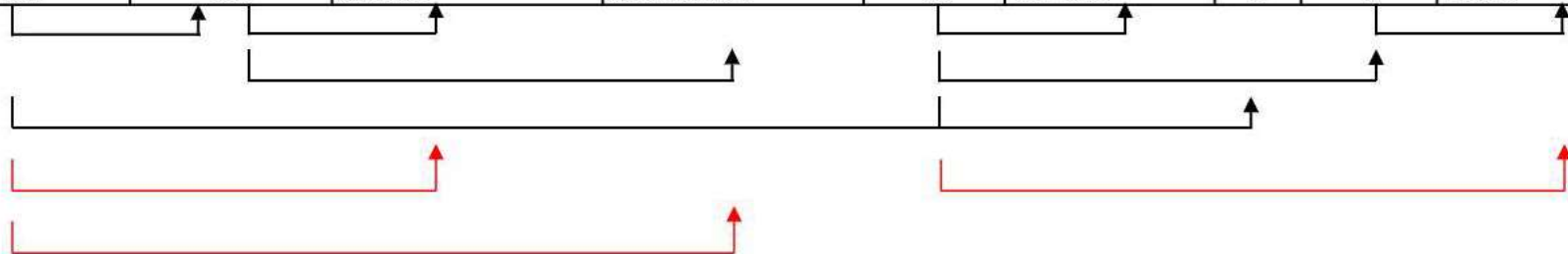
<u>KIERUNEK</u>	<u>WYDZIAŁ</u>	<u>ADRES_DZIEKANATU</u>
INFORMATYKA	ELEKTRYCZNY	PIOTROWO 3A
ELEKTRONIKA	ELEKTRYCZNY	PIOTROWO 3A
URBANISTYKA	ARCHITEKTURA	STRZELECKA 12
BUDOWA DRÓG	ARCHITEKTURA	STRZELECKA 12

Przechodnia zależność funkcyjna

Atrybut A w relacji R jest przechodnio funkcyjnie zależny od zbioru atrybutów X , jeśli $X \rightarrow A$ i istnieje zbiór atrybutów Z nie będący kluczem ani podzbiorem żadnego klucza relacji R , i zachodzi $X \rightarrow Z$ i jednocześnie $Z \rightarrow A$. Definicja może być uogólniona na zbiór atrybutów.

ZAMOWIENIA

NR_ZAMOWIENIA	ID_DOSTAWCY	NAZWA_DOSTAWCY	ADRES_DOSTAWCY	ID_CZESCI	NAZWA_CZESCI	ILOSC	MAGAZYN	ADRES_MAGAZYNU
1001	300	IBM	WARSZAWA	53	RDBMS	2	5	WARSZAWA
1001	300	IBM	WARSZAWA	55	T52	10	3	WARSZAWA
1001	300	IBM	WARSZAWA	57	R52	15	3	WARSZAWA
1002	400	ORACLE	POZNAN	34	iAS	5	2	SWADZIM
1002	400	ORACLE	POZNAN	36	JDEV	30	2	SWADZIM
1003	500	EMAX	KRAKOW	88	MYSZKA	100	4	KRAKOW
1004	600	ORACLE	WARSZAWA	53	RDBMS	5	5	WARSZAWA
1004	600	ORACLE	WARSZAWA	59	TOPLINK	2	8	PLOCK



Trzecia postać normalna (3NF)

Relacja R jest w trzeciej postaci normalnej jeśli dla każdej zależności funkcyjnej $X \rightarrow A$ w R spełniony jest jeden z warunków:

- X jest nadkluczem schematu relacji R,
- A jest atrybutem podstawowym schematu relacji R.

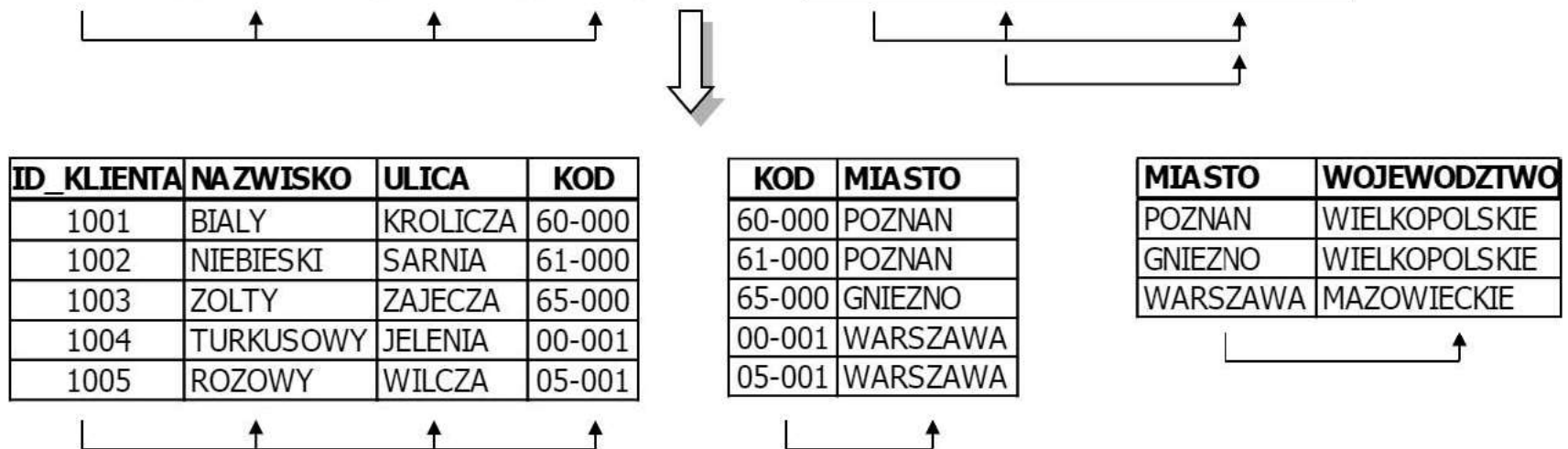
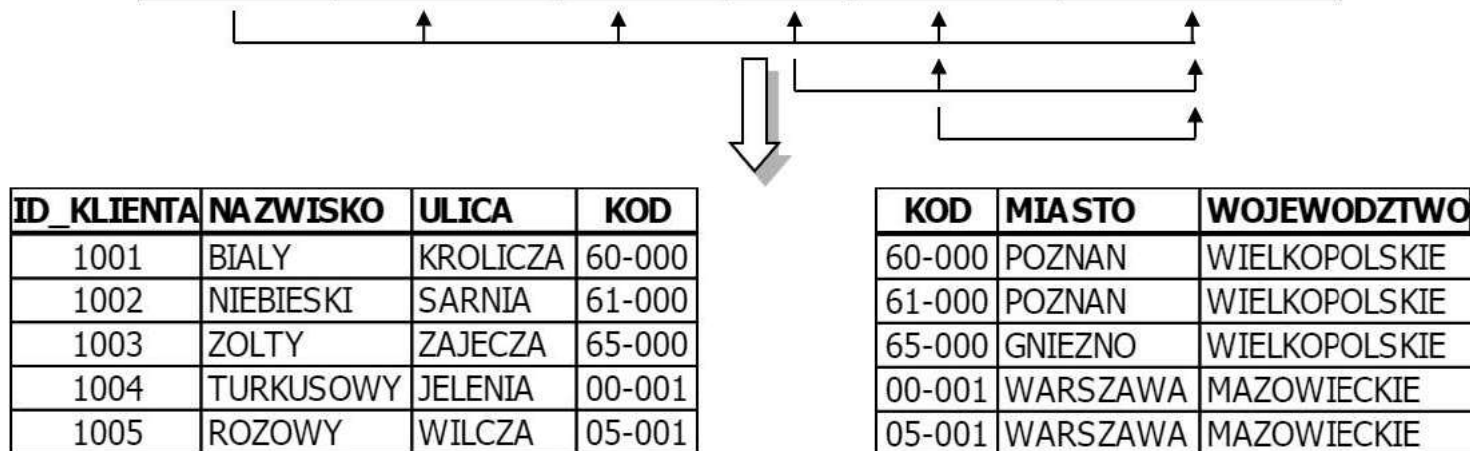
Relacja R jest w trzeciej postaci normalnej jeśli R jest w drugiej postaci normalnej i żaden atrybut wtórny relacji R nie jest przechodnio funkcyjnie zależny od klucza podstawowego schematu relacji R.

ID_KLIENTA	NAZWISKO	ULICA	KOD	MIASTO	WOJEWODZTWO
1001	BIALY	KROLICZA	60-000	POZNAN	WIELKOPOLSKIE
1002	NIEBIESKI	SARNIA	61-000	POZNAN	WIELKOPOLSKIE
1003	ZOLTY	ZAJECZA	65-000	GNIEZNO	WIELKOPOLSKIE
1004	TURKUSOWY	JELENIA	00-001	WARSZAWA	MAZOWIECKIE
1005	ROZOWY	WILCZA	05-001	SULECHOW	MAZOWIECKIE



Transformacja do 3NF

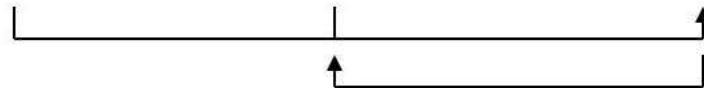
ID_KLIENTA	NAZWISKO	ULICA	KOD	MIASTO	WOJEWODZTWO
1001	BIALY	KROLICZA	60-000	POZNAN	WIELKOPOLSKIE
1002	NIEBIESKI	SARNIA	61-000	POZNAN	WIELKOPOLSKIE
1003	ZOLTY	ZAJECZA	65-000	GNIEZNO	WIELKOPOLSKIE
1004	TURKUSOWY	JELENIA	00-001	WARSZAWA	MAZOWIECKIE
1005	ROZOWY	WILCZA	05-001	WARSZAWA	MAZOWIECKIE



Postać normalna Boyce'a-Codda (BCNF)

Relacja R jest w postaci normalnej Boyce'a-Codda, jeśli dla każdej nietrywialnej zależności funkcyjnej $X \rightarrow A$ w R X jest nadkluczem schematu relacji R.

STUDENT	PRZEDMIOT	WYKŁADOWCA
MOCNY	BAZY DANYCH	NIEBIESKI
SILNY	BAZY DANYCH	ZIELONY
SLABY	SYSTEMY OPERACYJNE	POMARANCZOWY
SILNY	TEORIA PROGRAMOWANIA	FIOLETOWY
SLABY	BAZY DANYCH	NIEBIESKI
SREDNI	BAZY DANYCH	ZIELONY
BARDZOSLABY	SYSTEMY OPERACYJNE	POMARANCZOWY

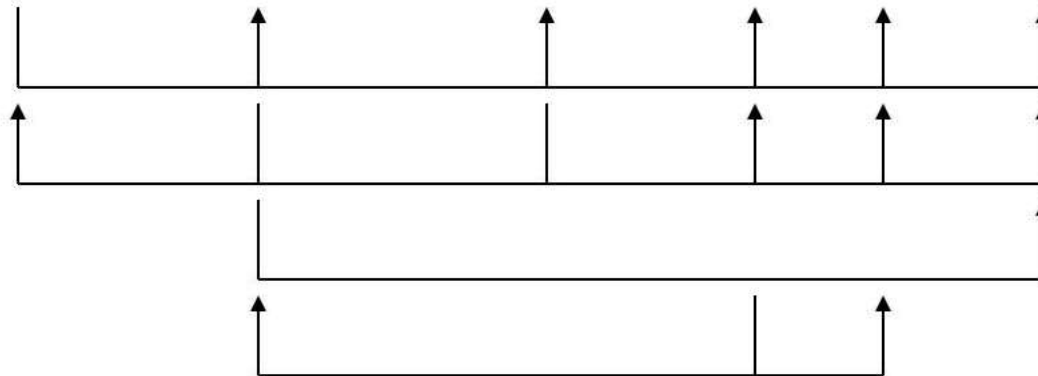


- Uwagi
 - każda relacja, która jest w BCNF, jest także w 3NF
 - dekompozycja relacji powinna prowadzić przynajmniej do 3NF lub BCNF

Zadanie

DZIALKI

ID_DZIALKI	GMINA	ID_OBSZARU	POW.	CENA	PODATEK
1001	WACHOCK	100	1000	50	2%
1002	WACHOCK	101	1000	50	2%
1003	PCIM	100	1250	45	1%
1004	PCIM	200	800	85	1%
1005	ROKIETNICA	101	1100	50	3%
1006	ROKIETNICA	102	1200	45	3%



- wyznacz klucz (klucze) relacji **DZIALKI**
- znormalizuj relację do 2NF, 3NF i BCNF

Zadania

- Dane są zbiory zależności funkcyjnych $F = \{A \rightarrow C, AC \rightarrow D, E \rightarrow AD, E \rightarrow H\}$ oraz $G = \{A \rightarrow CD, E \rightarrow AH\}$. Czy F i G są równoważne?
- Udowodnij, że dowolna relacja składająca się z dwóch atrybutów jest zawsze w BCNF.
- Dana jest relacja $R = \{A, B, C, D, E\}$ oraz zbiór zależności funkcyjnych $F = \{AB \rightarrow C, CD \rightarrow E, DE \rightarrow B\}$.
 - Czy AB jest kluczem relacji R ?
 - Czy ABD jest kluczem relacji R ?
- Dana jest relacja $R = \{A, B, C, D, E, F, G, H, I, J\}$ oraz zbiór zależności funkcyjnych $F = \{AB \rightarrow C, A \rightarrow DE, B \rightarrow F, F \rightarrow GH, D \rightarrow IJ\}$, podaj klucz (klucze) relacji R , zdekomponuj R do 2NF i następnie do 3NF.

Zadania

- Dane są relacje

`Orders(order#, date, customer#, total)`

`OrderItems(order#, item#, qty, price, discount)`

każdy element ma osobną zniżkę. Jak wygląda schemat po połączeniu naturalnym obu relacji? Co jest kluczem tego schematu? Wskaż zależności funkcyjne występujące w tym schemacie. Czy schemat jest w 2NF? Czy schemat jest w 3NF? Uzasadnij każdą odpowiedź.

- Dana jest relacja

`Cars(car#, salesman#, date, commission, discount)`

w której kluczem jest $\{c\#, \text{salesman}\# \}$, dane są zależności funkcyjne $\text{date} \rightarrow \text{discount}$ i $\text{salesman}\# \rightarrow \text{commission}$. Czy relacja `Cars` jest w 1NF, 2NF czy 3NF? Dokonaj pełnej normalizacji tej relacji.

Pożądane cechy dekompozycji

- Zachowanie atrybutów

Dekompozycja $D = \{R_1, \dots, R_n\}$ uniwersalnej relacji R posiada cechę **zachowywania atrybutów**, jeśli $\bigcup_i R_i = R$

- Zachowywanie zależności

Dane są zbiór zależności funkcyjnych F i uniwersalna relacja R .
Projekcją F na R_i $\pi_F(R_i)$, gdzie $R_i \subset R$, jest zbiór zależności $X \rightarrow Y$ w F^+ takich, że $X \cup Y$ są zawarte w R_i .

Dekompozycja $D = \{R_1, \dots, R_n\}$ uniwersalnej relacji R posiada cechę **zachowywania zależności w stosunku do zbioru zależności funkcyjnych F** , jeśli $F^+ = \pi_F(R_1) \cup \dots \cup \pi_F(R_n)$

Pożądane cechy dekompozycji

- Bezstratne połączenie

Dekompozycja $D = \{R_1, \dots, R_n\}$ uniwersalnej relacji R posiada cechę **bezstratnego połączenia w stosunku do zbioru zależności funkcyjnych F** , jeśli $\ast(\pi_{\langle R_1 \rangle}(r), \dots, \pi_{\langle R_n \rangle}(r)) = r$ dla każdego stanu r relacji R

cecha 1

Dekompozycja $D = \{R_1, R_2\}$ relacji R posiada cechę bezstratnego połączenia w stosunku do zbioru zależności funkcyjnych F , wtedy i tylko wtedy gdy:

- $(R_1 \cap R_2) \rightarrow (R_1 - R_2)$ znajduje się w F^+ , lub
- $(R_1 \cap R_2) \rightarrow (R_2 - R_1)$ znajduje się w F^+

cecha 2

Jeśli dekompozycja $D = \{R_1, \dots, R_m\}$ relacji R posiada cechę bezstratnego połączenia w stosunku do zbioru zależności funkcyjnych F i jeśli $D' = \{Q_1, \dots, Q_k\}$ relacji R_i posiada cechę bezstratnego połączenia w stosunku do projekcji zbioru zależności funkcyjnych F na R_i , to $D'' = \{R_1, \dots, R_{i-1}, Q_1, \dots, Q_k, R_{i+1}, \dots, R_m\}$ posiada cechę bezstratnego połączenia w stosunku do F .

Zależności wielowartościowe

LOT	DZIEŃ	SAMOLOT	LINIA	START	CEL
LO107	poniedziałek	ATR 72	LOT	Poznań	Monachium
LO107	środa	ATR 72	LOT	Poznań	Monachium
LO107	piątek	ATR 72	LOT	Poznań	Monachium
LO107	poniedziałek	EMBRAER 135	LOT	Poznań	Monachium
LO107	środa	EMBRAER 135	LOT	Poznań	Monachium
LO107	piątek	EMBRAER 135	LOT	Poznań	Monachium
UA112	wtorek	BOEING 737	United	Nowy Jork	Warszawa
UA112	czwartek	BOEING 737	United	Nowy Jork	Warszawa
UA112	wtorek	BOEING 747	United	Nowy Jork	Warszawa
UA112	czwartek	BOEING 747	United	Nowy Jork	Warszawa

United lata z Nowego Jorku do Warszawy we wtorki i czwartki, linię obsługują Boeingi 737 i 747

Nowak mówi po angielsku i niemiecku, potrafi programować w PHP i C#

NAZWISKO	JĘZYK	PROGR.
Kowalski	angielski	C++
Kowalski	angielski	Java
Kowalski	francuski	C++
Kowalski	francuski	Java
Nowak	angielski	C#
Nowak	niemiecki	PHP
Nowak	angielski	PHP
Nowak	niemiecki	C#

Zależności wielowartościowe

LOT	DZIEŃ	SAMOLOT	LINIA	START	CEL
LO107	poniedziałek	ATR 72	LOT	Poznań	Monachium
LO107	środa	ATR 72	LOT	Poznań	Monachium
LO107	piątek	ATR 72	LOT	Poznań	Monachium
LO107	poniedziałek	EMBRAER 135	LOT	Poznań	Monachium
LO107	środa	EMBRAER 135	LOT	Poznań	Monachium
LO107	piątek	EMBRAER 135	LOT	Poznań	Monachium
UA112	wtorek	BOEING 737	United	Nowy Jork	Warszawa
UA112	czwartek	BOEING 737	United	Nowy Jork	Warszawa
UA112	wtorek	BOEING 747	United	Nowy Jork	Warszawa
UA112	czwartek	BOEING 747	United	Nowy Jork	Warszawa
UA112	sobota	AIRBUS A340	United	Nowy Jork	Warszawa
UA112	wtorek	AIRBUS A340	United	Nowy Jork	Warszawa
UA112	czwartek	AIRBUS A340	United	Nowy Jork	Warszawa
UA112	sobota	BOEING 737	United	Nowy Jork	Warszawa
UA112	sobota	BOEING 747	United	Nowy Jork	Warszawa

United wprowadza nowe połączenie
w soboty i dodaje do obsługi linii
nowy typ samolotu

Zależności wielowartościowe

- Zależność wielowartościowa jest konsekwencją pierwszej postaci normalnej która nie dopuszcza atrybutów wielowartościowych.
- Wystąpienie zależności wielowartościowej $X \twoheadrightarrow Y$ w schemacie relacji $R(XYZ)$ wyraża
 - związek między zbiorami atrybutów X i Y ,
 - związek między zbiorami atrybutów X i Z ,
 - niezależność zbiorów atrybutów Y i Z .

NAZWISKO	JĘZYK	PROGR.
Kowalski	angielski	C++
Kowalski	angielski	Java
Kowalski	francuski	C++
Kowalski	francuski	Java
...

Zależność wielowartościowa

definicja

Dana jest relacja R oraz rozłączne parami zbiory atrybutów X, Y, Z , gdzie $Z = R - \{X \cup Y\}$.

Relacja R **spełnia zależność wielowartościową** $X \twoheadrightarrow Y$, jeżeli dla dowolnych dwóch krotek t_1 i t_2 z każdego stanu relacji $r(R)$ takich, że $t_1[X] = t_2[X]$, istnieją w $r(R)$ krotki t_3 i t_4 takie, że:

- $t_1[X] = t_2[X] = t_3[X] = t_4[X]$
- $t_1[Y] = t_3[Y]$ i $t_2[Y] = t_4[Y]$
- $t_2[Z] = t_3[Z]$ i $t_1[Z] = t_4[Z]$

Zależność wielowartościowa $X \twoheadrightarrow Y$ jest **trywialna**, jeśli

- $X \supseteq Y$, lub
- $X \cup Y = R$.

- krotki t_1, t_2, t_3, t_4 nie muszą być różne
- jeśli $X \twoheadrightarrow Y$, to również $X \twoheadrightarrow Z$, co oznaczamy $X \twoheadrightarrow Y|Z$
- trywialna zależność wielowartościowa $X \twoheadrightarrow Y$ jest spełniona dla dowolnego stanu relacji $r(R)$

Reguły wywodzenia wielowartościowych zależności funkcyjnych ³⁴

- Uniwersalne reguły wywodzenia zależności
 - IR₁ (reguła zwrotna): jeśli $X \supseteq Y$ to $X \rightarrow Y$
 - IR₂ (reguła zwiększania): $X \rightarrow Y \models XZ \rightarrow YZ$
 - IR₃ (reguła przechodnia): $X \rightarrow Y, Y \rightarrow Z \models X \rightarrow Z$
 - IR₄ (reguła uzupełnienia): $X \rightarrow\rightarrow Y \models X \rightarrow\rightarrow R - \{X \cup Y\}$
 - IR₅ (reguła zwiększania): jeśli $X \rightarrow\rightarrow Y$ i $W \supseteq Z$ to $WX \rightarrow\rightarrow YZ$
 - IR₆ (reguła przechodnia): $X \rightarrow\rightarrow Y, Y \rightarrow\rightarrow Z \models X \rightarrow\rightarrow Z - Y$
 - IR₇ (reguła replikacji): jeśli $X \rightarrow\rightarrow Y$ i istnieje W takie że:
 $W \cap Y = \emptyset, W \rightarrow Z, Y \supseteq Z,$
to wówczas $X \rightarrow Z$

Czwarta postać normalna (4NF)

Relacja R jest w czwartej postaci normalnej względem zbioru zależności wielowartościowych F, jeśli R jest w trzeciej postaci normalnej i dla każdej zależności wielowartościowej $X \twoheadrightarrow Y$ spełniony jest jeden z warunków:

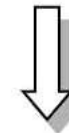
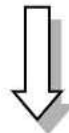
- X jest nadkluczem schematu relacji R, lub
- $X \twoheadrightarrow Y$ jest trywialna.

PRA COWNIK	PODWŁADNY	PROJEKT
Anielski	Lis	Manhattan
Anielski	Wilk	Manhattan
Anielski	Lis	Soho
Anielski	Wilk	Soho
Diabelski	Łoś	Tribeca
Diabelski	Łoś	Brooklyn
Diabelski	Jeleń	Tribeca
Diabelski	Jeleń	Brooklyn

PODWŁADNY \twoheadrightarrow PROJEKT

Transformacja do 4NF

PRA COWNIK	PODWŁADNY	PROJEKT
Anielski	Lis	Manhattan
Anielski	Wilk	Manhattan
Anielski	Lis	Soho
Anielski	Wilk	Soho
Diabelski	Łoś	Tribeca
Diabelski	Łoś	Brooklyn
Diabelski	Jeleń	Tribeca
Diabelski	Jeleń	Brooklyn



PRA COWNIK	PODWŁADNY
Anielski	Lis
Anielski	Wilk
Diabelski	Łoś
Diabelski	Jeleń

PRA COWNIK	PROJEKT
Anielski	Manhattan
Anielski	Soho
Diabelski	Tribeca
Diabelski	Brooklyn

Zależność połączeniowa

Za każdym razem gdy dealer d dostarcza auto a , i auto a występuje w typie t , i dealer d dostarcza jakiegokolwiek auto w typie t , to dealer d dostarcza auto a w typie t

PRZEDSTAWICIELE

DEALER	AUTO	TYP
Krotoski-Cichy	Audi	sedan
Krotoski-Cichy	Volkswagen	combi
Krotoski-Cichy	Volkswagen	sedan
Fietz	Toyota	sedan
Fietz	Toyota	combi
Pieluszyńska	Renault	sedan

R1

DEALER	AUTO
Krotoski-Cichy	Audi
Krotoski-Cichy	Volkswagen
Fietz	Toyota
Pieluszyńska	Renault

R2

AUTO	TYP
Audi	sedan
Volkswagen	sedan
Volkswagen	combi
Toyota	sedan
Toyota	combi
Renault	sedan

R3

DEALER	TYP
Krotoski-Cichy	sedan
Krotoski-Cichy	combi
Fietz	sedan
Fietz	combi
Pieluszyńska	sedan

Piąta postać normalna (5NF)

Niech $R = \{R_1, \dots, R_n\}$ oznacza zbiór schematów relacji zdefiniowanych nad zbiorem atrybutów $U = \{A_1, \dots, A_m\}$, takich, że $R_1 \cup \dots \cup R_n = U$.
Relacja $r(R)$ spełnia **zależność połączeniową** (ang. *join dependency*) $JD[R_1, \dots, R_n]$ jeśli można ją zdekomponować bez utraty informacji na podrelacje $r_1(R_1), \dots, r_n(R_n)$. Zachodzi wówczas:
 $* (r_1(R_1), \dots, r_n(R_n)) = r(R)$. Zależność $JD[R_1, \dots, R_n]$ jest **trywialna** jeśli którykolwiek ze schematów R_i jest równy R .

Relacja R jest w **piątej postaci normalnej** jeśli dla każdej zależności połączeniowej $JD[R_1, \dots, R_n]$ w schemacie R zachodzi:

- $JD[R_1, \dots, R_n]$ jest trywialna, i
- każdy schemat R_1, \dots, R_n jest nadkluczem schematu R .