

Bazy danych



Andrzej Łachwa, UJ, 2015

andrzej.lachwa@uj.edu.pl

8/14

Normalizacja

Proces prowadzi do kolejnych, coraz lepszych poziomów normalizacji zwanych postaciami normalnymi: pierwszą, drugą, trzecią, Boyce'a-Codda, czwartą i piątą (1NF, 2NF, 3NF, BCFN, ...).

Proces normalizacji realizujemy przez dekompozycję schematów relacji. Proces ten musi brać pod uwagę właściwość złączenia bezstratnego i właściwość zachowania zależności. Ta druga może być niekiedy pomijana.

Atrybut relacji nazywamy **podstawowym** (prymarnym), gdy należy do któregośkolwiek klucza kandydującego. Gdy nie należy do żadnego klucza, to nazywamy go **nieprymarnym**.

1NF

Każdy schemat relacji musi posiadać klucz główny.

Dziedzina każdego atrybutu musi zawierać wyłącznie wartości niepodzielne (atomowe), a wartość każdego atrybutu krotki musi być pojedynczą wartością z dziedziny.

W takiej sytuacji każdy atrybut jest funkcyjnie zależny od klucza.

BŁĄD 1

Macierze lub listy są ukryte w tabeli poprzez spłaszczenie faktycznej struktury.

BŁĄD 2

Używanie kilku kolumn, które mają tę samą wartość semantyczną.

BŁĄD 3

Lista wartości wpisana do łańcucha znaków.

2NF

Schemat jest w 1NF i żaden nieprimarny atrybut nie jest częściowo zależny od dowolnego klucza.

Mówimy krótko, że w tej postaci nie występują zależności częściowe.

Rozkład schematu do 2NF polega na pozbyciu się zależności częściowych.

3NF

Zależność $X \rightarrow A$ jest nazywana nietrywialną, gdy A nie jest elementem zbioru X .

Schemat jest w 2NF i dla każdej nietrywialnej zależności $X \rightarrow A$ albo X jest nadkluczem albo A jest prymarny.

Innymi słowy, żaden nieprymarny atrybut nie jest przechodnio zależny od klucza. Musi zależeć od klucza, całego klucza i tylko od klucza!

Rozkład schematu do 3NF polega na pozbyciu się zależności przechodnich.

BCNF

Jest to silniejsza wersja 3NF. Wymaga się, by dla każdej nietrywialnej zależności $X \rightarrow A$ zbiór atrybutów X był nadkluczem.

Przeczytaj rozdział 10 [Ramez Elmasri, Shamkant B. Navathe]

Zob. wykład o normalizacji na <http://wazniak.mimuw.edu.pl/>

NrPra	NrPro	Data	Godz	Imię	Nazwisko	Termin
128	54a	12.11	7	Jan	Adamski	luty 06
128	54a	13.11	8	Jan	Adamski	luty 06
128	54a	14.11	8	Jan	Adamski	luty 06
128	54a	15.11	6	Jan	Adamski	luty 06
128	54a	16.11	8	Jan	Adamski	luty 06
128	54a	17.11	6	Jan	Adamski	luty 06
128	54a	18.11	0	Jan	Adamski	luty 06
128	54b	19.11	4	Jan	Adamski	maj 06
128	54b	20.11	4	Jan	Adamski	maj 06
128	54b	21.11	10	Jan	Adamski	maj 06
128	54b	22.11	12	Jan	Adamski	maj 06
129	54a	12.11	8	Adam	Janowski	luty 06

Rozkład schematu na podschematy:

Praca (NrPracownika, NrProjektu, Data, IleGodzin, Imię, Nazwisko, Termin)



KartaPracy (NrPracownika, NrProjektu, Data, IleGodzin)

Pracownik (NrPracownika, Imię, Nazwisko)

Projekt (NrProjektu, Termin)

NrPra	NrPro	Data	Godz
128	54a	12.11	7
128	54a	13.11	8
128	54a	14.11	8
128	54a	15.11	6
128	54a	16.11	8
128	54a	17.11	6
128	54a	18.11	0
128	54b	19.11	4
128	54b	20.11	4
128	54b	21.11	10
128	54b	22.11	12
129	54a	12.11	8

NrPra	Imię	Nazwisko
128	Jan	Adamski
129	Adam	Janowski

NrPro	Termin
54a	02.06
54b	05.06.

2NF

NrUcznia*	NrKlasy	Wychowawca
118	4a	Anna Kowalska
119	4a	Anna Kowalska
120	4a	Anna Kowalska
121	4a	Anna Kowalska
122	4a	Anna Kowalska
123	4a	Anna Kowalska
124	4a	Anna Kowalska
125	4b	Marek Nowak
126	4b	Marek Nowak
127	4b	Marek Nowak
128	4b	Marek Nowak
129	4b	Marek Nowak

Rozkład schematu na podschematy:

Uczniowie (NrUcznia, NrKlasy, Wychowawca)



Uczniowie (NrUcznia, NrKlasy)

Klasy (NrKlasy, Wychowca)

NrUcznia*	NrKlasy
118	4a
119	4a
120	4a
121	4a
122	4a
123	4a
124	4a
125	4b
126	4b
127	4b
128	4b
129	4b

NrKlasy	Wychowawca
4a	Anna Kowalska
4b	Marek Nowak

3NF

Uczeń*	Język*	Nauczyciel
28	angielski	John Kowalsky
28	niemiecki	Karl Werner
28	rosyjski	Nina Woronin
29	rosyjski	Nina Woronin
30	rosyjski	Nina Woronin
30	niemiecki	Karl Werner
31	angielski	John Kowalsky
31	niemiecki	Karl Werner
32	niemiecki	Berta Werner
32	rosyjski	Nina Woronin
33	rosyjski	Nina Woronin
33	niemiecki	Karl Werner
33	angielski	John Kowalsky

{Uczeń*, Język*}
→ Nauczyciel

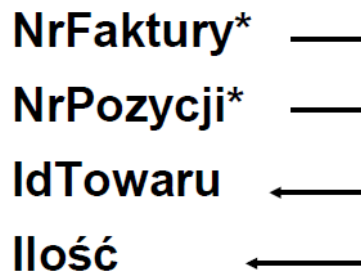
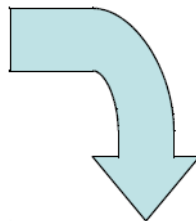
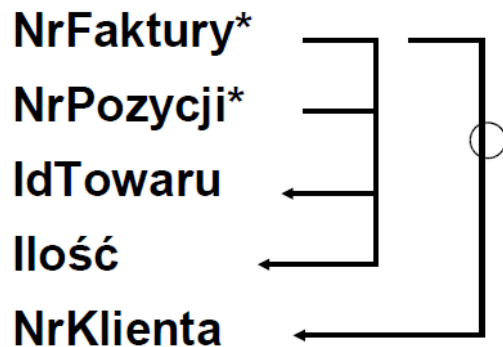
{Uczeń*,
Nauczyciel*} →
Język

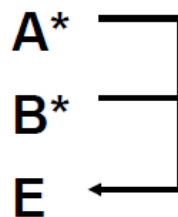
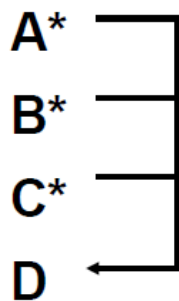
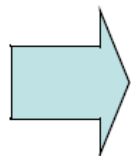
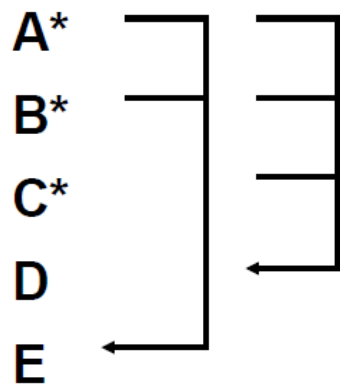
Nauczyciel →
Język

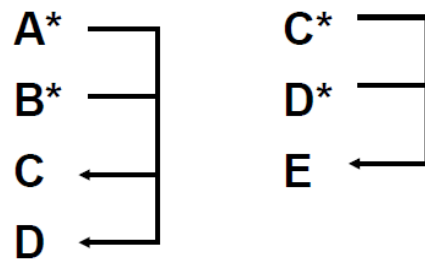
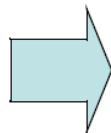
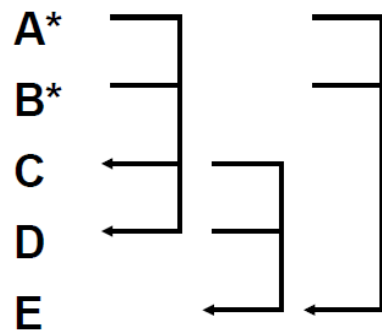
Uczeń*	Nauczyciel*
28	John Kowalsky
28	Karl Werner
28	Nina Woronin
29	Nina Woronin
30	Nina Woronin
30	Karl Werner
31	John Kowalsky
31	Karl Werner
32	Berta Werner
32	Nina Woronin
33	Nina Woronin
33	Karl Werner
33	John Kowalsky

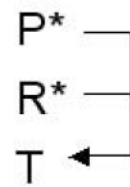
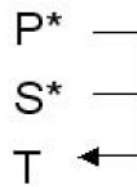
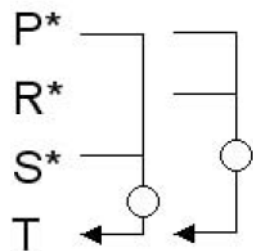
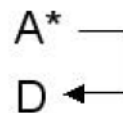
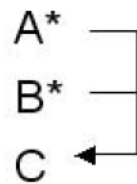
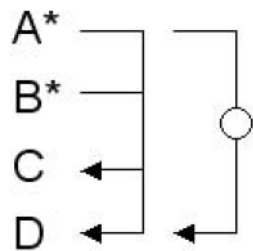
Język	Nauczyciel*
angielski	John Kowalsky
niemiecki	Karl Werner
rosyjski	Nina Woronin
niemiecki	Berta Werner

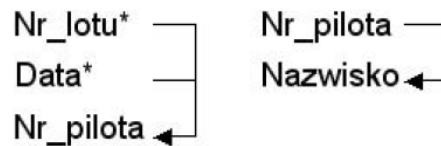
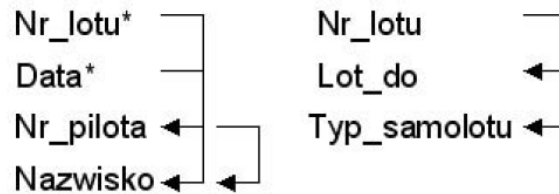
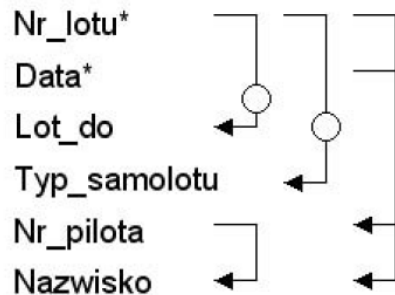
3½NF













Uwagi: kino może mieć kilka sal;
 różne filmy mogą być wyświetlane na różnych seansach;
 $\{T, M\}$, $\{K, T\}$ – klucze kandydujące;
 $K \rightarrow M$ i $K \not\subseteq KT$;
 BCNF gubi zależność $TM \rightarrow K$

Przykład

$\{\text{UCZEŃ}^*, \text{JĘZYK}^*, \text{NAUCZYCIEL}\}$

$\{\text{UCZEŃ}, \text{JĘZYK}\} \rightarrow \text{NAUCZYCIEL}$

$\text{NAUCZYCIEL} \rightarrow \text{JĘZYK}$

Są możliwe trzy rozkłady:

$\{\text{UCZEŃ}^*, \text{NAUCZYCIEL}^*\}, \{\text{UCZEŃ}^*, \text{JĘZYK}\}$

$\{\text{JĘZYK}, \text{NAUCZYCIEL}^*\}, \{\text{JĘZYK}^*, \text{UCZEŃ}^*\}$

$\{\text{NAUCZYCIEL}^*, \text{JĘZYK}\}, \{\text{NAUCZYCIEL}^*, \text{UCZEŃ}^*\}$

Tylko trzeci rozkład, po złączeniu, nie będzie generował fałszywych krotek!

Tabele pomocnicze

Tabele pomocnicze nie są częścią modelu danych, więc nie powinny pojawiać się na etapie projektowania koncepcyjnego.

Wyróżnimy dwa rodzaje takich tabel: tabele słownikowe i tabele funkcyjne. Rozważmy dwa przypadki tabel słownikowych: listę nazw województw oraz listę nazw państw świata. Obie listy powinny być używane zawsze wtedy, gdy użytkownik bazy danych będzie mógł wprowadzać te nazwy jako dane. Wprowadzanie to powinno być zawsze (w tego rodzaju przypadkach) ograniczone do wyboru elementu z listy!...

Sprawa nie jest jednak prosta (por. alfabetyczny wykaz krajów...).

Tabela funkcyjna jest sposobem zdefiniowania funkcji. Tabela ta jest wykorzystywana do wykonywania zapytań.

Często tworzoną tabelą funkcyjną jest kalendarz. Jest tak m. in. dlatego, że daty świąt są bardzo nieregularne (np. wyznaczanie świąt ruchomych, takich jak Wielkanoc czy Ramadan, przy użyciu funkcji byłoby bardzo trudne). Wielkość kalendarza na 20 lat to tylko około 7000 wierszy, czyli bardzo mała tabela. Tabelę taką zapełniamy danymi przygotowanymi w arkuszu kalkulacyjnych bądź pobranymi z Internetu i zabraniamy modyfikacji tych danych użytkownikom bazy (z wyłączeniem administratora danych, bo przecież mogą następować zmiany ustawowe regulujące liczbę i terminy świąt, terminy wakacji szkolnych itp.).

Por. kalendarz świąt ...