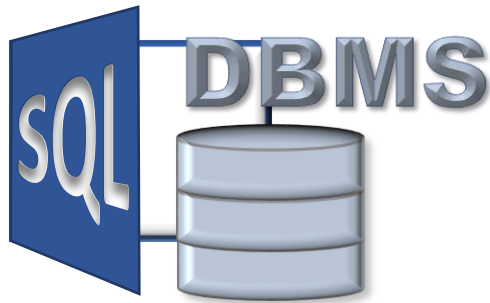


Bazy danych



Relacyjny model danych
zależności funkcjonalne
I, II, III forma normalna relacji
analiza przykładu

Zadanie: Należy zaprojektować system informatyczny (bazę danych) dla potrzeb placówki medycznej, który umożliwi gromadzenie i przetwarzanie danych skierowań na badania wystawianych przez lekarzy rodzinnych swoim pacjentom.

Przeznaczenie: Aplikacja przetwarza dane skierowań na badania.

Funkcje:

- Przechowuje dane osobowe pacjentów.
- Przechowuje dane lekarzy.
- Przechowuje dane skierowań na badania.
- Tworzy statystyki (np. skierowania pacjenta w określonym przedziale czasowym).

Schemat opisowy: Lekarz jest identyfikowany przez numer prawa wykonywania zawodu, ma przypisane imię i nazwisko. Pacjent jest identyfikowany przez numer PESEL, ma imię i nazwisko, jest przypisany do jednego lekarza POZ. Każde badanie ma unikalny kod (ICD-9), nazwę kategorii oraz cenę. Skierowanie zawiera informacje o dacie wystawienia, kodzie badania oraz identyfikator pacjenta i lekarza kierującego.

Modelowanie danych – proces tworzenia odpowiedniego odwzorowania danych dla potrzeb systemu informatycznego, polega na przejściu od swobodnego opisu fragmentu rzeczywistości (modelowanego systemu) do jego formalnej reprezentacji.

Elementy świata rzeczywistego

- ❑ **obiekt** – składnik rzeczywistego systemu postrzegany jako istotny przez jednostkę lub grupę, przyszłych użytkowników bazy danych,
- ❑ **powiązanie** – opis stanu, w którym znalazły się co najmniej dwa obiekty.

Obiekty i powiązania mogą być dodatkowo opisane za pomocą atrybutów.

Przykład

- **obiekt**: pacjent, **atrybuty**: PESEL, nazwisko, imię,
- **obiekt**: lekarz, **atrybuty**: PWZ (nr prawa wykonywania zawodu), nazwisko, imię,
- **obiekt**: badanie, **atrybuty**: ICD9 (kod badania), nazwa kategorii, cena,
- **powiązanie**: pacjent-lekarz,
- **powiązanie**: pacjent-lekarz-badanie, **atrybut**: data wystawienia.

Model relacyjny – definicje (1)

Relacją R na zbiorach D_1, D_2, \dots, D_n nazywamy dowolny podzbiór iloczynu kartezyjskiego tych zbiorów i zapisujemy:

$$R(D_1, D_2, \dots, D_n), R \subset D_1 \times D_2 \times \dots \times D_n.$$

Zapis postaci $R(D_1, D_2, \dots, D_n)$ jest **schematem relacji**, R nazwą relacji, a elementy D_1, D_2, \dots, D_n **atrybutami** lub **składnikami relacji**.

Krotka relacji – ciąg wartości atrybutów danego schematu relacji.

Alternatywna definicja relacji

Relacją R na zbiorach D_1, D_2, \dots, D_n nazywamy dowolny zbiór krotek postaci:

$$\langle d_1, d_2, \dots, d_n \rangle$$

takich, że: $d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$.

Przykład

Schemat relacji *Pacjent*:

Pacjent(PESEL, Nazwisko, Imię, Adres)

Krotka relacji *Pacjent*:

\langle 70052000234, Kowalski, Jan, 65-516 Zielona Góra Podgórna 50 \rangle

Model relacyjny – definicje (2)

Identyfikator relacji – atrybut lub ciąg atrybutów, których wartości określają w sposób jednoznaczny krotkę relacji.

Identyfikator kluczowy (klucz) relacji – jeden, dowolnie wybrany identyfikator relacji (zazwyczaj kryterium wyboru jest długość). Klucz w schemacie relacji jest zaznaczany przez podkreślenie odpowiednich atrybutów.

Klucz naturalny – klucz złożony z atrybutów relacji, których obecność wynika z przeprowadzonej analizy problemu.

Klucz sztuczny – sztucznie wprowadzony atrybut relacji, którego wartości gwarantują jednoznaczną identyfikację krotek. W roli klucza sztucznego najczęściej występuje liczba porządkowa.

Przykład

Klucz naturalny: *Pacjent(PESEL, Nazwisko, Imię, Adres)*

Klucz sztuczny: *Pacjent(IDP, PESEL, Nazwisko, Imię, Adres)*

IDP – identyfikator pacjenta, unikalna wartość liczbowa.

Przykład – rozwiązanie trywialne

Schemat relacji zaproponowany na podstawie analizy zadania:

Skierowanie(PESEL, NazwiskoP, ImięP, PWZ, NazwiskoL, ImięL, ICD9, Kategoria, Cena, Data)

Atrybuty:

- *PESEL* – pesel pacjenta
- *NazwiskoP* – nazwisko pacjenta,
- *ImięP* – imię pacjenta,
- *AdresP* – adres pacjenta
- *PWZ* – numer prawa wykonywania zawodu lekarza,
- *NazwiskoL* – nazwisko lekarza,
- *ImięL* – imię lekarza,
- *ICD9* – kod badania,
- *Kategoria* – nazwa kategorii badania,
- *Cena* – cena badania
- *Data* – data wystawienia skierowania.

Uwaga: kompletny schemat relacji wymaga określenia klucza, odpowiednia analiza zostanie przedstawiona na kolejnych slajdach.

Przykładowy zestaw danych

Przykładowy zbiór krotek relacji *Skierowanie*

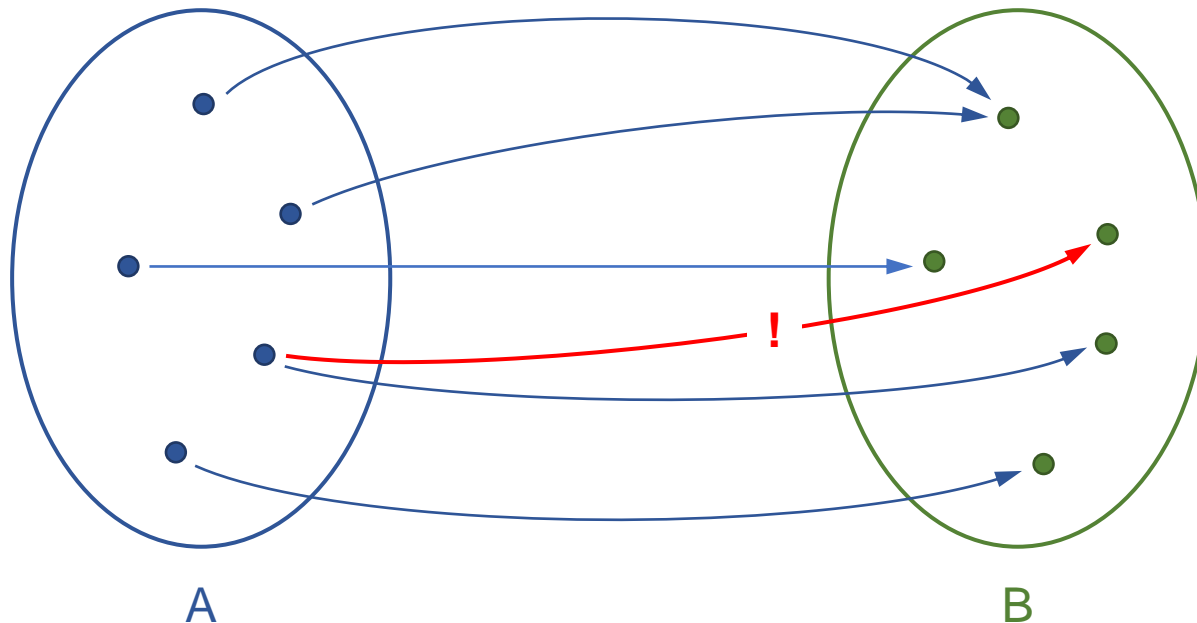
<i>PESEL</i>	<i>NazwiskoP</i>	<i>ImięP</i>	<i>PWZ</i>	<i>NazwiskoL</i>	<i>ImięL</i>	<i>ICD9</i>	<i>Kategoria</i>	<i>Cena</i>	<i>Data</i>
70052000234	Nowak	Andrzej	1234567	Adamska	Alicja	A14	Leukocyty	25	15.01.21
70052000234	Nowak	Andrzej	1234567	Adamska	Alicja	A15	Glukoza	40	15.01.21
82012090803	Kowalski	Jan	7654321	Wójcik	Zofia	A14	Leukocyty	25	18.01.21
82012090803	Kowalski	Jan	7654321	Wójcik	Zofia	C09	Erytrogram	37	18.01.21
82012090803	Kowalski	Jan	7654321	Wójcik	Zofia	C19	Hemoglobina	28	18.01.21
70052000234	Nowak	Andrzej	1234567	Adamska	Alicja	C09	Erytrogram	37	18.01.21
82012090803	Kowalski	Jan	1234567	Adamska	Alicja	A14	Leukocyty	25	11.02.21
70052000234	Nowak	Andrzej	1234567	Adamska	Alicja	C41	Limfocyty T	45	25.02.21

Uwaga: Skierowanie na każde badanie przechowywane jest jako osobna krotka relacji. Umieszczenie kilku kodów badań w tym samym polu (np. „po przecinku”) jest fundamentalnym błędem modelu i uniemożliwi efektywne przetwarzanie danych w docelowej aplikacji.

1. Anomalie przy wstawianiu – w bazie nie można umieścić informacji o pacjentach, którym nie wystawiono żadnego skierowania (pola ICD9, Kategoria, Data nie mogą być wypełnione).
2. Redundancja – wiele informacji powtarza się: nazwiska i imiona pacjentów (każdy z pacjentów czterokrotnie), dane lekarzy (Alicja Adamska pięciokrotnie, Zofia Wójcik trzykrotnie), opis badania (A14 trzykrotnie, C09 dwukrotnie).
3. Niespójność danych – wielokrotne powtarzanie informacji prowadzi do sprzecznych informacji: do 18.01 lekarzem Jana Kowalskiego była Zofia Wójcik, po tym dniu pacjent zmienił lekarza i na skierowaniu z 11.02 występuje Alicja Adamska (ustalenie lekarza rodzinnego pacjenta wymaga analizy zawartości całej bazy danych).
4. Anomalie przy aktualizacji – zmiana danych osobowych (np. zmiana nazwiska) wymaga aktualizacji wszystkich skierowań związanych z daną osobą (bez tej operacji wystąpi niespójność danych).
5. Anomalie przy usuwaniu – informacje o lekarzach nie są odseparowane od skierowań, nie można usunąć danych lekarza, który zmienił miejsce pracy, nie można ustalić listy lekarzy zatrudnionych w placówce.

Zależność funkcjonalna

Składnik B jest **funkcjonalnie zależny** od składnika A w relacji $R(A, B, C, D)$ jeżeli każdej wartości $a \in A$ jest przyporządkowana tylko jedna wartość $b \in B$. Zależność funkcjonalną B w stosunku do A zapisujemy $A \rightarrow B$.



Uwaga: B jest zależny funkcjonalnie od A jeżeli występują wyłącznie zależności reprezentowane przez niebieskie strzałki. Jeżeli występuje zależność przedstawiona czerwoną strzałką jednemu elementowi z A odpowiada kilka elementów z B, więc składnik B nie jest zależny funkcjonalnie od A.

Zależność funkcjonalna – przykłady

Skierowanie(*PESEL, NazwiskoP, ImięP, PWZ, NazwiskoL, ImięL, ICD9, Kategoria, Data*)

PESEL → *NazwiskoP*

Jest zależnością funkcjonalną:

Każda osoba ma przypisany unikalny numer *PESEL*, więc każdemu numerowi *PESEL* odpowiada dokładnie jedno *Nazwisko*. Zależność odwrotna nie będzie zależnością funkcjonalną, ponieważ w zbiorze może wystąpić kilka osób o tym samym *Nazwisku*, a w takim przypadku jednemu *Nazwisku* będzie odpowiadać kilka różnych numerów *PESEL*.

PESEL → *ICD9*

Nie jest zależnością funkcjonalną:

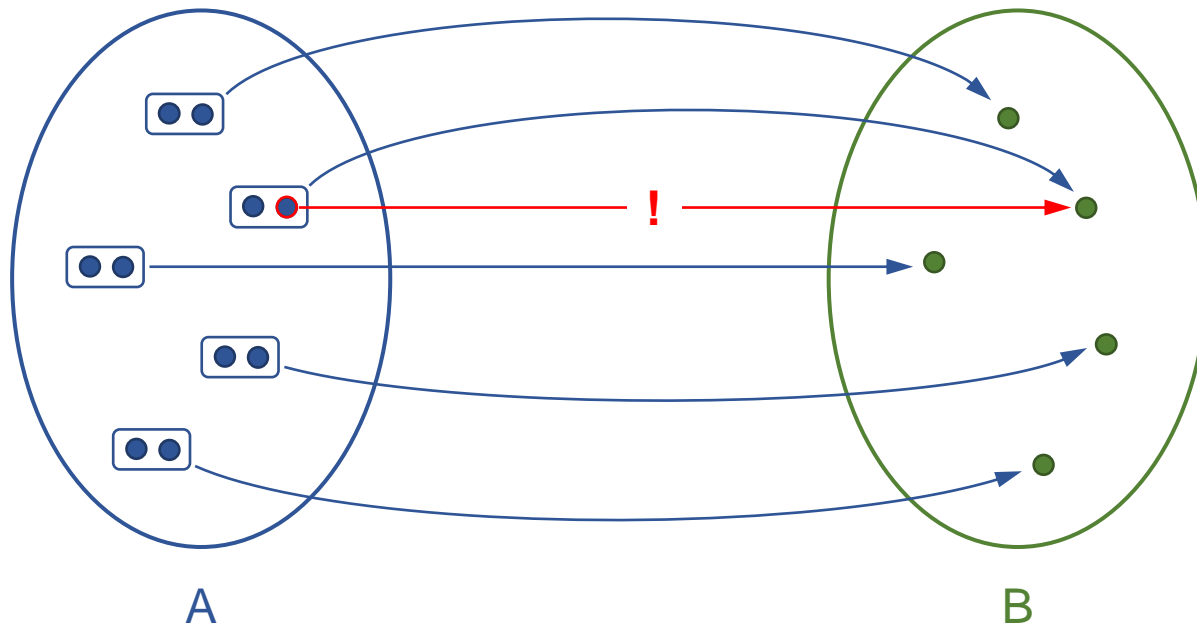
Jedna osoba może mieć kilka zleceń na różne badania (każde z nich ma inny kod), więc jednemu numerowi *PESEL* może odpowiadać wiele kodów *ICD9*.

Uwaga: analizując zależność funkcjonalną należy rozpatrywać wszystkie możliwości, uwzględniając również przypadki które nie występują w aktualnym zbiorze danych, ale mogą pojawić się w przyszłości.

Zależność funkcjonalna elementarna

Składnik B jest w **zależności funkcjonalnej elementarnej** od składnika A w relacji $R(A, B, C, D)$ jeżeli jest funkcjonalnie zależny od całości A i nie jest funkcjonalnie zależny od części A.

Uwaga: Rozpatrywanie zależności funkcjonalnej elementarnej ma sens gdy składnik A można podzielić na mniejsze części. Jeżeli A stanowi jedną, niepodzielną całość każda zależność funkcjonalna jest jednocześnie elementarna.



Zależność funkcjonalna elementarna – przykłady

Skierowanie(*PESEL*, *NazwiskoP*, *ImięP*, *PWZ*, *NazwiskoL*, *ImięL*, *ICD9*, *Kategoria*, *Data*)

PESEL, *Data*, *ICD9* → *PWZ*

Jest zależnością funkcjonalną elementarną:

PESEL pacjenta z *Datą* skierowania i kodem badania jednoznacznie identyfikują lekarza, który wystawił skierowanie (założenie: w tym samym dniu jeden pacjent nie dostaje dwóch skierowań na to samo badanie). *PWZ* nie jest funkcjonalnie zależny od numeru *PESEL* (wielu lekarzy może wystawiać skierowanie pacjentowi), *Daty* (tego samego dnia skierowanie może wystawić wielu lekarzy) lub kodu *ICD9* (wielu lekarzy wystawia skierowanie na to samo badanie), nie ma również zależności *PWZ* od żadnej z par składników (*PESEL-Data*, *PESEL-ICD9*, *Data-ICD9*).

PESEL, *Data* → *NazwiskoP*

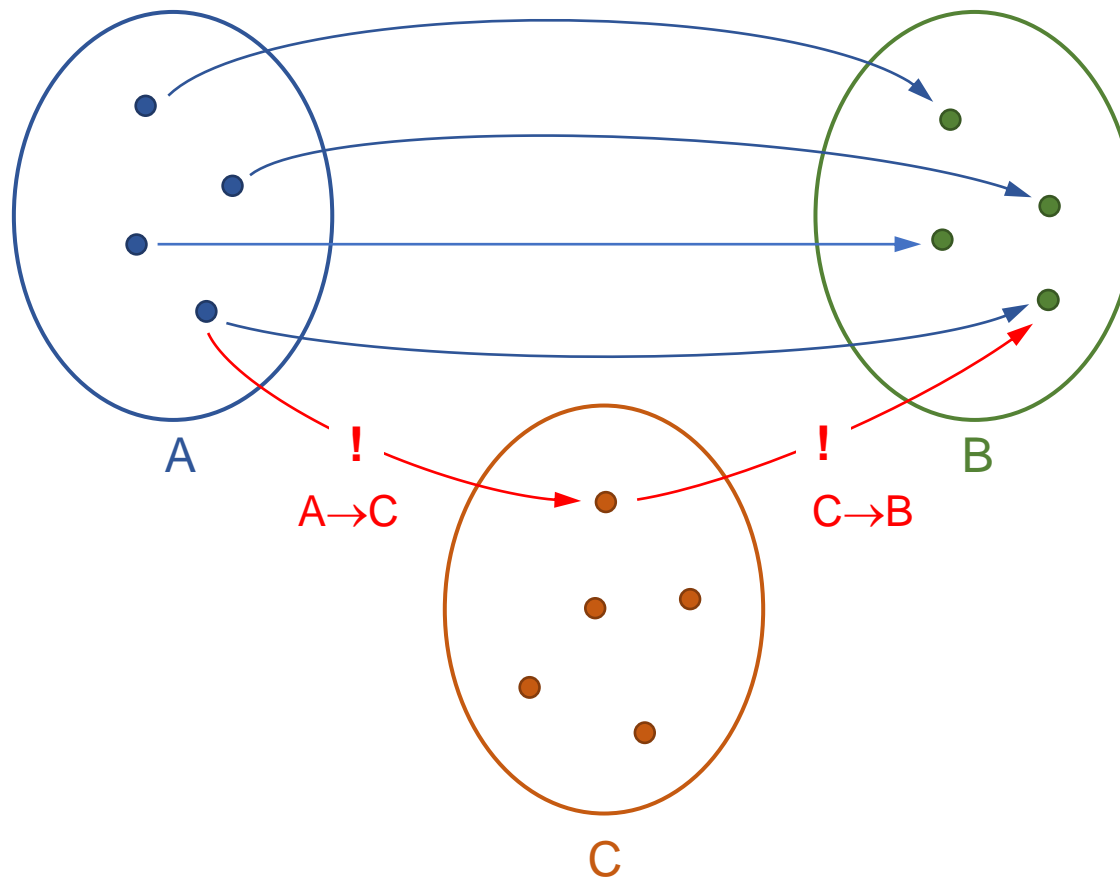
Nie jest zależnością funkcjonalną elementarną.

Nazwisko pacjenta zależy funkcjonalnie od całości, ale jest zależne również od samego numeru *PESEL*, czyli jest zależne od części (patrz przykład na s. 10).

Uwaga: Jeżeli $A \rightarrow B$, to zachodzi $A, C \rightarrow B$ dla dowolnego składnika C w relacji R .

Zależność funkcjonalna bezpośrednia

Składnik B jest w **zależności funkcjonalnej bezpośredniej** od składnika A w relacji $R(A, B, C, D)$ jeżeli jest funkcjonalnie zależny od A i nie istnieje taki składnik C dla którego zachodzi: $A \rightarrow C$ i $C \rightarrow B$.



Zależność funkcjonalna bezpośrednia

Skierowanie(*PESEL*, *NazwiskoP*, *ImięP*, *PWZ*, *NazwiskoL*, *ImięL*, *ICD9*, *Kategoria*, *Data*)

PESEL → *ImięP*

Jest zależnością funkcjonalną bezpośrednią.

Poza numerem *PESEL* w relacji nie istnieje składnik, od którego *Imię* jest zależne funkcjonalnie (*Nazwisko* może powtórzyć się u kilku pacjentów, pozostałe składniki nie mają bezpośredniego związku z *Imieniem* pacjenta), nie można więc wskazać składnika, poprzez który uzyskuje się zależność pośrednią.

PESEL, *Data*, *ICD9* → *NazwiskoL*

Nie jest zależnością bezpośrednią.

Nazwisko lekarza kierującego na badanie jest zależne funkcjonalnie od numeru *PESEL*, *Daty* skierowania i kodu *ICD9*. Jednocześnie ta trójka atrybutów wyznacza jednoznacznie *PWZ* lekarza (*PWZ* jest od nich zależne funkcjonalnie, patrz s.2-9), a *PWZ* określa *Nazwisko* lekarza (każdy lekarz ma unikalny *PWZ*), stąd zachodzi:

PESEL, *Data*, *ICD9* → *PWZ*, *PWZ* → *NazwiskoL*,

istnieje więc składnik (*PWZ*), poprzez który uzyskuje się zależność pośrednią.

I forma normalna relacji

Relacja R jest w **pierwszej formie normalnej (IFN)** jeżeli każdy ze składników, który nie jest elementem klucza, jest w zależności funkcjonalnej od klucza.

Uwaga: Spełnienie warunku IFN gwarantuje prawidłowy wybór klucza relacji.

Graf zależności funkcjonalnych – graf skierowany, którego wierzchołkami są atrybuty (składniki) relacji, a krawędzie skierowane wyznaczają zależności funkcjonalne.

Konstrukcja grafu zależności funkcjonalnych

- Zapisać wszystkie atrybuty relacji,
- Wyznaczyć zależności funkcjonalne, których źródłem są pojedyncze składniki relacji i połączyć odpowiednie wierzchołki grafu krawędziami odpowiadającymi zależnościom,
- Jeżeli istnieją wierzchołki grafu niepołączone z pozostałymi wyznaczyć zależności, których źródłem są pary atrybutów i zaznaczyć je na grafie,
- Powtarzać operację do chwili wyznaczenia grupy atrybutów będących źródłem zależności dla wszystkich pozostałych, taka grupa jest kluczem relacji.

atrybut1 → atrybut2

*Pojedyncze źródło
zależności funkcjonalnej*

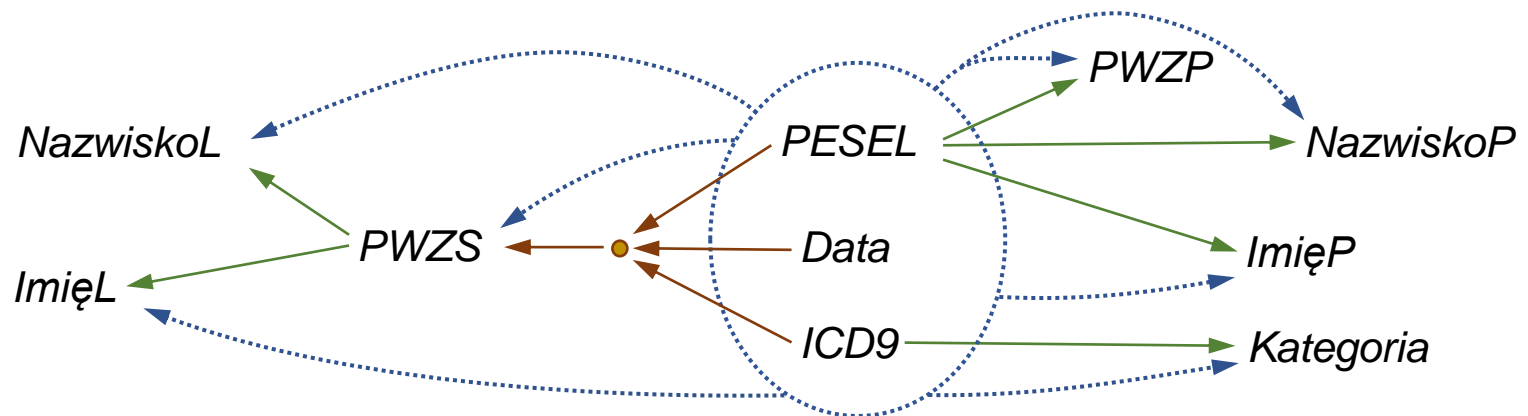
atrybut1 → ● → atrybut3
atrybut2 → ●

*Wielokrotne źródło
zależności funkcjonalnej*

Graf zależności funkcjonalnych – przykład

Skierowanie(PESEL, NazwiskoP, ImięP, PWZ, NazwiskoL, ImięL, ICD9, Kategoria, Data)

Uwaga: należy uwzględnić informację o lekarzach rodzinnych pacjentów stąd wprowadza się: PWZP – id. lekarza rodzinnego, PWZS – id. lekarza kierującego na badanie)



Zależności funkcjonalne

- $PESEL \rightarrow NazwiskoP, ImięP, PWZP, PWZ \rightarrow NazwiskoL, ImięL, ICD9 \rightarrow Kategoria,$
- $PESEL, Data, ICD9 \rightarrow PWZS,$
- $PESEL, Data, ICD9$ są źródłem zależności dla pozostałych atrybutów (założenie: w tym samym dniu ten sam pacjent nie dostaje dwóch skierowań na to samo badanie).

Relacja *Skierowanie* w IFN

Skierowanie(PESEL, *NazwiskoP*, *ImięP*, *PWZP*, *PWZS*, *NazwiskoL*, *ImięL*,
ICD9, *Kategoria*, Data)

PESEL określa jednoznacznie dane pacjenta, każde badanie ma unikalny kod, w konkretnym dniu dany pacjent jest przypisany do jednego lekarza rodzinnego i może mieć wystawione jedno skierowanie na konkretne badanie (możliwe kilka skierowań, ale na różne badania).

Stąd:

- *PESEL* → *NazwiskoP*, *ImięP*, *PWZP*, ponieważ *PESEL* jednoznacznie określa osobę,
- *ICD9* → *Kategoria*, ponieważ badanie ma unikalny kod,
- *PESEL*, *Data*, *ICD9* → *PWZS*, *NazwiskoL*, *ImięL*, ponieważ w danym pacjent może mieć wystawione jedno skierowanie na konkretne badanie,

Wszystkie składniki relacji *Skierowanie* zależą funkcjonalnie od atrybutów *PESEL*, *Data*, *ICD9* więc są one prawidłowym kluczem relacji. Relacja o takim schemacie jest w IFN.

II forma normalna relacji

Relacja R jest w **drugiej formie normalnej (IIFN)** jeżeli jest w IFN i każdy ze składników, który nie jest elementem klucza, jest w zależności funkcjonalnej elementarnej od klucza (zależy funkcjonalnie od klucza i nie jest zależny od jego części).

Uwaga: Sprawdzenie IIFN dotyczy wyłącznie relacji, których klucz składa się z kilku atrybutów. W przypadku relacji z kluczem jednoelementowym warunek zależności elementarnej wynika z istnienia zależności funkcjonalnej.

Transformacja relacji z IFN do IIFN

- Wyznaczenie wszystkich zależności elementarnych od składników klucza,
- Dekompozycja (podział) relacji pierwotnej poprzez wyłączenie atrybutów, które nie były zależne elementarnie od klucza,
- Utworzenie nowej relacji dla każdego źródła zależności elementarnej zawierającej fragment klucza relacji pierwotnej oraz wszystkie atrybuty, które były od niego zależne elementarnie.

Uwaga: Jedna z relacji powstających w wyniku dekompozycji powinna mieć klucz zgodny z relacją pierwotną. W szczególnym przypadku może być pozbawiona dodatkowych atrybutów, nie może jednak zanikać, ponieważ opisuje istotne z punktu widzenia modelu powiązanie pomiędzy rzeczywistymi obiektami.

II forma normalna relacji – przykład

Relacja *Skierowanie* w IFN

Skierowanie(PESEL, *NazwiskoP*, *ImięP*, *PWZP*, *PWZS*, *NazwiskoL*, *ImięL*,
ICD9, *Kategoria*, Data)

Źródła zależności elementarnych od składników klucza (patrz graf zależności s.16):

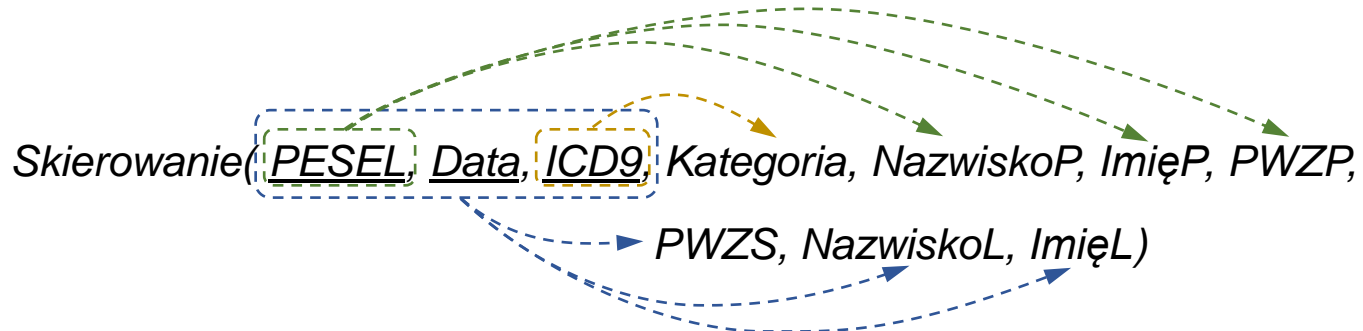
- *PESEL* → *NazwiskoP*, *ImięP*, *PWZP*,
- *PESEL*, *Data*, *ICD9* → *PWZS*, *NazwiskoL*, *ImięL*,
- *ICD9* → *Kategoria*.

Wnioski

- Relacja nie jest w IIFN, ponieważ atrybuty *NazwiskoP*, *ImięP*, *PWZP*, *PWZS*, *NazwiskoL*, *ImięL*, *Kategoria* nie są zależne elementarnie od klucza.
- Należy wykonać dekompozycję pierwotnej relacji *Skierowanie* tworząc trzy relacje pochodne (istnieją trzy źródła zależności elementarnych).

II forma normalna – dekompozycja

W celu doprowadzenia relacji do IIFN należy dokonać dekompozycji na trzy relacje, których kluczem zostaną składniki będące źródłem zależności elementarnych (s.19).



Dekompozycja

- pierwsze źródło zależności elementarnej (*PESEL*, *Data*, *ICD9*):

Skierowanie1(PESEL, Data, ICD9, *PWZS*, *NazwiskoL*, *ImięL*)

- drugie źródło zależności elementarnej (*PESEL*):

Skierowanie2(PESEL, *NazwiskoP*, *ImięP*, *PWZP*)

- trzecie źródło zależności elementarnej (*ICD9*):

Skierowanie3(ICD9, *Kategoria*)

Relacja R jest w **trzeciej formie normalnej (III FN)** jeżeli jest w IIFN i każdy ze składników, który nie jest elementem klucza jest w zależności funkcjonalnej bezpośredniej od klucza.

Transformacja relacji z IIFN do III FN

- Wyznaczenie wszystkich zależności bezpośrednich pomiędzy atrybutami relacji (niezależnie czy są one składnikami klucza),
- Dekompozycja (podział) relacji pierwotnej poprzez wyłączenie atrybutów, które nie były zależne bezpośrednio,
- Utworzenie nowej relacji dla każdego źródła zależności bezpośredniej zawierającej źródło zależności (klucz) oraz wszystkie atrybuty, które były od niego zależne bezpośrednio.

Uwaga: Źródła zależności bezpośrednich stają się kluczami nowopowstałych relacji, jednak pozostają również w relacji pierwotnej. Ich obecność gwarantuje utrzymanie powiązania pomiędzy tworzonymi relacjami.

III forma normalna relacji – przykład

Zestaw relacji w IIFN

Skierowanie1(PESEL, Data, ICD9, PWZS, NazwiskoL, ImięL)

Skierowanie2(PESEL, NazwiskoP, ImięP, PWZP)

Skierowanie3(ICD9, Kategoria)

- Relacje *Skierowanie2* i *Skierowanie3* są w IIIFN (nie ma zależności pośrednich).
- Relacja *Skierowanie1* nie spełnia warunków IIIFN (istnieje zależność pośrednia).

Istnieje zależność funkcjonalna (patrz graf. s.16):

$PWZS \rightarrow \text{NazwiskoL}, \text{ImięL},$

stąd:

$PESEL, Data, ICD9 \rightarrow PWZS, \quad PWZS \rightarrow \text{NazwiskoL}, \text{ImięL},$

więc istnieją składniki (*NazwiskoL*, *ImięL*), które nie są zależne bezpośrednio.

Wniosek

Należy wykonać dekompozycję relacji *Skierowanie1* wyodrębniając nazwisko i imię lekarza do osobnej relacji.

III forma normalna – dekompozycja

W celu doprowadzenia relacji do IIIFN należy dokonać dekompozycji na dwie relacje, których kluczem zostaną składniki będące źródłem zależności bezpośrednich (patrz graf zależności funkcjonalnych s.16).



Dekompozycja

- pierwsze źródło zależności bezpośredniej (PESEL, Data, ICD9):

Skierowanie1-1(PESEL, Data, ICD9, PWZS)

- drugie źródło zależności bezpośredniej (PZWS):

Skierowanie1-2(PZWS, NazwiskoL, ImięL)

Proces kolejnych transformacji, którym poddano pierwotną relację *Skierowanie* (s.6) określa się normalizacją danych. Jego celem jest uzyskanie optymalnego modelu danych, pozbawionego redundancji i pozostałych anomalii przedstawionych na s.8. Wynikiem normalizacji jest zestaw relacji w IIFN postaci:

Skierowanie1-1(PESEL, Data, ICD9, PWZS)

Skierowanie1-2(PZWS, NazwiskoL, ImięL)

Skierowanie2(PESEL, NazwiskoP, ImięP, PWZP)

Skierowanie3(ICD9, Kategoria)

Ostatnim krokiem modelowania jest przyjęcie nazw relacji, które określają ich faktyczną zawartość. W tym przypadku warto również zauważyć, że atrybuty *PWZP* i *PWZS* (wprowadzone ze względów formalnych) odnoszą się do tego samego atrybutu i ich nazwy mogą być ujednoczone do *PWZ*. Ostateczny zestaw relacji przyjmuje postać:

Skierowanie(PESEL, Data, ICD9, PWZ)

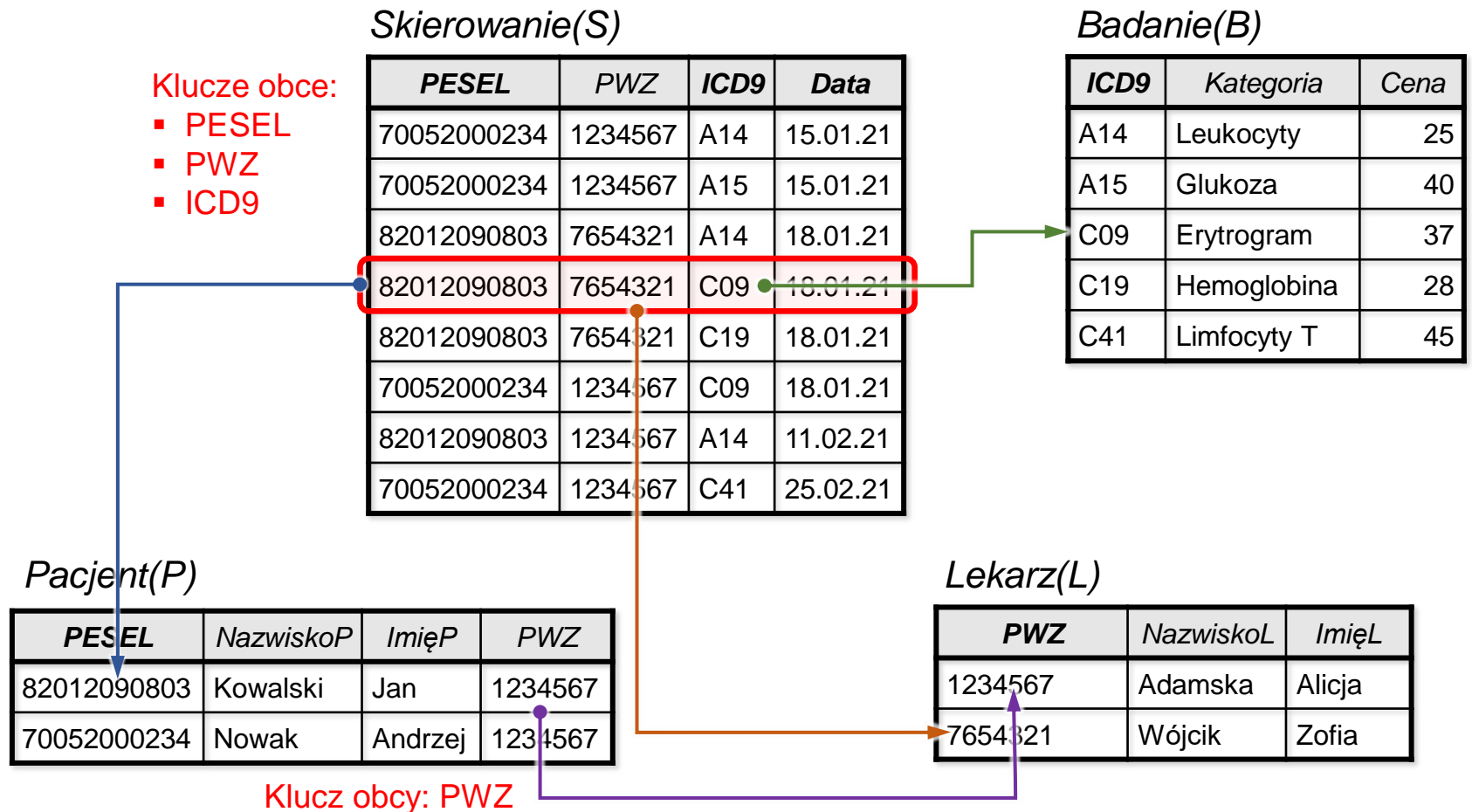
Lekarz(PZW, NazwiskoL, ImięL)

Pacjent(PESEL, NazwiskoP, ImięP, PWZ)

Badanie(ICD9, Kategoria)

Przykładowy zestaw danych

Przykładowy zestaw danych (s.7) zapisany w nowych strukturach. Powiązania pomiędzy relacjami zrealizowane przez klucze obce.



Analiza rozwiązania (1)

Pacjent (P)

PESEL	NazwiskoP	ImięP	PWZ
82012090803	Kowalski	Jan	1234567
70052000234	Nowak	Andrzej	1234567

Lekarz (L)

PWZ	NazwiskoL	ImięL
1234567	Adamska	Alicja
7654321	Wójcik	Zofia

Badanie (B)

ICD9	Kategoria	Cena
A14	Leukocyty	25
A15	Glukoza	40
C09	Erytrogram	37
C19	Hemoglobina	28
C41	Limfocyty T	45

Skierowanie (S)

PESEL	PWZ	ICD9	Data
70052000234	1234567	A14	15.01.21
70052000234	1234567	A15	15.01.21
82012090803	7654321	A14	18.01.21
82012090803	7654321	C09	18.01.21
82012090803	7654321	C19	18.01.21
70052000234	1234567	C09	18.01.21
82012090803	1234567	A14	11.02.21
70052000234	1234567	C41	25.02.21

Przyjmując rozmiary atrybutów (w znakach): PESEL – 11, Nazwisko, Imię – 25, PWZ – 7, ICD9 – 6, Kategoria – 50, Cena – 2, Data – 8, otrzymujemy:

- **pierwotna relacja:** $(11+25+25+7+25+25+6+50+2+8) \times 8 = 184 \times 8 = 1472$
- **zestaw relacji:** $(11+25+25+7) \times 2 + (7+25+25) \times 2 + (6+50+2) \times 5 + (11+7+6+8) \times 8 =$
 $68 \times 2 + 57 \times 2 + 58 \times 5 + 32 \times 8 =$
 $136 + 114 + 290 + 256 = 796$

Uwaga: Każde kolejne badanie powiększa rozmiar relacji pierwotnej o 182 znaki, znormalizowany zestaw relacji o 32 znaki.

Cechy uzyskanego modelu danych:

1. Zoptymalizowany zestaw relacji, który pozwala wykonać aplikację zgodną z pierwotnym przeznaczeniem, realizującą zaplanowane funkcje (s.2).
2. Większa elastyczność modelu danych, łatwa rozbudowa schematów relacji bez nadmiernego powiększania rozmiaru bazy danych (np. uzupełnienie relacji *Pacjent* o dodatkowe dane osobowe).
3. Wyeliminowanie redundancji, znacząca redukcja rozmiaru bazy danych w porównaniu do pierwotnej formy relacji (różnice tym większe im większa liczba skierowań).
4. Wyeliminowanie niespójności danych, określenie lekarza rodzinnego pacjenta nie wymaga analizy całej bazy.
5. Brak anomalii przy wstawianiu, aktualizacji i usuwaniu, baza zawiera dane wszystkich pacjentów (niezależnie czy otrzymali skierowanie na badanie), aktualizacja danych osobowych nie wymaga zmian w wielu wierszach, dane lekarzy odseparowane od skierowań.