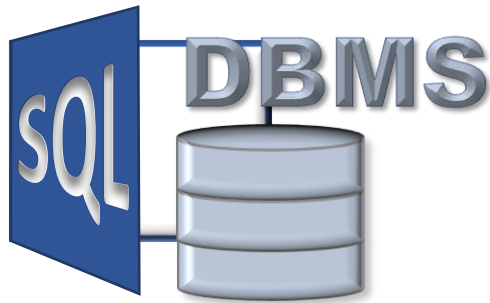


Bazy danych



Bazy danych – wprowadzenie

pojęcia podstawowe
tworzenie projektu informatycznego
modelowanie i modele danych
architektura systemów baz danych


Język SQL

- ❑ L. Beighley, *SQL*, Helion, Gliwice 2011.
- ❑ M.J. Hernandez, *Zapytania SQL dla zwykłych śmiertelników : praktyka obróbki danych w języku SQL*, Mikom, 2001.
- ❑ L. Rockoff, *Język SQL : przyjazny podręcznik*, Helion, 2017.
- ❑ R.K. Stephens, R.R. Plew, B. Morgan, J. Perkins, *SQL w 3 tygodnie*, LT&P, 1999.
- ❑ R.K. Stephens, A. Jones, R.R. Plew, *SQL w 24 godziny*, Helion, 2016.


Bazy danych

- ❑ P. Beynon-Davies, *Systemy baz danych*, WNT, 2003.
- ❑ M.J. Hernandez, *Bazy danych dla zwykłych śmiertelników*, Mikom, 2000.
- ❑ J.D.Ullman, J.Widom, *Podstawowy wykład z systemów baz danych*, WNT, 2001.
- ❑ M. Whitehorn, *Relacyjne bazy danych : teoria i praktyka projektowania relacyjnego baz danych*, Helion, 2011.

Dostęp do zasobów Springer'a z wyszukiwarki EBSCO (<http://www.bu.uz.zgora.pl>)




BIBLIOTEKA UNIWERSYTETU ZIELONOGÓRSKIEGO



Multiwyszukiwarka E-czasopisma **E-książki** Przejdź do listy A-Z

Przeszukuj tytuły książek elektronicznych



Instrukcja konfiguracji przeglądarki: <https://www.ck.uz.zgora.pl/index.php?w3cacheproxy>

- ❑ J. R. Russo, *SQL by Example*, New York : Momentum Press. 2018.
- ❑ A. Kriegel, *Discovering SQL : A Hands-On Guide for Beginners*, Ind: Wrox., 2011.
- ❑ J.L. Harrington, *SQL Clearly Explained*, 2nd ed., Morgan Kaufmann, 2003.
- ❑ J.L. Harrington, *Relational Database Design and Implementation: Clearly Explained*, 3rd ed. Amsterdam : Morgan Kaufmann. 2009.

System informatyczny – zespół systemów komputerowych (komputery, urządzenia peryferyjne) i oprogramowania służący do przetwarzania informacji.

Baza danych – uporządkowany zbiór danych, dostępnych dla licznych użytkowników, w którym można przeprowadzić efektywne wyszukiwanie i aktualizowanie informacji.

System zarządzania bazą danych (SZBD, ang.DBMS) – program zawierający narzędzia umożliwiające budowę i przetwarzanie bazy danych o dowolnej strukturze.

System bazy danych – baza danych wraz z jej systemem zarządzania.

Architektura systemu baz danych

- Poziom fizyczny – pliki dyskowe i algorytmy zapewniające dostęp do danych,
- Poziom logiczny – konceptualny model bazy danych, obejmuje całą zawartość bazy,
- Poziom zewnętrzny (użytkowy) – zbiór widoków poprzez które użytkownicy uzyskują dostęp do odpowiednich fragmentów bazy danych.

Etapy budowy projektu informatycznego

Analiza

Definiowanie przeznaczenia i funkcjonalności aplikacji.

Projekt

Projekt struktury bazy danych i procesów aplikacji niezbędnych do zaimplementowania żądanych funkcji.

Budowa (implementacja)

Przekształcenie projektu w aplikację poprzez utworzenie odpowiednich składników bazy danych i programu.

Testowanie

Sprawdzenie aplikacji pod kątem zgodności z założeniami (wymaganiami użytkownika), przeznaczeniem i zakresem funkcji.

Instalacja (wdrożenie)

Uruchomienie aplikacji w środowisku użytkownika, szkolenia z zakresu obsługi.

Etapy projektu na podstawie: K.Henderson, Bazy danych w architekturze klient-serwer, Robomatic 1998

Analiza wymaga kontaktu z użytkownikiem końcowym i rozpoznania wymagań. Należy unikać zagadnień związanych z wyglądem lub działaniem aplikacji i skupić się na oczekiwaniach użytkownika związanych z funkcjonalnością.

Krok 1: Definiowanie przeznaczenia

Przeznaczenie należy wyrazić jednym, prostym zdaniem zawierającym: podmiot (opisuje aplikację), orzeczenie (podstawowe zadanie aplikacji) oraz dopełnienie (opisuje obiekt, którego dotyczy zadanie). Sformułowane przeznaczenie nie zawiera listy funkcji, powinno jednak obejmować wszystkie aspekty przyszłego zastosowania programu.

Krok 2: Definiowanie funkcji

Listę należy ograniczyć do najważniejszych funkcji aplikacji. Każda powinna bliżej definiować przeznaczenie i nie może wykraczać poza główny cel (kroku 1). Konstrukcja zdań przypomina zdanie wyrażające przeznaczenie całej aplikacji.

Krok 3: Schemat opisowy

Syntetyczny opis prezentujący funkcjonowanie fragmentu rzeczywistości, stanowi uzupełnienie listy funkcji i powinien zawierać opis danych, które będą przetwarzane. Schemat opisowy jest podstawą do stworzenia modelu danych.

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 danych – dobrze zdefiniowany sposób opisu świata rzeczywistego.

Najważniejsze modele danych

- ❑ **kartotekowy** – dane zapisywane są w pojedynczej tablicy, każda tablica jest samodzielnym dokumentem i nie może współpracować z innymi tablicami,
- ❑ **sieciowy** – oparty na strukturze grafu, obiekty są reprezentowane przez wierzchołki grafu, powiązania przez jego krawędzie,
- ❑ **hierarchiczny** – modyfikacja modelu sieciowego, oparty na strukturze drzewa (graf, który nie zawiera cykli),
- ❑ **relacyjny** – wykorzystuje matematyczne pojęcie relacji, przy jej pomocy reprezentuje zarówno obiekty jak i powiązania,
- ❑ **obiektyowy** – wprowadza rozszerzone pojęcie obiektu (zgodne z podejściem stosowanym w programowaniu obiektowym), w którym uwzględniane są atrybuty (reprezentujące cechy) oraz metody (reprezentujące zachowania/umiejętności) rzeczywistych obiektów występujących w modelowanym systemie.

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.

Powiązanie (związek) – opis stanu, który dotyczy dwóch lub więcej obiektów.

Typy powiązań

- jeden-do-jeden – krotce z relacji pierwszej odpowiada dokładnie jedna krotka z relacji drugiej i odwrotnie (zazwyczaj eliminowane z modelu danych),
- jeden-do-wielu – krotce z relacji pierwszej odpowiada wiele krotek z relacji drugiej, ale jednej krotce z relacji drugiej odpowiada jedna krotka w relacji pierwszej,
- wiele-do-wielu – krotce z relacji pierwszej odpowiada wiele krotek z relacji drugiej i krotce z relacji drugiej odpowiada wiele krotek z relacji pierwszej.

Klucz obcy – atrybut lub ciąg atrybutów, który występuje w pewnej relacji i jednocześnie jest identyfikatorem kluczowym innej relacji. Tworzy powiązanie pomiędzy relacjami.

Przykład

Lekarz(PWZ, NazwiskoL, ImięL)

Pacjent(PESEL, NazwiskoP, ImięP, PWZ)

PWZ jest kluczem obcym w relacji *Pacjent*, tworzy powiązanie typu jeden-do-wielu pomiędzy *Pacjentem* i *Lekarzem* (np. lekarz rodzinny danego pacjenta).

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 stronie następnej.

Przykładowy zbiór krotek relacji *Skierowanie*

PESEL	NazwiskoP	ImięP	PWZ	NazwiskoL	ImięL	ICD9	Kategoria	Cena	Data
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

Wybór klucza naturalnego:

- Pojedyncze atrybuty nie identyfikują jednoznacznie krotki relacji.
- PESEL+ICD9 – nie jest unikalny (pacjent może powtarzać badanie).
- ICD9+Data – nie jest unikalny (to samo badanie w tym samym dniu).
- PESEL+ICD9+Data – unikalny (w danym dniu jedno zlecenie na badanie dla pacjenta).

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.

Przykład – rozwiązanie znormalizowane

Obiekty: pacjent, lekarz, badanie.

Powiązania: pacjent-lekarz, pacjent-lekarz-badanie.

Relacje reprezentujące obiekty

Pacjent(PESEL, NazwiskoP, ImięP)

Lekarz(PWZ, NazwiskoL, ImięL)

Badanie(ICD9, Kategoria, Cena)

Relacje reprezentujące powiązania

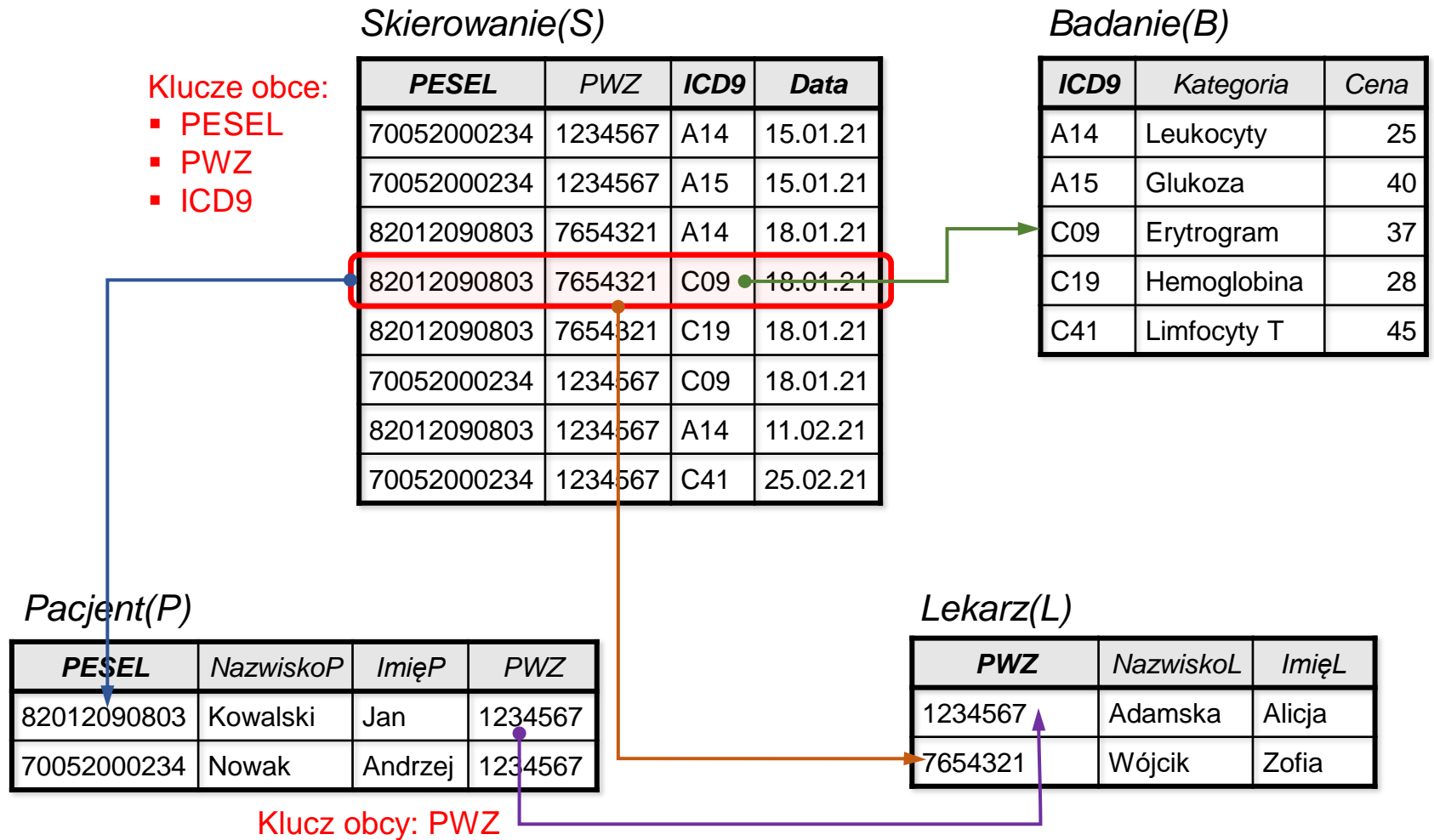
- pacjent-lekarz – jeden pacjent ma dokładnie jednego lekarza, dodatkowa relacja nie jest wymagana, powiązanie jest realizowane przez uzupełnienie danych pacjenta o identyfikator lekarza:

Pacjent(PESEL, NazwiskoP, ImięP, PWZ)

- pacjent-lekarz-badanie – jeden pacjent ma wiele badań, jeden lekarz wystawia wiele skierowań, to samo badanie występuje na wielu skierowaniach (dotyczy wielu pacjentów, jest zlecane przez wielu lekarzy), konieczna dodatkowa relacja:

Skierowanie(PESEL, PWZ, ICD9, Data)

Przykładowy zestaw danych



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:** $(11+25+25+7) \times 2 + (7+25+25) \times 2 + (6+50+2) \times 5 + (11+7+6+8) \times 8 =$
 $68 \times 2 + 57 \times 2 + 58 \times 5 + 32 \times 8 =$
 $136 + 114 + 290 + 256 = 796$

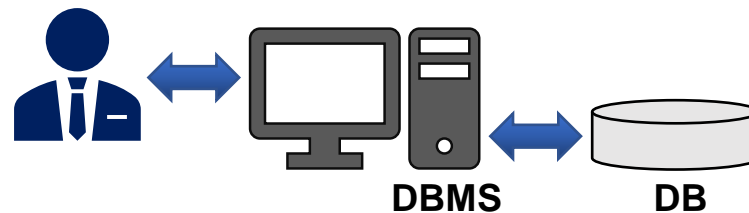
Uwaga: Każde kolejne badanie powiększa rozmiar relacji pierwotnej o 182 znaki, znormalizowany zestaw relacji o 32 znaki.

Cechy uzyskanego modelu danych:

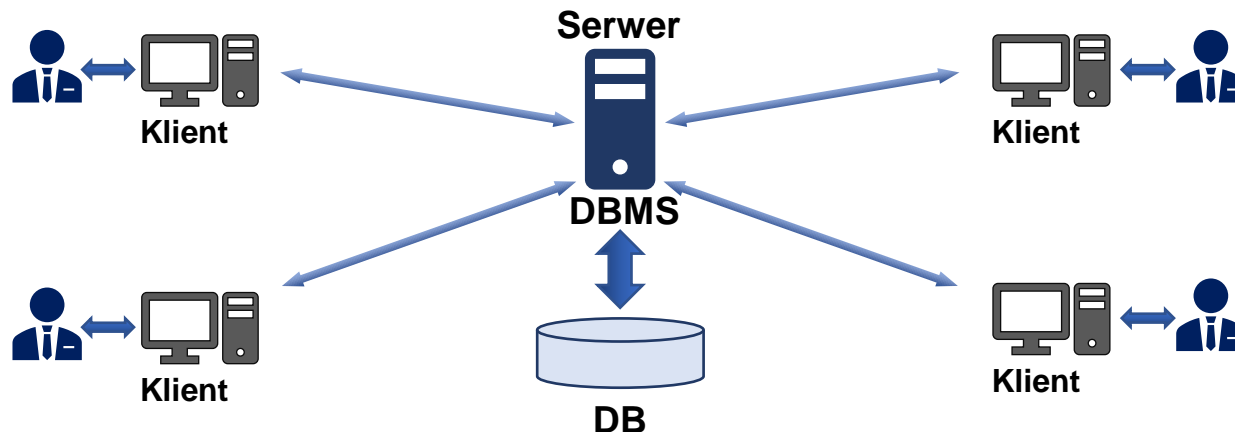
1. Zoptymalizowany zestaw relacji, który pozwala wykonać aplikację zgodną z pierwotnym przeznaczeniem, realizującą zaplanowane funkcje (s.1-7).
2. Większa elastyczność modelu danych, łatwa rozbudowa schematów relacji bez nadmiernego powiększania rozmiaru bazy danych (np. uzupełnienie relacji *Pacjent* 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.

Architektura systemów baz danych

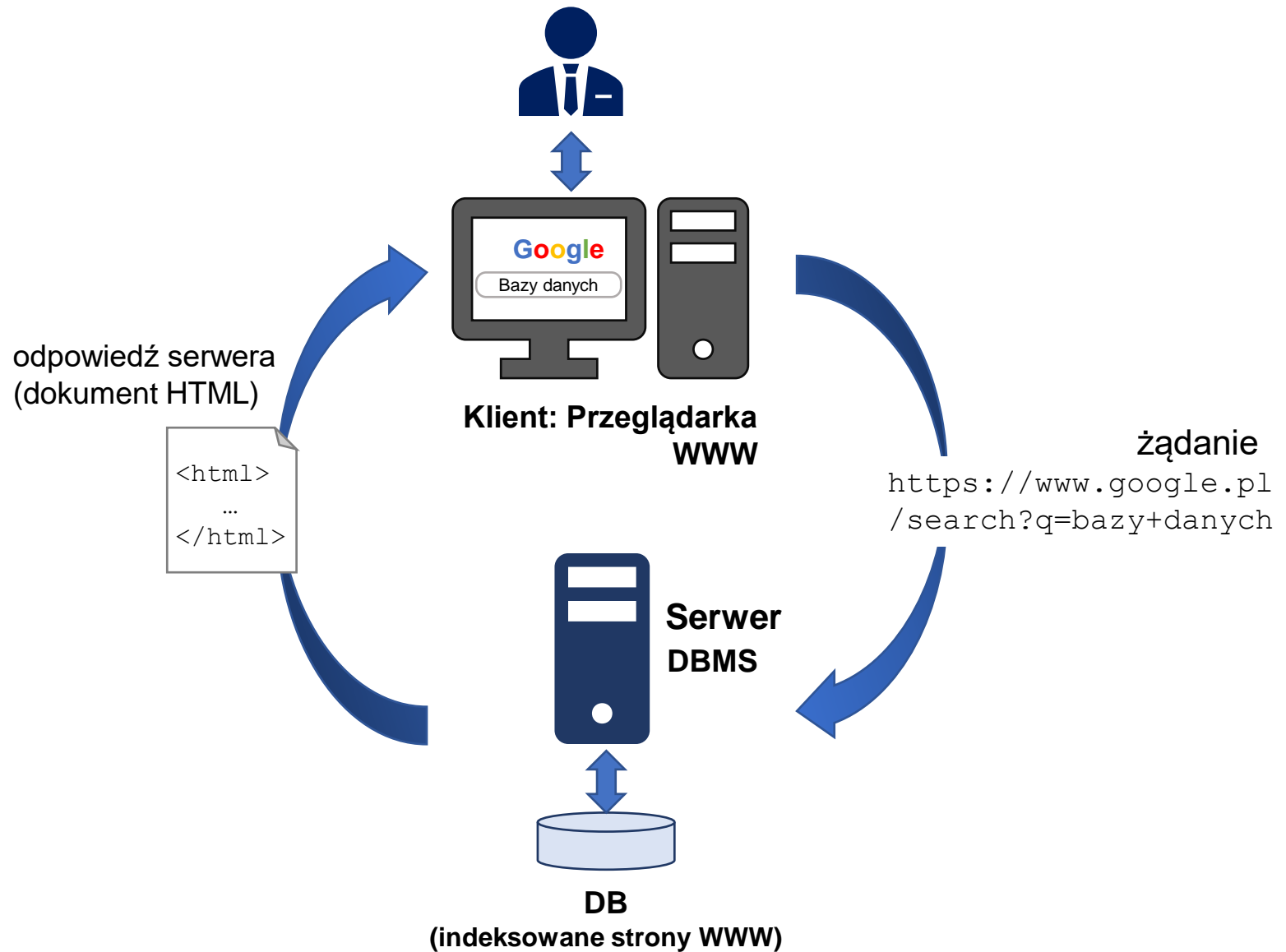
Lokalna baza danych – rodzaj bazy danych do użytku indywidualnego, system zarządzania bazą danych oraz zbiór danych, który przetwarza znajdują się na jednym komputerze, dane nie są dostępne z innych jednostek.



Baza klient-serwer – składa się z wyróżnionego węzła nazywanego serwerem i szeregu podłączonych do niego węzłów nazywanych klientami. Serwer zawiera bazę danych i system zarządzania, i wykonuje usługi zlecane przez klientów (przetwarzanie danych). Rolą klienta jest przesłanie żądania i oczekiwanie na odpowiedź serwera. Architektura klient-serwer pozwala na współużytkowanie bazy przez wielu użytkowników (klientów).



Przykład – wyszukiwarka Google



The screenshot displays the phpMyAdmin interface. The top window shows the 'Struktura' (Structure) view for the 'skierowania' database. It lists four tables: B, L, P, and S. Below this, a second window shows the 'Przeglądaj' (Browse) view for the table 'P' (Dane pacjentów). A SQL query is entered and ready to be executed.

Tabela	Działanie	Rekordy	Typ	Metoda porównywania napisów	Rozmiar	Nadmiar
<input type="checkbox"/> B	Przeglądaj Struktura Szukaj Wstaw Opróżnij Usuń	5	InnoDB	utf8mb4_0900_ai_ci	16.0 KB	-
<input type="checkbox"/> L	Przeglądaj Struktura Szukaj Wstaw Opróżnij Usuń	2	InnoDB	utf8mb4_0900_ai_ci	16.0 KB	-
<input type="checkbox"/> P	Przeglądaj Struktura Szukaj Wstaw Opróżnij Usuń	4	InnoDB	utf8mb4_0900_ai_ci	16.0 KB	-
<input type="checkbox"/> S	Przeglądaj Struktura Szukaj Wstaw Opróżnij Usuń	8	InnoDB	utf8mb4_0900_ai_ci	16.0 KB	-
4 tabel	Suma	19	InnoDB	utf8mb4_0900_ai_ci	64.0 KB	0 B


```

1 SELECT *
2 FROM P INNER JOIN S ON P.PESEL=S.PESEL
3 WHERE S.ICD9 = 'A14'
    
```

Columns: PESEL, NazwiskoP, ImieP, PWZ

Buttons: SELECT *, SELECT, INSERT, UPDATE, DELETE, Wyczyść, Format, Wykonaj

Options: Parametry wiązania, Pokaż to zapytanie tutaj ponownie, Zachowaj pole zapytania, Przywróć po zakończeniu, Włącz sprawdzanie kluczy obcych

phpMyAdmin w trybie podglądu struktury bazy i wprowadzania kwerendy SQL