

Relacyjny model danych

**Wykład przygotował:
Robert Wrembel**



Plan wykładu

- Relacyjny model danych
- Struktury danych
- Operacje
- Ograniczenia integralnościowe

W ramach drugiego wykładu z baz danych zostanie przedstawiony relacyjny model danych, który w praktyce jest najczęściej stosowany. W szczególności wykład omówi: struktury danych tego modelu, operacje modelu i ograniczenia integralnościowe.



Model danych

- Definiuje
 - struktury danych
 - operacje
 - ograniczenia integralnościowe
- Relacyjny model danych
 - relacje
 - selekcja, projekcja, połączenie, operacje na zbiorach
 - klucz podstawowy, klucz obcy, zawężenie dziedziny, unikalność, wartość pusta/niepusta

W ogólności model danych definiuje:

- struktury wykorzystywane do reprezentowania danych,
- operacje na danych,
- ograniczenia integralnościowe, czyli reguły poprawności danych.

Jednym z fundamentalnych modeli jest model relacyjny. Jest on wykorzystywany w większości komercyjnych i niekomercyjnych systemów baz danych. W modelu tym, strukturą danych jest relacja; operacje na danych obejmują selekcję, projekcję, połączenie i operacje na zbiorach. Ograniczenia integralnościowe w tym modelu to: klucz podstawowy, klucz obcy, zawężenie dziedziny, unikalność wartości, możliwość nadawania wartości pustych/niepustych.



Struktury danych (1)

- Baza danych jest zbiorem relacji
- Schemat relacji R , oznaczony przez $R(A_1, A_2, \dots, A_n)$, składa się z nazwy relacji R oraz listy atrybutów A_1, A_2, \dots, A_n
- Liczbę atrybutów składających się na schemat relacji R nazywamy stopniem relacji
- Każdy atrybut A_i schematu relacji R posiada domenę, oznaczoną jako $\text{dom}(A_i)$
- Domena definiuje zbiór wartości atrybut relacji poprzez podanie typu danych

BD – wykład 2 (4)

W modelu relacyjnym, baza danych jest zbiorem relacji. Każda relacja posiada swój tzw. schemat, który składa się z listy atrybutów. Schemat relacji R jest często oznaczany jako $R(A_1, A_2, \dots, A_n)$, gdzie A_1, A_2, \dots, A_n oznaczają atrybuty. Liczbę atrybutów składających się na schemat relacji R nazywamy stopniem relacji.

Każdy atrybut posiada swoją domenę, zwaną także dziedziną. Definiuje ona zbiór wartości jakie może przyjmować atrybut poprzez określenie tzw. typu danych, np. liczba całkowita, data, ciąg znaków o długości 30.



Struktury danych (2)

- Relacją r o schemacie $R(A_1, A_2, \dots, A_n)$, oznaczoną $r(R)$, nazywamy zbiór n -tek (krotek) postaci $r = \{t_1, t_2, \dots, t_m\}$.
- Pojedyncza krotka t jest uporządkowaną listą n wartości $t = \langle v_1, v_2, \dots, v_n \rangle$, gdzie $v_i, 1 \leq i \leq n$, jest elementem $\text{dom}(A_i)$ lub specjalną wartością pustą (NULL)
- i -ta wartość krotki t , odpowiadająca wartości atrybutu A_i , będzie oznaczana przez $t[A_i]$
- Relacja $r(R)$ jest relacją matematyczną stopnia n zdefiniowaną na zbiorze domen $\text{dom}(A_1), \text{dom}(A_2), \dots, \text{dom}(A_n)$ będącą podzbiorem iloczynu kartezjańskiego domen definiujących R :
$$r(R) \subseteq \text{dom}(A_1) \times \text{dom}(A_2) \times \dots \times \text{dom}(A_n)$$

BD – wykład 2 (5)

Formalna definicja relacji jest następująca:

Relacją r o schemacie $R(A_1, A_2, \dots, A_n)$, oznaczoną $r(R)$, nazywamy zbiór n -tek (krotek) postaci $r = \{t_1, t_2, \dots, t_m\}$.

Pojedyncza krotka t jest uporządkowaną listą n wartości $t = \langle v_1, v_2, \dots, v_n \rangle$, gdzie $v_i, 1 < i < n$, jest elementem $\text{dom}(A_i)$ lub specjalną wartością pustą (NULL).

i -ta wartość krotki t , odpowiadająca wartości atrybutu A_i , będzie oznaczana przez $t[A_i]$.

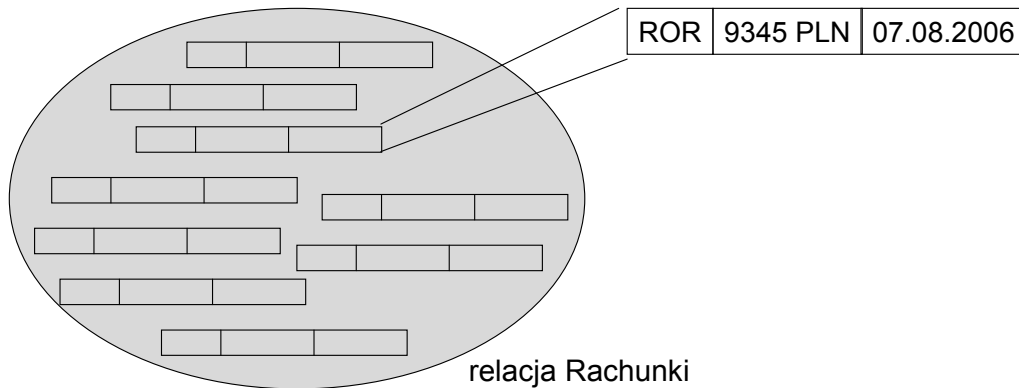
Matematyczna definicja relacji jest następująca:

Relacja $r(R)$ jest relacją matematyczną stopnia n zdefiniowaną na zbiorze domen $\text{dom}(A_1), \text{dom}(A_2), \dots, \text{dom}(A_n)$ będącą podzbiorem iloczynu kartezjańskiego domen definiujących R .



Struktury danych (3)

Relacja jest zbiorem krotek (k-wartości), które są listami wartości



Innymi słowy, relacja jest zbiorem krotek (k-wartości), które są listami wartości. Przykładowo, relacja Rachunki jest złożona ze zbioru krotek. Każda z nich przechowuje trzy wartości, tj. rodzaj rachunku, saldo i datę jego ważności.



Alternatywna definicja relacji

- Wyświetlana relacja ma postać tabeli
 - krotki są wierszami tej tabeli
 - nagłówki kolumn są atrybutami

IdPrac	Nazwisko	Etat	Szef	Zatrudniony	Płaca	IdZesp
7369	SMITH	SPRZEDAWCA	7902	1980-12-17	2048	20
7499	ALLEN	SPRZEDAWCA	7698	1981-02-20	4096	10
7566	JONES	KIEROWNIK	7839	1981-04-02	7616	20
7654	MARTIN	SPRZEDAWCA	7698	1981-09-28	3200	30
7782	CLARK	KIEROWNIK	7839	1981-06-09	6272	10
7788	SCOTT	KIEROWCA	7566	1982-12-09	7680	20
7839	KING	DYREKTOR		1981-11-17	12800	10
7844	TURNER	SPRZEDAWCA	7698	1981-09-08	3840	30

Intuicyjnie, relacja ma postać klasycznej tabeli z kolumnami i wierszami. Kolumny odpowiadają atrybutom relacji, a wiersze (zwane również rekordami) odpowiadają krotkom.



Baza danych

- Baza danych = zbiór relacji
- Schemat bazy danych = zbiór schematów relacji
- Schemat relacji = zbiór {atrybut, dziedzina, [ograniczenia integralnościowe]}
- Relacja = zbiór krotek
- Krotka = lista wartości atomowych

Slajd ten podsumowuje omówione wcześniej definicje.

Baza danych jest zbiorem relacji.

Schemat relacji jest zbiorem {atrybut, dziedzina, [ograniczenia integralnościowe]}.

Schemat bazy danych jest zbiorem schematów relacji.

Relacja jest zbiorem krotek.

Krotka jest listą wartości atomowych.



Charakterystyka relacji

- Każdy atrybut relacji ma unikalną nazwę
- Porządek atrybutów w relacji nie jest istotny
- Porządek krotek w relacji nie jest istotny i nie jest elementem definicji relacji
- Wartości atrybutów są atomowe (elementarne)
- Relacja nie zawiera rekordów powtarzających się

Relacja posiada następujące cechy:

- każdy atrybut relacji ma unikalną nazwę,
- porządek atrybutów w relacji nie jest istotny,
- porządek krotek w relacji nie jest istotny i nie jest elementem definicji relacji,
- wartości atrybutów są atomowe (elementarne),
- relacja nie zawiera rekordów powtarzających się. Ponieważ relacja jest zbiorem krotek, więc, z definicji zbioru, wszystkie krotki relacji muszą być unikalne.



Unikalność krotek relacji - klucze (1)

- Ograniczenie na unikalność krotek relacji
 - Każdy podzbiór S atrybutów relacji R, taki że dla każdych dwóch krotek ze zbioru $r(R)$ zachodzi $t_1[S] \neq t_2[S] \Rightarrow$ superkluczem (super key) R
 - Superklucz
 - cały schemat relacji

Każdy podzbiór S atrybutów relacji R, jest nazywany superkluczem (ang. super key) relacji R jeżeli dla każdych dwóch krotek ze zbioru $r(R)$ zachodzi $t_1[S] \neq t_2[S]$. W ogólności, cały schemat relacji jest superkluczem.



Unikalność krotek relacji - klucze (2)

- Superklucz może posiadać nadmiarowe atrybuty
- Kluczem K schematu relacji R nazywamy superklucz schematu R o takiej własności, że usunięcie dowolnego atrybutu A z K powoduje, że $K' = K - A$ nie jest już superkluczem
- Klucz jest minimalnym superkluczem zachowującym własność unikalności krotek relacji
- Schemat relacji może posiadać więcej niż jeden klucz

Superklucz może posiadać nadmiarowe atrybuty. Kluczem K schematu relacji R nazywamy superklucz schematu R o takiej własności, że usunięcie dowolnego atrybutu A z K powoduje, że $K' = K - A$ nie jest już superkluczem.

Klucz jest minimalnym superkluczem zachowującym własność unikalności krotek relacji.

Schemat relacji może posiadać więcej niż jeden klucz.



Unikalność krotek relacji - klucze (3)

- Wyróżniony klucz \Rightarrow klucz podstawowy
- Pozostałe klucze \Rightarrow klucze wtórne lub kandydujące

Jeden z kluczy relacji może być wyróżniony jako tzw. klucz podstawowy, który jednoznacznie identyfikuje krotki relacji. W związku z tym, klucz podstawowy nie może przyjmować wartości pustych. Pozostałe klucze schematu relacji nazywamy kluczami wtórnymi lub kandydującymi.



Ograniczenie integralnościowe

- Mechanizm (reguła), który gwarantuje że dane wpisane to relacji spełnią nałożone na nie warunki
 - czuwa nad tym SZBD
- Definiuje się na poziomie
 - pojedynczego atrybutu
 - całej relacji
- Rodzaje
 - klucz podstawowy (primary key)
 - klucz obcy (foreign key)
 - unikalność (unique)
 - zawężenie domeny/dziedziny (check)
 - wartość pusta/niepusta (NULL/NOT NULL)

BD – wykład 2 (13)

Każda relacja może posiadać jawnie zdefiniowane ograniczenia integralnościowe. Ograniczenie integralnościowe jest pewną regułą gwarantującą, że dane znajdujące się w relacji spełniają tę regułę. W praktyce nad zapewnieniem integralności danych czuwa SZBD. Ograniczenie integralnościowe definiuje się albo dla pojedynczego atrybutu albo dla całej relacji.

Wyróżnia się następujące ograniczenia integralnościowe:

- klucz podstawowy (primary key),
- klucz obcy (foreign key),
- unikalność (unique),
- zawężenie domeny/dziedziny (check),
- wartość pusta/niepusta (NULL/NOT NULL).



Klucz podstawowy

- Klucz podstawowy relacji (primary key)
 - atrybut (lub zbiór atrybutów), którego wartość jednoznacznie identyfikuje krotkę
 - wartość ta jest unikalna w obrębie całej relacji i jest niepusta
- Przykłady:
 - adres e-mail, NIP, PESEL, nr dowodu, nr paszportu

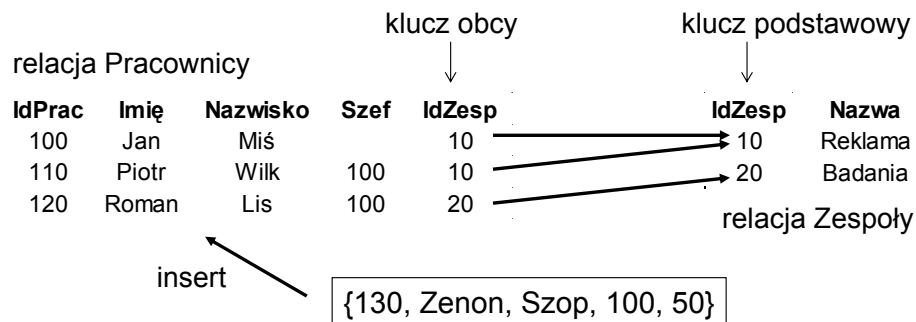
Klucz podstawowy relacji (ang. primary key) jest to atrybut lub zbiór atrybutów, którego wartość jednoznacznie identyfikuje krotkę relacji. Z definicji, wartość atrybutu, który zdefiniowano jako klucz podstawowy jest unikalna w obrębie całej relacji i jest niepusta.

Przykładami atrybutów, które mogły by być kluczami podstawowymi są np. adres e-mail, NIP, PESEL, nr dowodu, nr paszportu.



Klucz obcy (1)

- Klucz obcy relacji (foreign key)
 - atrybut (lub zbiór atrybutów), który wskazuje na klucz podstawowy
 - służy do reprezentowania powiązań między danymi (łączenia relacji)



Klucz obcy relacji (ang. foreign key) jest atrybutem lub zbiorem atrybutów, który wskazuje na klucz podstawowy w innej relacji. Klucz obcy służy do reprezentowania powiązań między danymi (łączenia relacji). Dziedziną wartości klucza obcego jest dziedzina wartości klucza podstawowego, na który ten klucz obcy wskazuje.

W przykładzie ze slajdu, w relacji Zespoły kluczem podstawowym jest atrybut IdZesp. W relacji Pracownicy kluczem obcym jest IdZesp i wskazuje on na IdZesp w relacji Zespoły. Wartościami atrybutu IdZesp w relacji Pracownicy mogą być tylko te wartości, które przyjmuje IdZesp w relacji Zespoły.

Przykładowy rekord {130, Zenon, Szop, 100, 50} nie zostanie wstawiony do relacji Pracownicy, ponieważ wartość atrybutu IdZesp (50) nie znajduje się w relacji Zespoły. Naruszono w tym przypadku ograniczenie integralnościowe klucza obcego.



Klucz obcy (2)

- Dane są relacje R1 i R2. Podzbiór FK atrybutów relacji R1 nazywany jest kluczem obcym R1 jeżeli:
 - atrybuty w FK mają taką samą domenę jak atrybuty klucza podstawowego PK relacji R2
 - dla każdej krotki t1 relacji R1 istnieje dokładnie jedna krotka t2 relacji R2, taka że $t1 [FK] = t2 [PK]$, lub $t1 [FK] = null$
- Klucz obcy (ograniczenie referencyjne) gwarantuje, że rekordy z tabeli R1 występują w kontekście związanego z nim rekordu z tabeli R2

Bardziej formalna definicja klucza obcego jest następująca.

Dane są relacje R1 i R2. Podzbiór FK atrybutów relacji R1 nazywany jest kluczem obcym R1 jeżeli:

- atrybuty w FK mają taką samą domenę jak atrybuty klucza podstawowego PK relacji R2,
- dla każdej krotki t1 relacji R1 istnieje dokładnie jedna krotka t2 relacji R2, taka że $t1 [FK] = t2 [PK]$, lub $t1 [FK] = null$.

Klucz obcy, zwany również ograniczeniem referencyjnym, gwarantuje, że rekordy z tabeli R1 występują w kontekście związanego z nim rekordu z tabeli R2.



Zawężenie dziedziny

- Zawężenie dziedziny (ograniczenie domeny) atrybutu (check)
 - ograniczenie dozwolonych wartości do pewnego podzbioru przez wyrażenie logiczne określające przedział lub za pomocą wyliczeniowej listy wartości
- Przykłady:
 - płeć: K, M, nieznana, N/A
 - pensja: wartości dodatnie
 - kolor oczu: niebieskie, szare, piwne

Zbiór wartości domeny atrybutu może być zawężony przez wyrażenie logiczne do pewnego podzbioru: przedziału lub wyliczeniowej listy wartości. Jest to tzw. ograniczenie integralnościowe zawężenia dziedziny (domeny). Przykładami tego typu ograniczenia są np.

- ograniczenie dopuszczalnych wartości atrybutu płeć do: K, M, nieznana, N/A (zgodnie ze standardem ISO),
- zagwarantowanie dodatnich wartości atrybutu pensja,
- ograniczenie dopuszczalnych wartości atrybut kolor_oczu do trzech wartości: niebieskie, szare, piwne.



Zawężenie dziedziny - przykład

- Etat - dziedzina: {'Analityk', 'Dyrektor', 'Referent', 'Kierownik', 'Sekretarka'}
- Płaca - dziedzina: $placa > 500$
- IdPrac - klucz podstawowy relacja Pracownicy

IdPrac	Nazwisko	Etat	Płaca	Szef	IdZesp
120	Kowalski	Analityk	850	100	10
100	Tarzan	Dyrektor	1700		10
130	Nowak	Referent	600	100	10
110	Józek	Kierownik	1200	100	20
140	Nowacki	Analityk	800	110	20
150	Bunio	Sekretarka	700	100	10

insert

{200, 'Szop', 'Księgowy', 900, 10}
 {130, 'Borsuk', 'Kierownik', 1000, 20}
 {210, 'Rosomak', 'Kierownik', 400, 20}

BD – wykład 2 (18)

Jako przykład rozważmy relację Pracownicy ze slajdu. Przyjmijmy, że dla atrybutu Etat zdefiniowano ograniczenie zawężające zbiór jego wartości do analityka, dyrektora, referenta, kierownika i sekretarki. Dla atrybutu płaca określono dziedzinę wartości większych niż 500. Atrybut IdPrac zdefiniowano jako klucz podstawowy relacji Pracownicy. Do tak zdefiniowanej relacji nie da się wstawić żadnej z trzech krotek. Pierwsza z nich narusza integralność etatu, druga narusza integralność klucza podstawowego, a trzecia - integralność płacy.



Podstawowe operacje algebry relacji

- Selekcja (SELECT)
- Projektcja (PROJECT)
- Połączenie (JOIN)
 - Iloczyn kartezjański
- Operacje na zbiorach
 - suma (UNION)
 - część wspólna (INTERSECTION)
 - różnica (MINUS, DIFFERENCE)

W modelu relacyjnym wykorzystuje się tzw. algebrę relacji, definiującą zbiór operacji na danych i semantykę tych operacji. Operacjami tymi są: selekcja, projektcja, połączenie, iloczyn kartezjański jako specjalny przypadek połączenia, operacje na zbiorach (suma, część wspólna i różnica).



Operacja selekcji

- Przeznaczenie:
 - wyodrębnienie podzbioru krotek relacji, które spełniają warunek selekcji
- Notacja: $\sigma_{\langle \text{warunek selekcji} \rangle} (\langle \text{Nazwa relacji} \rangle)$
 - warunek selekcji jest zbiorem predykatów postaci
 - $\langle \text{atrybut} \rangle \langle \text{operator relacyjny} \rangle \langle \text{litera} \rangle$
 - lub
 - $\langle \text{atrybut} \rangle \langle \text{operator relacyjny} \rangle \langle \text{atrybut} \rangle$
 - predykaty są łączone operatorami logicznymi: AND lub OR
- Własności: operacja selekcji jest komutatywna:

$$\sigma_{\langle \text{war1} \rangle} (\sigma_{\langle \text{war2} \rangle} (R)) = \sigma_{\langle \text{war2} \rangle} (\sigma_{\langle \text{war1} \rangle} (R))$$

Operacja selekcji umożliwia wyodrębnienie podzbioru krotek relacji, które spełniają warunek selekcji.

Operacja ta jest oznaczana symbolem sigma z pewnym warunkiem selekcji. Operacja ta działa na relacji o pewnej nazwie. Warunek selekcji jest zbiorem predykatów postaci $\langle \text{atrybut} \rangle \langle \text{operator relacyjny} \rangle \langle \text{litera} \rangle$ lub $\langle \text{atrybut} \rangle \langle \text{operator relacyjny} \rangle \langle \text{atrybut} \rangle$.

Predykaty są łączone operatorami logicznymi: AND lub OR.

Rozważmy dwie operacje selekcji. Operacja S1 jest realizowana jako pierwsza. S1 posiada warunek W1 i jest realizowana na relacji R. Operacja S2 jest realizowana jako druga. S2 posiada warunek W2 i jest realizowana na wyniku operacji S1. Przyjmijmy, że wynik operacji S1 i S2 wykonanych w takiej kolejności jest zbiorem krotek $\{k1, k2, k3\}$. Jeżeli zamienimy kolejność wykonywania operacji selekcji, tzn. najpierw zostanie wykonana operacja S2 z warunkiem W2 na relacji R, a następnie S1 z warunkiem W1 na wyniku działania operacji S2, to w wyniku końcowym otrzymamy identyczny zbiór krotek jak poprzednio. Taką własność operacji selekcji nazywamy komutatywnością.



Operacja selekcji - przykłady (1)

- $\sigma_{\text{IdZesp} = 10}(\text{Pracownicy})$

```
select IdPrac, Nazwisko, Etat, Szef,  
       Zatrudniony, Płaca, IdZesp  
from Pracownicy  
where IdZesp=10
```

- $\sigma_{\text{Płaca} > 7000}(\text{Pracownicy})$

```
select IdPrac, Nazwisko, Etat, Szef,  
       Zatrudniony, Płaca, IdZesp  
from Pracownicy  
where Płaca>10
```

Na slajdzie przedstawiono dwa przykłady operacji selekcji. Pierwszy z nich wybiera z relacji Pracownicy te rekordy, dla których wartość atrybut IdZesp jest równa 10. Drugi przykład wybiera z relacji Pracownicy tylko tych pracowników których wartość atrybutu Płaca jest większa niż 7000. Obie operacje wyrażono w notacji ogólnej i w języku SQL.



Operacja selekcji - przykłady (2)

- $\sigma_{(IdZesp=10 \text{ AND } Płaca>7000) \text{ OR } (IdZesp=20) \text{ AND } Płaca>8000}$
(Pracownicy)

```
select IdPrac, Nazwisko, Etat, Szef,
       Zatrudniony, Płaca, IdZesp
from Pracownicy
where (IdZesp=10 and Płaca>7000)
or     (IdZesp=20 and Płaca>8000)
```

- $\sigma_{Etat='Księgowy' \text{ AND } (Płaca \geq 6000 \text{ AND } Płaca < 9000)}$ (Pracownicy)

```
select IdPrac, Nazwisko, Etat, Szef,
       Zatrudniony, Płaca, IdZesp
from Pracownicy
where Etat='KSIĘGOWY'
and     (Płaca >= 6000 and Płaca < 9000)
```

BD – wykład 2 (22)

W pierwszym przykładzie z tego slajdu operacja selekcji wybiera z relacji Pracownicy krotki dla których wartość atrybutu IdZesp=10 i Płaca>7000 lub IdZesp=20 i Płaca>8000. Należy zwrócić tu uwagę na priorytety operatorów. AND ma wyższy priorytet niż OR, co dodatkowo zostało zaznaczone za pomocą nawiasów.

Drugi przykład ilustruje selekcję z relacji Pracownicy wszystkich księgowych zarabiających w przedziale między 6000 i 9000.



Operacja projekcji

- Przeznaczenie:
 - wyodrębnienie wybranych atrybutów relacji
- Notacja: $\pi_{\langle \text{atrybuty} \rangle} (\langle \text{Nazwa relacji} \rangle)$
 - atrybuty jest podzbiorem atrybutów ze schematu relacji
- Własności: operacja projekcji nie jest komutatywna
- Składanie operacji projekcji jest możliwe jeżeli lista2 zawiera wszystkie atrybuty lista1

$$\pi_{\langle \text{lista1} \rangle} (\pi_{\langle \text{lista2} \rangle} (R)) = \pi_{\langle \text{lista1} \rangle} (R)$$

Drugą operacją modelu relacyjnego jest projekcja. Umożliwia ona wyodrębnienie (wybór) tylko określonych atrybutów relacji.

Operacja ta jest oznaczana symbolem pi z podzbiorem wybieranych atrybutów z całego zbioru atrybutów relacji. Operacja ta działa na relacji o pewnej nazwie.

Operacja projekcji nie jest komutatywna, a składanie operacji projekcji jest możliwe jeżeli lista2 zawiera wszystkie atrybuty lista1. Notację operacji składania projekcji przedstawiono na slajdzie.



Operacja projekcji - przykłady

- π_{Nazwisko} (Pracownicy)

```
select Nazwisko  
from Pracownicy
```

- $\pi_{\text{Nazwisko, Etat, Płaca}}$ (Pracownicy)

```
select Nazwisko, Etat, Płaca  
from Pracownicy
```

Na slajdzie przedstawiono dwa przykłady projekcji. W pierwszym, ze zbioru atrybutów relacji Pracownicy jest wybierany tylko atrybut Nazwisko. Wynikiem tej operacji projekcji jest zbiór nazwisk wszystkich pracowników. W drugim przykładzie, ze zbioru atrybutów relacji Pracownicy są wybierane atrybuty Nazwisko, Etat i Płaca. W tym przypadku, wynikiem jest zbiór krotek wszystkich pracowników, ale każda z krotek posiada tylko 3 wartości: nazwiska, etatu i pensji.

Obie przykładowe operacje projekcji wyrażono w notacji ogólnej i w języku SQL.

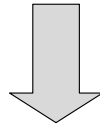


Składanie operacji

- Wynik danej operacji może być zbiorem wejściowym dla innej operacji

$\sigma_{\text{IdZesp} = 10} (\text{Pracownicy}) \Leftrightarrow \text{PracZesp10}$

$\pi_{\text{IdPrac}, \text{Nazwisko}} (\text{PracownicyZesp10}) \Leftrightarrow \text{PracZesp10Wynik}$



$\text{PracZesp10Wynik} = \pi_{\text{IdPrac}, \text{Nazwisko}} (\sigma_{\text{IdZesp} = 10} (\text{Pracownicy}))$

Sekwencja wielu operacji, w której kolejne operacje są wykonywane na pośrednich wynikach operacji poprzednich, może być zastąpiona pojedynczą operacją złożoną, powstałą przez zagnieżdżenie operacji elementarnych.

Jako przykład rozważmy operację selekcji z warunkiem $\text{IdZesp}=10$. Przyjmijmy, że jej wynikiem jest relacja tymczasowa o nazwie PracZesp10 . Następnie na tej relacji wykonujemy operację projekcji atrybutów IdPrac i Nazwisko .

Przyjmijmy, że jej wynikiem jest relacja tymczasowa o nazwie PracZesp10Wynik .

Obie operacje można złożyć w jedną, której wynik będzie identyczny z zawartością relacji PracZesp10Wynik , jak pokazano na slajdzie.



Operacje na zbiorach (1)

- Kompatybilność relacji
 - Dwie relacje: $R(A_1, \dots, A_n)$ i $S(B_1, \dots, B_n)$ są kompatybilne, jeżeli mają ten sam stopień i jeżeli $\text{dom}(A_i) = \text{dom}(B_i)$ dla $1 \leq i \leq n$
- Operacje na zbiorach
 - dla dwóch kompatybilnych relacji: $R(A_1, \dots, A_n)$ i $S(B_1, \dots, B_n)$

W modelu relacyjnym są dostępne operacje na zbiorach o takiej samej semantyce, jak standardowe operacje na zbiorach znane z kursu matematyki. W modelu relacyjnym operacje te są wykonywane na relacjach, które jak wiemy są zbiorami krotek. Relacje te muszą być kompatybilne.

Dwie relacje są kompatybilne jeśli mają ten sam stopień i dziedziny odpowiadających sobie atrybutów są takie same.

Operacje sumy, iloczynu i różnicy dwóch kompatybilnych relacji R i S są zdefiniowane następująco.



Operacje na zbiorach (2)

- Suma:
 - Wynikiem tej operacji, oznaczanej przez $R \cup S$, jest relacja zawierająca wszystkie krotki, które występują w R i wszystkie krotki, które występują w S, z wyłączeniem duplikatów krotek
 - Operacja sumy jest operacją komutatywną: $R \cup S = S \cup R$
- Iloczyn:
 - Wynikiem tej operacji, oznaczonej przez $R \cap S$, jest relacja zawierająca krotki występujące zarówno w R i S
 - Operacja iloczynu jest operacją komutatywną:
 $R \cap S = S \cap R$

BD – wykład 2 (27)

Suma: wynikiem tej operacji, oznaczanej przez $R \text{ SUMA } S$, jest relacja zawierająca wszystkie krotki, które występują w R i wszystkie krotki, które występują w S, z wyłączeniem duplikatów krotek. Operacja sumy jest operacją komutatywną: $R \text{ SUMA } S = S \text{ SUMA } R$.

Iloczyn: wynikiem tej operacji, oznaczonej przez $R \text{ ILOCZYN } S$, jest relacja zawierająca krotki występujące zarówno w R i S. Operacja iloczynu jest operacją komutatywną: $R \text{ ILOCZYN } S = S \text{ ILOCZYN } R$.



Operacje na zbiorach (3)

- Różnica:
 - Wynikiem tej operacji, oznaczonej przez $R-S$, jest relacja zawierająca wszystkie krotki, które występują w R i nie występują w S
 - Operacja różnicy nie jest operacją komutatywną:
 $R - S \neq S - R$

Różnica: wynikiem tej operacji, oznaczonej przez $R-S$, jest relacja zawierająca wszystkie krotki, które występują w R i nie występują w S . Operacja różnicy nie jest operacją komutatywną: $R - S \neq S - R$.



Operacje na zbiorach - przykłady

Uczniowie

Imię	Nazwisko
Ala	Kusiak
Edek	Musiał
Adam	Zając
Olek	Struś
Ola	Buba

Instruktorzy

Imię	Nazwisko
Jan	Kuc
Edek	Musiał
Wacek	Misiek

Uczniowie \cup Instruktorzy

Imię	Nazwisko
Ala	Kusiak
Edek	Musiał
Adam	Zając
Olek	Struś
Ola	Buba
Jan	Kuc
Wacek	Misiek

Uczniowie - Instruktorzy

Imię	Nazwisko
Ala	Kusiak
Adam	Zając
Olek	Struś
Ola	Buba

Uczniowie \cap Instruktorzy

Imię	Nazwisko
Edek	Musiał

Instruktorzy - Uczniowie

Imię	Nazwisko
Jan	Kuc
Wacek	Misiek

BD – wykład 2 (29)

Na slajdzie przedstawiono dwie kompatybilne relacje Uczniowie i Instruktorzy oraz wyniki operacji sumy, iloczynu i różnicy tych relacji.



Operacje na zbiorach - SQL

```
select Imię, Nazwisko  
from Uczniowie  
UNION  
select Imię, Nazwisko  
from Instruktorzy;
```

```
select Imię, Nazwisko  
from Uczniowie  
MINUS  
select Imię, Nazwisko  
from Instruktorzy;
```

```
select Imię, Nazwisko  
from Uczniowie  
INTERSECT  
select Imię, Nazwisko  
from Instruktorzy;
```

```
select Imię, Nazwisko  
from Instruktorzy  
MINUS  
select Imię, Nazwisko  
from Uczniowie;
```

Na slajdzie przedstawiono polecenia zapisane w języku SQL realizujące operacje sumy, iloczynu i różnicy relacji z poprzedniego slajdu.



Iloczyn kartezjański

- Dane są dwie relacje: $R(A_1, \dots, A_n)$ i $S(B_1, \dots, B_m)$
 - Wynikiem iloczynu kartezjańskiego relacji R i S , oznaczonym przez $R \times S$, jest relacja Q stopnia $n+m$ i schemacie: $Q(A_1, \dots, A_n, B_1, \dots, B_m)$
- Krotkom w relacji Q odpowiadają wszystkie kombinacje krotek z relacji R i S
- Jeżeli relacja R ma N krotek, a relacja S ma M krotek, to relacja Q będzie miała $N \cdot M$ krotek

Kolejną operacją modelu relacyjnego jest połączenie. Szczególnym przypadkiem połączenia jest tzw. iloczyn kartezjański, zdefiniowany następująco.

Dane są dwie relacje: $R(A_1, \dots, A_n)$ i $S(B_1, \dots, B_m)$. Wynikiem iloczynu kartezjańskiego relacji R i S , oznaczonym przez $R \times S$, jest relacja Q stopnia $n+m$ i schemacie: $Q(A_1, \dots, A_n, B_1, \dots, B_m)$. Krotkom w relacji Q odpowiadają wszystkie kombinacje krotek z relacji R i S . Jeżeli relacja R ma N krotek, a relacja S ma M krotek, to relacja Q będzie miała $M \cdot N$ krotek. Innymi słowy, iloczyn kartezjański polega na połączeniu każdej krotki z relacji R z każdą krotką relacji S .



Iloczyn kartezjański - przykład

Pracownicy

Imię	Nazwisko
Ala	Kusiak
Edek	Musiał
Adam	Zajęc

Zespoły

Nazwa	Lokalizacja
Reklama	Krucza 10
Badania	Piotrowo 3A

Pracownicy x Zespoły

Imię	Nazwisko	Nazwa	Lokalizacja
Ala	Kusiak	Reklama	Krucza 10
Edek	Musiał	Reklama	Krucza 10
Adam	Zajęc	Reklama	Krucza 10
Ala	Kusiak	Badania	Piotrowo 3A
Edek	Musiał	Badania	Piotrowo 3A
Adam	Zajęc	Badania	Piotrowo 3A

Na slajdzie przedstawiono dwie relacje, tj. Pracownicy i Zespoły oraz wynik iloczynu kartezjańskiego tych relacji.



Operacja połączenia (1)

- Przeznaczenie:
 - łączenie na podstawie warunku połączeniowego wybranych krotek z dwóch relacji w pojedynczą krotkę
- Notacja: operacja połączenia relacji $R(A_1, \dots, A_n)$ i $S(B_1, \dots, B_m)$, jest oznaczona jako:

$$R \bowtie_{\langle \text{warunek połączeniowy} \rangle} S$$

- warunek połączeniowy jest zbiorem predykatów połączonych operatorami logicznymi AND
- predykaty są postaci: $A_i \theta B_j$
 - A_i i B_j są atrybutami połączeniowymi
 - A_i jest atrybutem R , B_j jest atrybutem S
 - $\text{dom}(A_i) = \text{dom}(B_j)$,
 - θ jest operatorem relacyjnym ze zbioru $\{ =, \neq, <, \leq, >, \geq \}$

BD – wykład 2 (33)

Operacja połączenia umożliwia łączenie wybranych krotek z dwóch relacji w pojedynczą krotkę. Krotki są łączone na podstawie podanego warunku połączeniowego.

Notację operacji łączenia relacji R i S przedstawiono na slajdzie. Warunek połączeniowy jest zbiorem predykatów połączonych operatorami logicznymi AND. Predykaty te są postaci: $A_i \text{ THETA } B_j$, gdzie

- A_i i B_j są atrybutami połączeniowymi,
- A_i jest atrybutem R , B_j jest atrybutem S ,
- $\text{dom}(A_i) = \text{dom}(B_j)$,
- THETA jest operatorem relacyjnym ze zbioru $\{ =, \neq, <, >, \leq, \geq \}$.



Operacja połączenia (2)

- Ogólna postać operacji połączenia (theta join)
 - $R \bowtie_{\theta} S$
- Połączenie równościowe (equi join)
 - θ jest operatorem =
- Połączenie nierównościowe (non-equi join)
 - θ jest operatorem różnym od =

Ogólna postać operacji połączenia, gdzie THETA jest dowolnym operatorem relacyjnym jest nazywana połączeniem typu THETA (ang. theta join).

Operacja połączenia, dla której THETA jest operatorem =, nazywana jest połączeniem równościowym (ang. equi join).

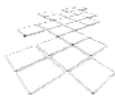
Operacja połączenia, dla której THETA jest operatorem różnym od =, nazywana jest połączeniem nierównościowym (ang. non-equi join).



Operacja połączenia (3)

- Połączenie naturalne (natural join)
 - połączenie równościowe
 - jeden z atrybutów połączeniowych jest usunięty ze schematu relacji wynikowej
 - oznaczane jako: $R * S$
 - atrybuty połączeniowe w obu relacjach muszą mieć taką samą nazwę

Operacja połączenia równościowego, w której jeden z atrybutów połączeniowych jest usunięty ze schematu relacji wynikowej, jest nazywana połączeniem naturalnym (ang. natural join). Połączenie naturalne jest oznaczane jako: $R * S$, przy czym wymagane jest, by atrybuty połączeniowe w obu relacjach miały taką samą nazwę.




Operacja połączenia - przykłady

Pracownicy

IdPrac	Imię	Nazwisko	Szef	IdZesp
100	Jan	Miś		10
110	Piotr	Wilk	100	10
120	Roman	Lis	100	20

Zespoły

IdZesp	Nazwa
10	Reklama
20	Badania

Pracownicy  Szef=IdPrac Pracownicy

IdPrac	Imię	Nazwisko	Szef	IdZesp	IdPrac	Imię	Nazwisko	Szef	IdZesp
110	Piotr	Wilk	100	10	100	Jan	Miś		10
120	Roman	Lis	100	20	100	Jan	Miś		10

Pracownicy * Zespoły

IdPrac	Imię	Nazwisko	Szef	IdZesp	Nazwa
100	Jan	Miś		10	Reklama
110	Piotr	Wilk	100	10	Reklama
120	Roman	Lis	100	20	Badania

Na slajdzie przedstawiono dwie relacje, tj. Pracownicy i Zespoły oraz wynik połączenia równościowego i naturalnego tych relacji.



Operacja połączenia - SQL

```
select *  
from pracownicy p join zespoly z  
on p.id_zesp=z.id_zesp
```

połączenie równościowe
(niestandardowe)

```
select nazwisko, nazwa  
from pracownicy p join zespoly z  
on p.id_zesp=z.id_zesp
```

połączenie równościowe
(standardowe)

połączenie naturalne
(standardowe)

```
select *  
from pracownicy p natural join zespoly z
```

BD – wykład 2 (37)

Na slajdzie przedstawiono polecenia zapisane w języku SQL realizujące operacje połączenia równościowego i naturalnego relacji z poprzedniego slajdu.

Połączenie równościowe zapisano w dwóch postaciach, pierwsza nie jest zgodna ze standardem SQL, ale jest wspierana przez wiele SZBD. Druga notacja jest zgodna ze standardem języka. Połączenie naturalne wyspecyfikowano zgodnie ze standardem.