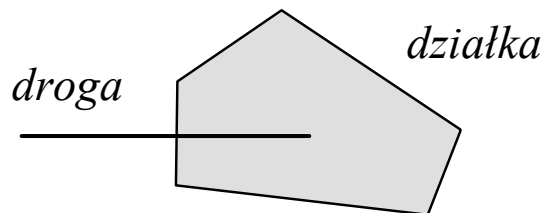


Wielowymiarowe bazy danych

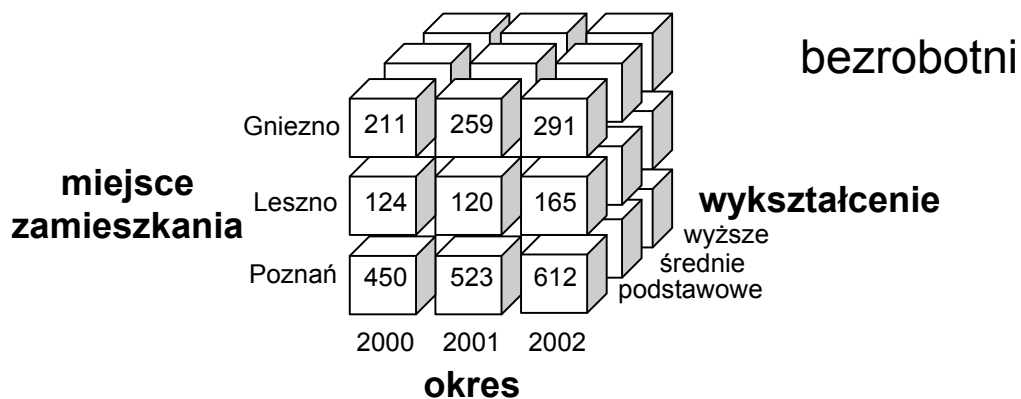
Wielowymiarowe bazy danych

Dziedziny zastosowań

- **Multimedialne bazy danych** – dane medialne przechowywane jako wielowymiarowe wektory danych
- **Systemy geograficzne, systemy wspomaganie projektowania** (VLSI CAD, mechanical CAD), **medyczne bazy danych** – dane w bazie danych mają charakterystykę przestrzenną



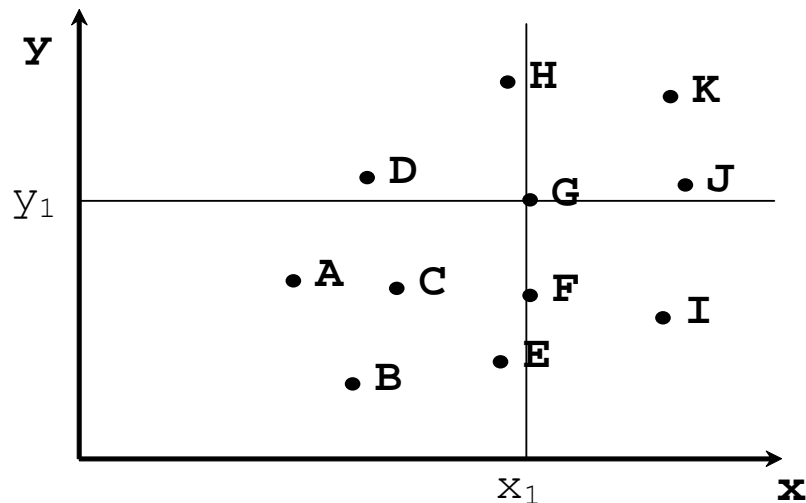
- **Przetwarzanie OLAP** – analiza danych w przestrzeni wielowymiarowej



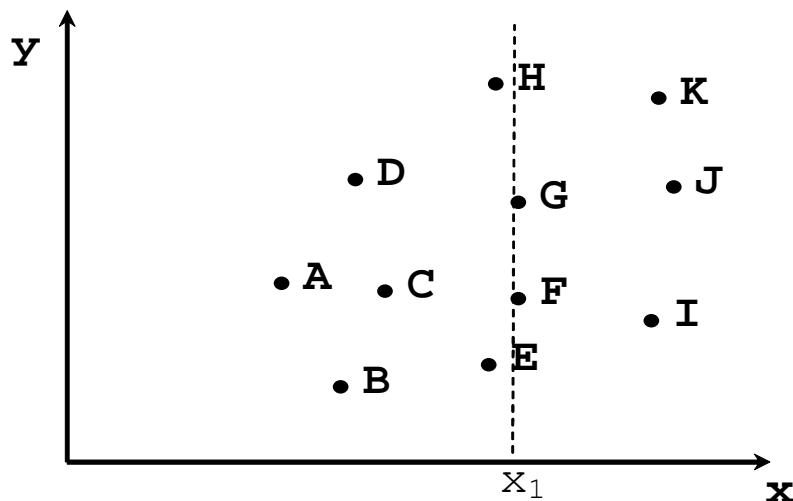
- **Eksploracja danych** – wyznaczanie grup danych charakteryzujących się podobnymi cechami
- **Klasyczne bazy danych** – wydajne wyszukiwanie danych według kilku niezbyt selektywnych atrybutów

Klasy zapytań wielowymiarowych

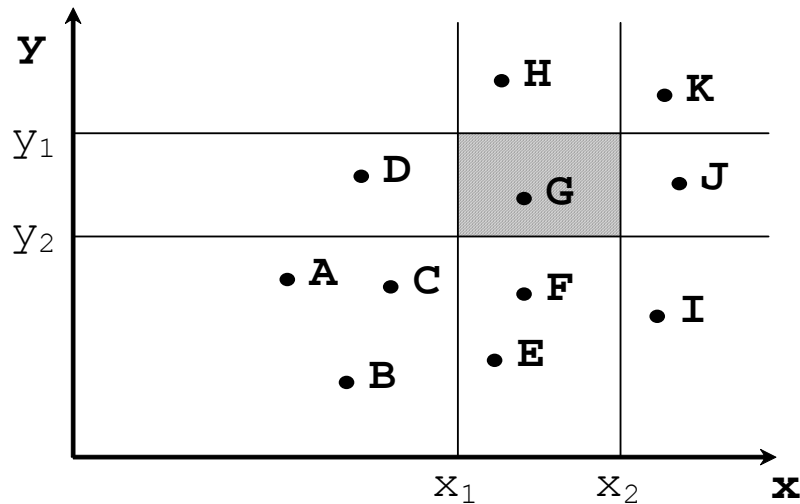
- **Zapytania punktowe** – odwołanie do dokładnych wartości w poszczególnych wymiarach;



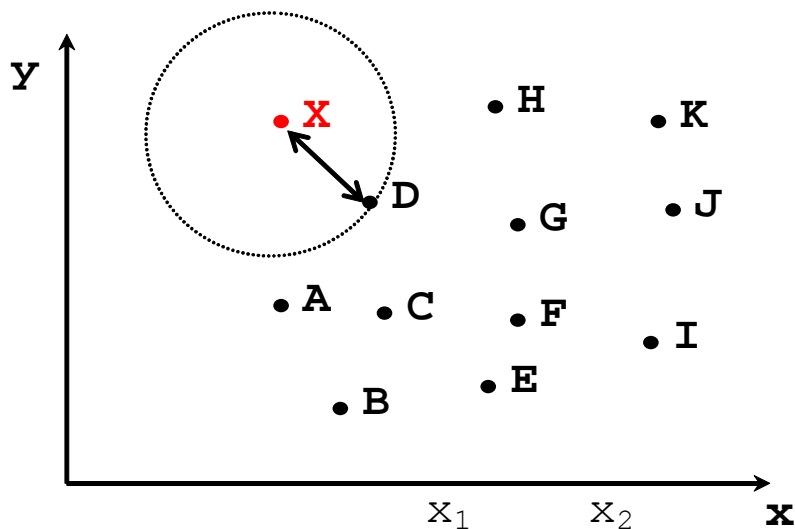
- **Zapytania o częściowym dopasowaniu** – wartości niektórych wymiarów są określone, a innych nie, np. wielowymiarowa analiza danych w magazynach danych;



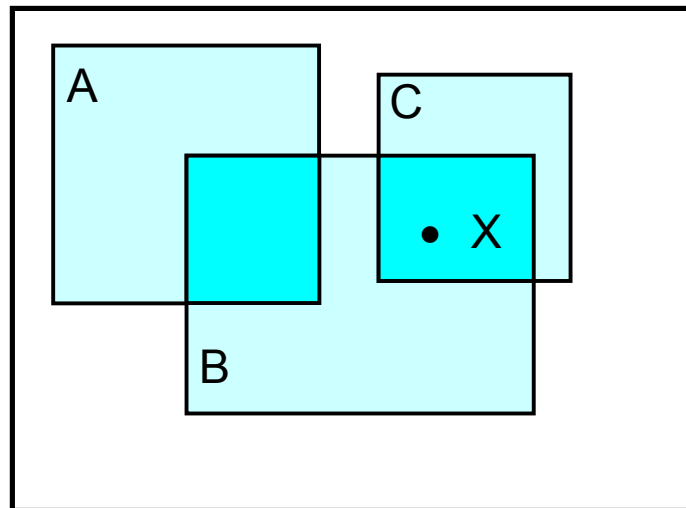
- **Zapytania o zakres** – odwołania do zakresów wartości poszczególnych wymiarów; np. poszukiwanie zbioru figur geometrycznych znajdujących się w pewnym obszarze;



- **Zapytania o najbliższego sąsiada** – szukanie obiektów w przestrzeni wielowymiarowej najbliższych danemu; np. szukanie obiektów geograficznych najbliższych danej lokalizacji;



- **Przestrzenne połączenie** (ang. spatial join) – łączenie obiektów z różnych zbiorów na podstawie wzajemnych odległości. Dla dwóch zbiorów obiektów przestrzennych **A** i **B**, funkcji odległości **L** i danej wartości odległości **r** operacja przestrzennego połączenia wyznacza zbiór par $\{ \langle a, b \rangle \mid a \in A \text{ i } b \in B \text{ takich, że } L(a,b) \leq r \}$.
- **Zapytania typu „gdzie ja jestem”** - poszukiwanie dla danego punktu figur, które go zawierają.

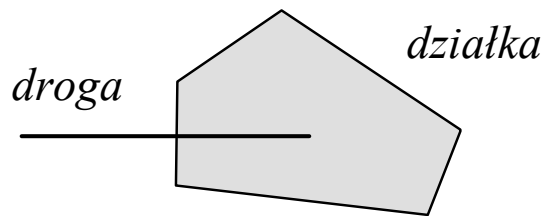


Operatory danych przestrzennych

Operatory dla danych punktowych:

$>$, $<$, $=$, \neq , \geq , \leq

Operatory dla danych przestrzennych:



droga ? działka

- droga dochodzi do działki,
- droga przecina działkę,
- ...

Operatory danych przestrzennych

Każdy obiekt przestrzenny o zawarty w przestrzeni wielowymiarowej Ω ma zdefiniowane:

- wnętrze $i(o)$ – zbiór wszystkich punktów tworzących wnętrze obiektu,
- granicę $b(o)$ – zbiór wszystkich punktów tworzących krawędź obiektu,
- zewnątrz $e(o)$ – zbiór wszystkich punktów znajdujących się na zewnątrz obiektu.

Dla każdego obiektu:

- $i(o)$, $b(o)$, $e(o)$ parami są rozłączne
- $i(o) \cup b(o) \cup e(o) = \Omega$

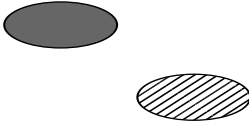
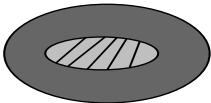
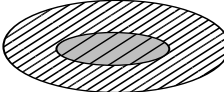

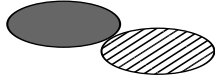

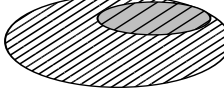
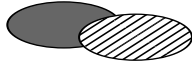
Z definicji:

- dla punktów: $i(o) = \emptyset$
- dla odcinków i łamanych granicami są wierzchołki będące ich końcami, a wnętrze stanowią wszystkie pozostałe ich punkty.

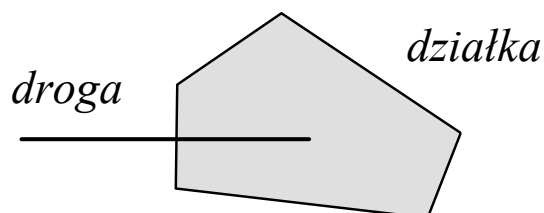
Definicja relacji przestrzennych

Zależności między wnętrzami, granicami i zewnętrżami danych przestrzennych definiują zachodzącą między nimi relację przestrzenną.

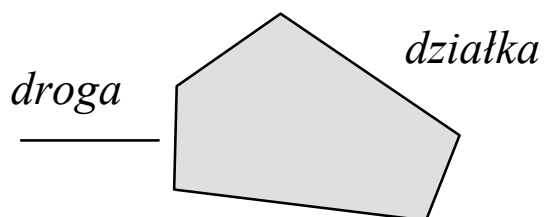
	$I(o_1)$	$B(o_1)$	$E(o_1)$
$I(o_2)$	$I(o_1) \cap I(o_2)$	$B(o_1) \cap I(o_2)$	$E(o_1) \cap I(o_2)$
$B(o_2)$	$I(o_1) \cap B(o_2)$	$B(o_1) \cap B(o_2)$	$E(o_1) \cap B(o_2)$
$E(o_2)$	$I(o_1) \cap E(o_2)$	$B(o_1) \cap E(o_2)$	$E(o_1) \cap E(o_2)$

			
$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ <p>disjoint</p>	$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ <p>contains</p>	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ <p>inside</p>	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ <p>equal</p>
			
$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ <p>meet</p>	$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ <p>covers</p>	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ <p>coveredBy</p>	$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ <p>overlap</p>

Relacje przestrzenne między wielokątem, a linią



	I(dr)	B(dr)	E(dr)
I(dz)	$\neq \emptyset$	$\neq \emptyset$	$\neq \emptyset$
B(dz)	$\neq \emptyset$	\emptyset	$\neq \emptyset$
E(dz)	$\neq \emptyset$	$\neq \emptyset$	$\neq \emptyset$



	I(dr)	B(dr)	E(dr)
I(dz)	\emptyset	\emptyset	$\neq \emptyset$
B(dz)	\emptyset	\emptyset	$\neq \emptyset$
E(dz)	$\neq \emptyset$	$\neq \emptyset$	$\neq \emptyset$

Wielowymiarowe zapytania w SQL

Dana jest baza danych punktów ulokowanych na płaszczyźnie – w wymiarach x i y. Znajdź punkt ulokowany najbliżej danego punktu o współrzędnych $x=100$ i $y=100$.

```
select *
from punkty p
where not exists (
select *
    from punkty q
    where power(q.x-100)+power(q.y-100) <
        power(p.x-100)+power(p.y-100))
```

Dana jest baza danych prostokątów ulokowanych na płaszczyźnie reprezentowanych przez współrzędne dwóch przeciwległych narożników (xlg, ylg, xpd i ypd). Znajdź wszystkie prostokąty zawierające dany punkt o współrzędnych $x=100$ i $y=100$.

```
select *
from prostokaty
where xlg >= 100 and xpd <= 100 and
    ylg >= 100 and ypd <= 100
```

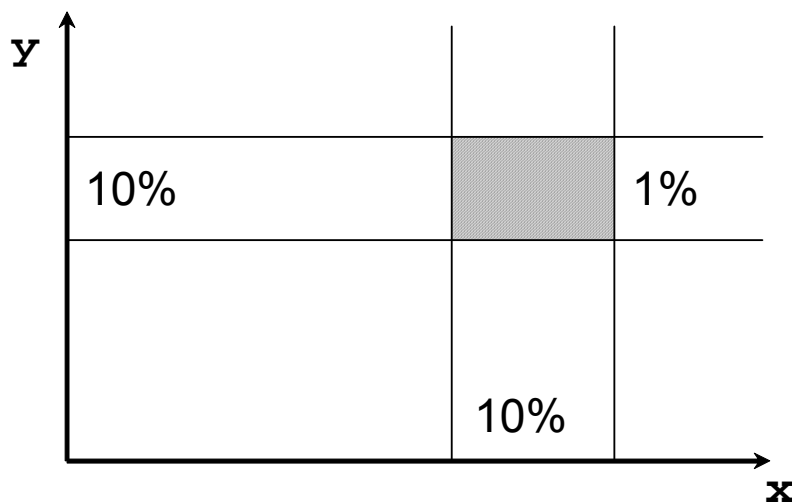
Dana jest baza danych sprzedaży opisanej wymiarami: czasu, lokalizacji i typu asortymentu. Wyznacz sumaryczną sprzedaż artykułów spożywczych w roku 2006 w sklepie „Biedronka”.

```
select sum(cena)
from (sprzedaż natural join sklepy)
    natural join towary
where year(data) = 2006 and
    sklepy.nazwa = 'Biedronka'
    and towary.branża='spożywcza'
```

Ograniczenia klasycznych struktur danych

1. Mała wydajność klasycznych struktur danych:

- selektywne zapytanie o zakres w przestrzeni trójwymiarowej $\sigma=0,001$ w poszczególnych wymiarach jest mało selektywne $\sigma=0,1$.



Przykład:

liczba wymiarów	$n=3$
liczba rekordów	$r=1.000.000$
liczba bloków	$b=100.000$
selektywność atrybutów	$\sigma=0,1$
wysokość indeksów	$h=4$
średnie wypełnienie liści	$q=50$

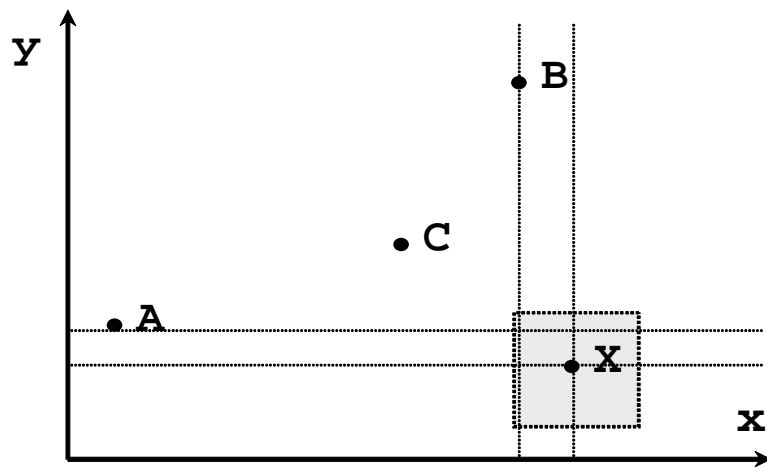
$$\text{koszt} = h + \sigma * r / q + \sigma * r = 4 + 2000 + 100000 = \mathbf{102004}$$

Jednowymiarowe indeksy nie będą używane przez optymalizator zapytań !!! (indeksy bitmapowe)

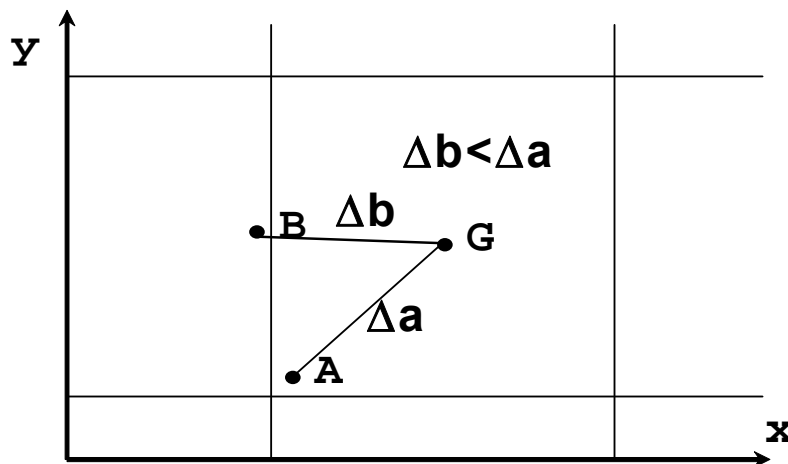
2. Dla indeksów złożonych brak symetrii traktowania poszczególnych wymiarów.

3. Brak obsługi zapytań o relacje przestrzenne, np. znajdź najbliższego sąsiada. Szukaj najbliższych punktów w zadanych zakresach:

- zadany zakres może być pusty;



- punkt znaleziony w danym zakresie może nie być najbliższym położonym punktem.



Wielowymiarowe struktury danych

1. Pliki haszowe

- Pliki kratowe
- Haszowanie podzielone

2. Struktury drzewiaste

- kd-drzewa – wielowymiarowe drzewa binarne
- kB-drzewa – wielowymiarowe B-drzewa
- hB-drzewa – modyfikacja kB-drzew
- Quad-drzewa
- R-drzewa