

**hB-tree**

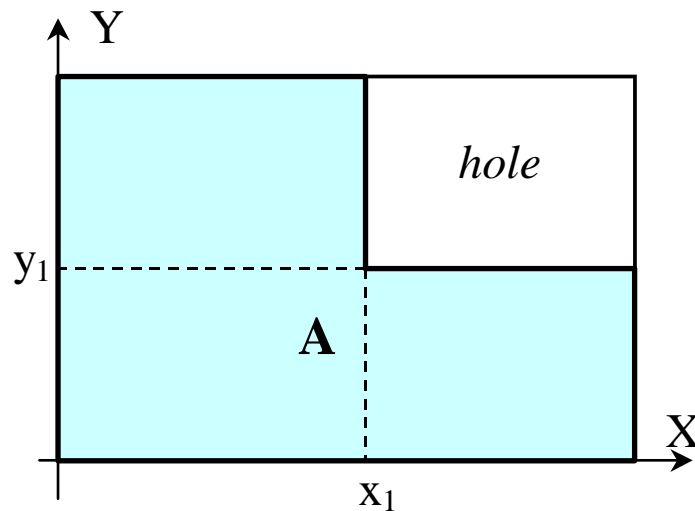
**Holey Brick Tree**

# hB-drzewo - idea

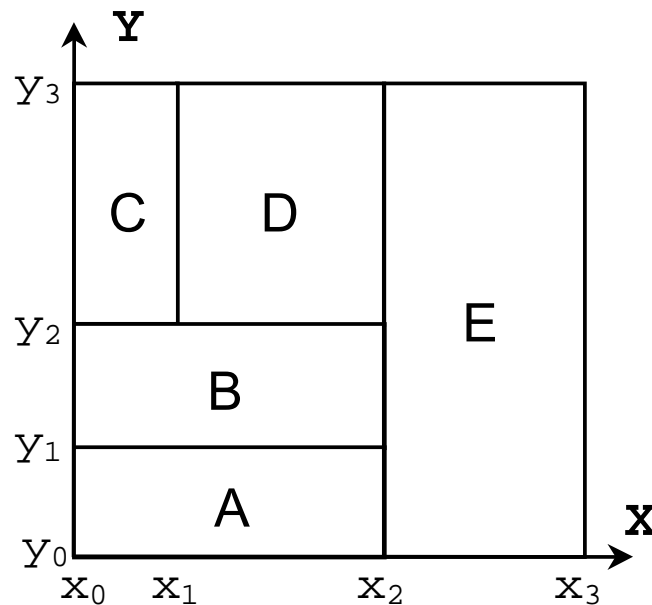
Modyfikacja struktury kdB-drzewa dla uzyskania następujących własności:

- równomierne wypełnienie dzielonych regionów
- uniknięcie dzielenia podregionów
- zwiększenie rzędu drzewa (ang. fan-out) i uniezależnienie go od liczby wymiarów

**Zastąpienie regionów – *regionami nieregularnymi* (ang. holey brick).**



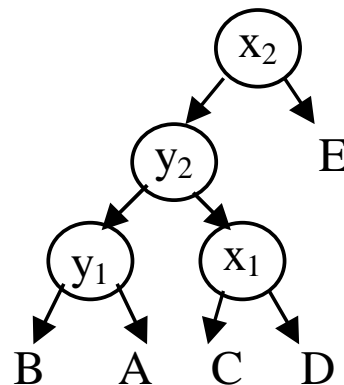
# Struktura węzłów hB-drzewa



Alternatywne reprezentacje regionów w węzłach wewnętrznych indeksu:

a) za pomocą granic regionów b) za pomocą kd-drzew

- A:  $(x_0, x_2), (y_0, y_1)$ ;
- B:  $(x_0, x_2), (y_1, y_2)$ ;
- C:  $(x_0, x_1), (y_2, y_3)$ ;
- D:  $(x_1, x_2), (y_2, y_3)$ ;
- E:  $(x_2, x_3), (y_0, y_3)$ ;



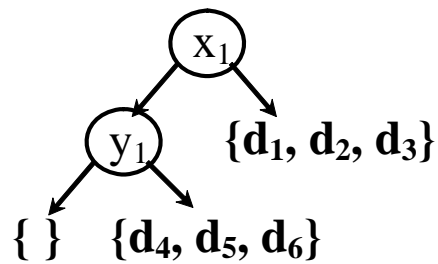
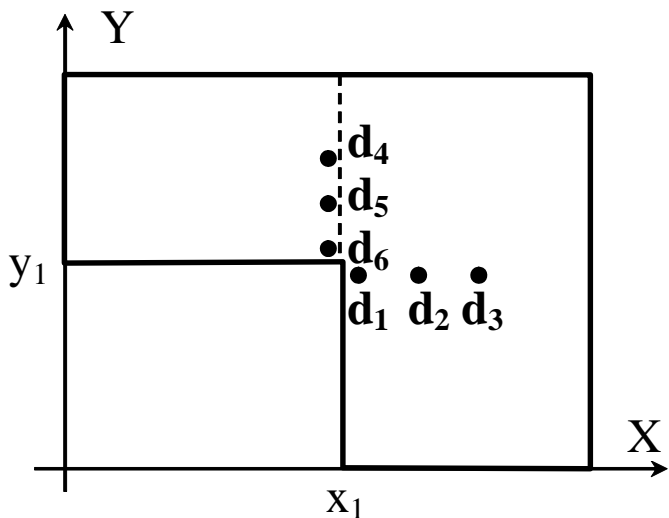
Implementacja węzłów kd-drzewa:

- dziedzina atrybutu,
- wartość atrybutu,
- typ warunku ( $\leq, >$ ) lub ( $<, \geq$ ),
- dwa wskaźniki.

# Reprezentacja nieregularnych regionów hB-drzewa

Do reprezentacji nieregularnych regionów wykorzystano kd-drzewa o zmodyfikowanej topologii grafów acyklicznych skierowanych (DAG).

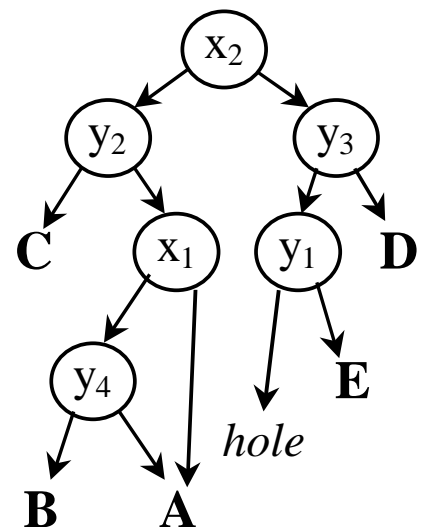
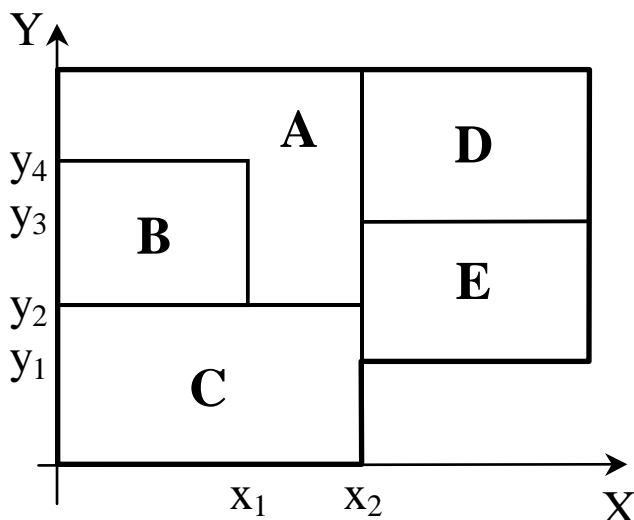
## Struktura liścia:



lokalne kd-drzewo

dane: {(punkt, wskaźnik), ...}

## Struktura węzła wewnętrznego:

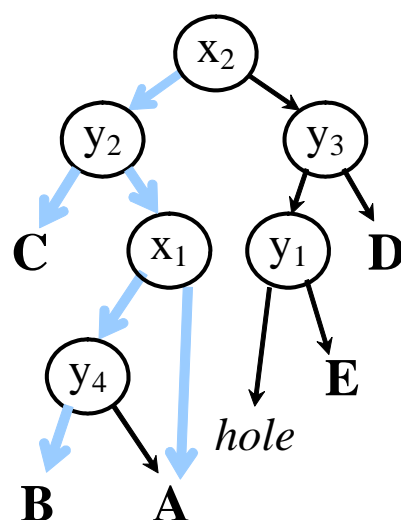
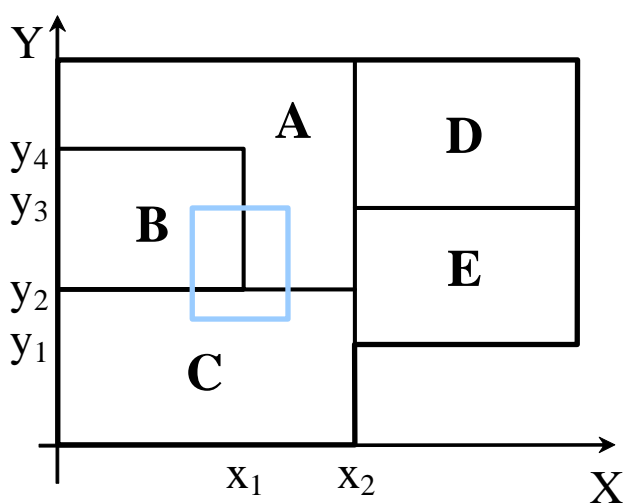


# Operacje na hB-drzewach

## wyszukiwanie danych

Podobne jak w kdB-drzewach; jedyna różnica polega na sposobie wyszukiwania odpowiednich wskaźników wewnątrz węzłów.

- Dla zapytań punktowych wyszukiwanie odbywa się wzdłuż jednej ścieżki prowadzącej od korzenia do liścia (hB-drzew i kd-drzew) indeksu poprzez węzły reprezentujące regiony zawierające dany punkt.
- Dla zapytań nie-punktowych przeszukiwanych może być równoległe wiele ścieżek. Jeżeli wartość danego atrybutu w węźle kd-drzewa jest większa od obydwu granic przedziału zapytania, dalsze poszukiwanie kontynuowane jest w poddrzewie wskazywanym przez lewy wskaźnik węzła. Jeżeli wartość tego atrybutu jest mniejsza od obydwu granic przedziału zapytania dalsze poszukiwanie kontynuowane jest w poddrzewie wskazywanym przez prawy wskaźnik węzła. W pozostałych przypadkach przeszukiwanie kontynuowane jest równoległe w obydwu poddrzewach.

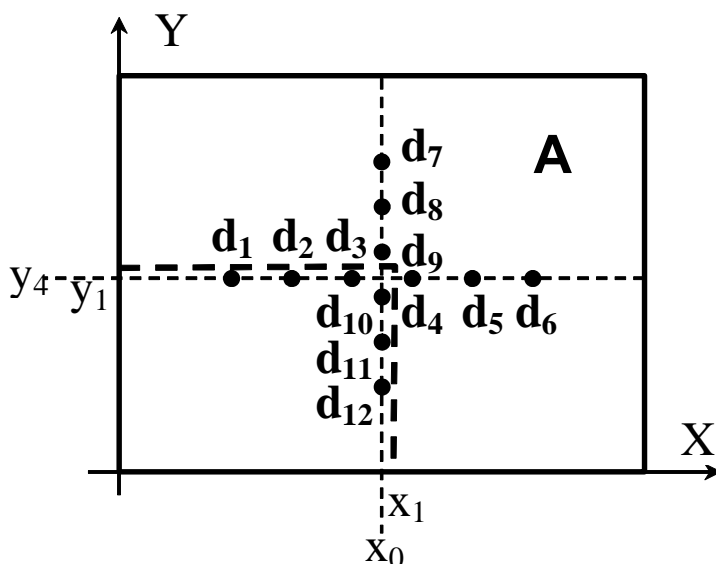


# Ogólny algorytm wstawiania rekordu indeksu

1. Znajdź w hB-drzewie liść indeksu, do którego powinien być wstawiony nowy rekord indeksu.
2. Znajdź w kd-drzewie wewnątrz znalezionej liścia właściwe miejsce dla wstawianego rekordu i spróbuj go tam wstawić. Jeżeli wstawiany rekord zmieścił się w wolnym obszarze liścia, jest to koniec operacji wstawiania. W przeciwnym wypadku dokonaj podziału liścia:
  - 2.1. Utwórz nowy pusty liść i rozdziel równomiernie zawartość przepełnionego liścia powiększoną o wstawiany rekord między stary i nowy liść.
  - 2.2. Prześlij odpowiedni fragment kd-drzewa adresujący obydwa węzły do węzła wyższego poziomu indeksu i wkomponuj go odpowiednio w kd-drzewo tego węzła. Jeżeli w węźle tym jest wolne miejsce - zakończ procedurę.
3. Dokonaj podziału węzła wewnętrznego.
  - 3.1. Utwórz nowy pusty węzeł i rozdziel równomiernie zawartość przepełnionego węzła.
  - 3.2. Prześlij odpowiedni fragment kd-drzewa adresujący obydwa węzły do węzła wyższego poziomu indeksu jeżeli stary węzeł jest korzeniem – utwórz nowy korzeń) i wkomponuj go odpowiednio w kd-drzewo tego węzła. Jeżeli w węźle tym jest wolne miejsce - zakończ procedurę.

# Podział liścia hB-drzewa

Liść przed podziałem:



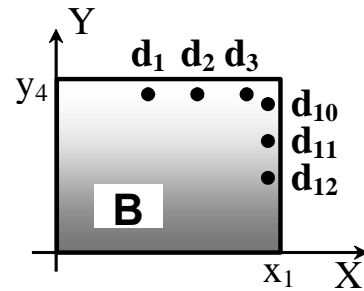
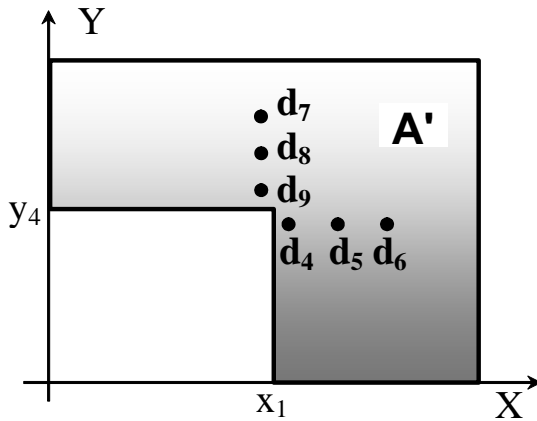
Dzielenie względem jednej z osi  $x_0$  lub  $y_0$  powoduje nierównomierne wypełnienie regularnych regionów; w najlepszym wypadku w proporcji 1:3 ( $\frac{1}{4}$  i  $\frac{3}{4}$ ).

Podział względem więcej niż jednej osi daje znacznie lepsze wyniki. Podział względem osi  $x_1$  i  $y_1$  daje dla powyższego przykładu równomierne wypełnienie regionów w proporcji 1:1 ( $\frac{1}{2}$  i  $\frac{1}{2}$ ).

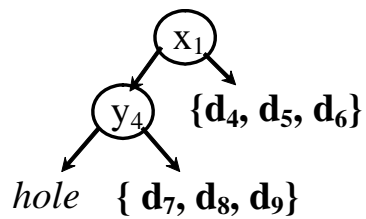
**W najgorszym ogólnym przypadku podział względem  $k$  osi dla przestrzeni  $k$ -wymiarowej daje wypełnienie w proporcji 1:2.**

# Podział liścia hB-drzewa

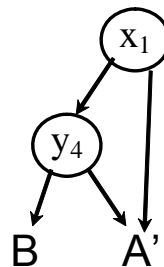
Liść po podziale:



$\{d_1, d_2, d_3, d_{10}, d_{11}, d_{12}\}$



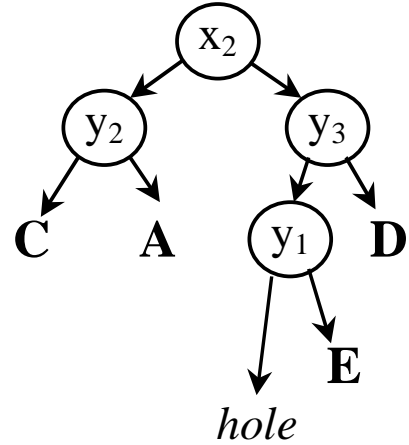
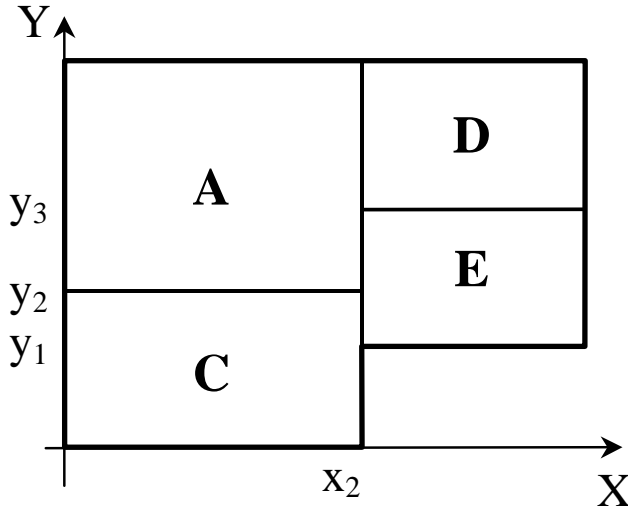
Na wyższy poziom indeksu przekazywane jest do wstawienia kd-drzewo:





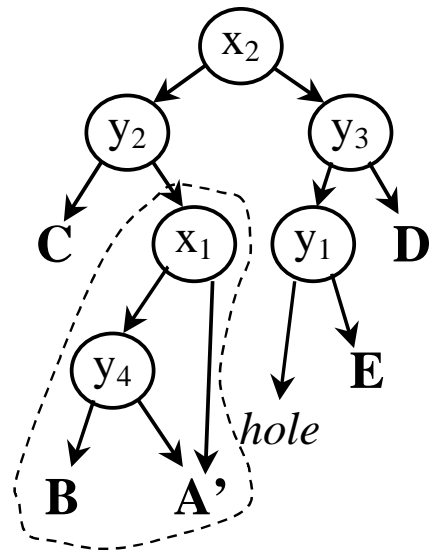
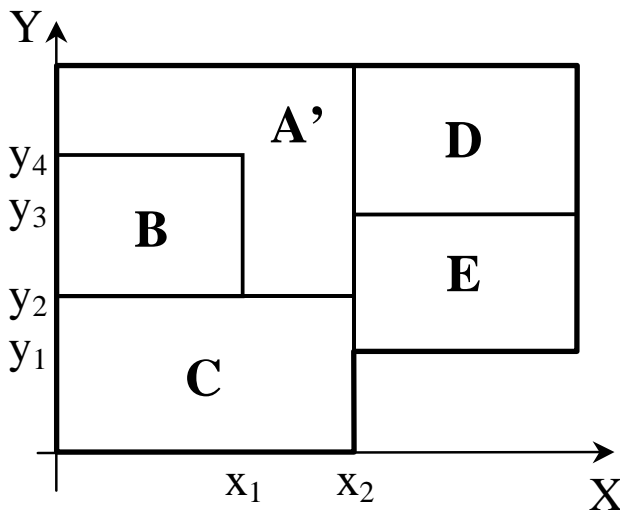
# Modyfikacja węzła wewnętrznego

Stan przed podziałem:



Stan po podziale:

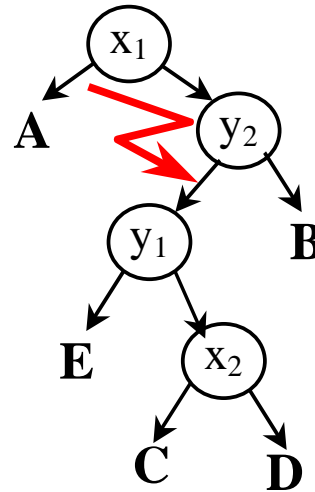
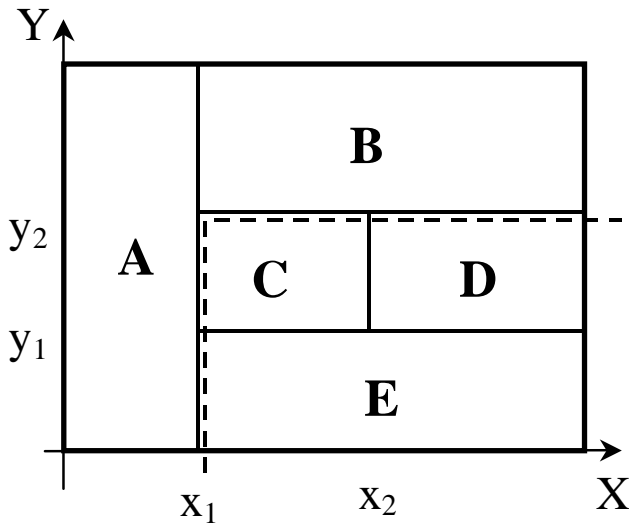
(podział regionu A na regiony A' i B)



W szczególnym przypadku wynikiem podziału węzła może być sytuacja, w której dwa węzły wewnętrzne adresują ten sam węzeł niższego poziomu (kd-DAG).

# Podział węzła wewnętrznego

## Stan przed podziałem



## Stan po podziale

