

## Zadanie 2: Złożone struktury danych

---

### Część programistyczna

Należy zaimplementować następujące dynamiczne struktury danych przechowujące zbiór różnowartościowych liczb naturalnych:

1. Drzewo BST (Binary Search Tree) konstruowane metodą FCFS (kolejne wstawianie elementów),
2. Drzewo AVL konstruowane metodą wykorzystującą bisekcję (rekurencyjny wybór elementu środkowego) oraz jeden z algorytmów sortowania o złożoności logarytmicznej (z Zadania 1),
3. Kopiec minimalny (HMIN) na tablicy jednowymiarowej.

Program powinien umożliwić wykonanie następujących operacji na w/w strukturach danych:

1. Utworzenie struktury z  $n$ -elementowego nieuporządkowanego ciągu danych.
2. Wyszukanie w strukturze elementu o najmniejszej oraz największej wartości:
  - wypisanie ścieżki poszukiwania (od korzenia do ostatniego elementu na ścieżce),
  - podanie liczby wykonanych porównań.
3. Wyszukanie elementu o kluczu podanym przez użytkownika oraz:
  - podanie poziomu, na którym znajduje się węzeł,
  - wypisanie wszystkich elementów znajdujących się na tym samym poziomie,
  - podanie wysokości poddrzewa, którego korzeniem jest dany element,
  - wypisanie w porządku pre-order poddrzewa, którego korzeniem jest dany element.
4. Wypisanie wszystkich elementów struktury w porządku malejącym.
5. Usunięcie elementu o kluczu wskazanym przez użytkownika (pojedynczo lub wraz z jego poddrzewem).
6. Równoważenie drzewa BST jedną z poniższych metod (należy wybrać jedną):
  - algorytm DSW (Day–Stout–Warren),
  - równoważenie przez iteracyjne usuwanie korzenia.

Przed i po zrównoważeniu należy wypisać drzewo w porządku pre-order.

Dane wejściowe:

1. program powinien umożliwić wczytanie z klawiatury  $n$ -elementowego ciągu liczb naturalnych (dla  $n \leq 12$  – tryb demonstracyjny),
2. generowanie danych testowych przy użyciu wbudowanego generatora.

W trybie demonstracyjnym, po wczytaniu danych program powinien sprawdzić czy ciąg jest różnowartościowy; jeśli jest – wyświetli strukturę drzewa w postaci czytelnej reprezentacji tekstowej, w przeciwnym przypadku ponownie poprosić o dane wejściowe.

Dane wyjściowe: dla każdej operacji program powinien podawać czas wykonania operacji oraz liczbę porównań kluczy (jeżeli operacja ich wymaga). Ponadto po utworzeniu struktury program ma wyświetlić wysokość struktury i liczbę jej węzłów. Program powinien działać w trybie menu (w pętli), aż do wyboru opcji zakończenia działania.

### Część eksperymentalna

Generowanie danych: dla każdego rozmiaru  $n$  należy wygenerować po 10 niezależnych różnowartościowych ciągów liczb naturalnych dla każdej z następujących postaci danych: losowe, rosnące, malejące.

Dwa niezależne eksperymenty:

1. Eksperyment 1 – wzrost liniowy. Należy wybrać 10–15 różnych wartości  $n$  rosnących liniowo, np.: 1000, 2000, 3000, 4000, ... Zakres należy dobrać tak, aby można było zaobserwować wzrost czasu obliczeń wraz ze wzrostem  $n$  oraz aby pomiary były możliwe do wykonania w rozsądnym czasie.
2. Eksperyment 2 – wzrost wykładniczy. Należy wybrać 10–15 wartości  $n$  rosnących wykładniczo, np.: 1000, 2000, 4000, 8000, 16000, 32000, ... lub  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ , ... Zakres należy dobrać tak, aby

## Zadanie 2: Złożone struktury danych

---

możliwe było zaobserwowanie charakteru wzrostu złożoności oraz wyraźnych różnic między strukturami dla dużych danych.

Procedura pomiarowa (dla obu eksperymentów): dla każdej struktury, typu danych oraz rozmiaru  $n$ , należy wygenerować 10 niezależnych ciągów danych i zmierzyć czasy tworzenia struktury, wyszukiwania elementu maksymalnego, oraz równoważenia drzewa. Następnie należy obliczyć średni czas każdej z w/w operacji (punkt na wykresie) oraz odchylenie standardowe. Każdy punkt pomiarowy musi być średnią z 10 uruchomień. Wszystkie czasy należy zapisać w tabelach wyników.

### Sprawozdanie

Sprawozdanie powinno zawierać:

1. Wykresy zależności czasu od liczby elementów (według struktury i typu danych). Należy przygotować wykresy funkcji  $t = f(n)$  pokazujące czasy: tworzenia struktury (BST, AVL, HMIN), wyszukiwania elementu maksymalnego (BST, AVL, HMIN) oraz równoważenia (BST), z uwzględnieniem typu danych wejściowych (losowe, rosnące, malejące). Wykresy powinny być oddzielne dla eksperymentu 1 (wzrost liniowy) i eksperymentu 2 (wzrost wykładniczy). Na każdym wykresie należy zaznaczyć odchylenie standardowe (np. w postaci słupków błędów). Podsumowując, należy przygotować następujące wykresy:
  - a. Dla operacji tworzenia struktury należy wygenerować 3 wykresy dla eksperymentu 1 oraz 3 wykresy dla eksperymentu 2. W każdej trójce wykresów powinien być jeden wykres dla danych losowych, jeden dla danych rosnących i jeden dla danych malejących. Na każdym z wykresów należy przedstawić trzy krzywe – po jednej dla każdej struktury (BST, AVL, HMIN).
  - b. Dla operacji wyszukiwania maksimum należy przygotować wykresy w taki sam sposób jak dla operacji tworzenia struktury (6 wykresów).
  - c. Dla operacji równoważenia drzewa BST należy przygotować po jednym wykresie dla każdego eksperymentu (łącznie 2 wykresy). Na każdym wykresie należy przedstawić trzy krzywe odpowiadające różnym typom danych wejściowych (losowe, rosnące, malejące).
2. Analizę porównawczą obu eksperymentów. Należy przeprowadzić bezpośrednie porównanie wyników uzyskanych w eksperymencie 1 i 2 oraz odpowiedzieć na pytania:
  - W którym eksperymencie różnice między strukturami są bardziej widoczne?
  - Jak zmienia się kształt wykresów przy wzroście wykładniczym?
  - Czy AVL utrzymuje stabilną wysokość niezależnie od typu danych?
  - Jak typ danych wejściowych wpływa na czasy operacji dla poszczególnych struktur i dla których struktur ten wpływ jest największy, a dla których najmniejszy? Z czego to wynika?
  - Czy dla danych rosnących i malejących obserwowane są podobne efekty? Jeśli nie, z czego mogą wynikać różnice?
  - Jak wpływ typu danych wejściowych przekłada się na wysokość drzewa i w konsekwencji na koszt operacji?
3. Analizę złożoności obliczeniowej. Dla każdej struktury należy podać złożoność czasową operacji tworzenia struktury i wyszukiwania elementu, a dla BST - równoważenia. Należy porównać wyniki teoretyczne z wynikami eksperymentalnymi i wskazać ewentualne rozbieżności. Czy eksperymentalna złożoność czasowa zależy od typu danych wejściowych? Które przypadki odpowiadają złożoności średniej, a które pesymistycznej (najgorszej)?
4. Wnioski końcowe. W podsumowaniu należy wskazać strukturę najbardziej efektywną w praktyce, określić strukturę najbardziej wrażliwą na postać danych wejściowych, porównać zachowanie drzewa zrównoważonego i niezrównoważonego, ocenić, który sposób skalowania danych (liniowy czy wykładniczy) lepiej ujawnia charakter złożoności asymptotycznej.