264511 Ivan Hancharyk

Piątek 9:15 TN

Algorytmy i Złożoność Obliczeniowa

Sprawozdanie z projektu nr. 1

Temat: "Badanie efektywności wybranych algorytmów sortowania ze względu na złożoność obliczeniową"

1. Wprowadzenie

W ramach niniejszego sprawozdania badamy wydajność trzech podstawowych algorytmów sortowania: sortowania przez wstawianie (insertion sort), sortowania bąbelkowego (bubble sort) oraz sortowania przez scalanie (merge sort). Te algorytmy są fundamentalne w dziedzinie informatyki i często stanowią punkt wyjścia do nauczania technik sortowania ze względu na ich różnorodne zastosowania oraz charakterystyczne własności złożoności. Poniżej przedstawiono krótki opis każdego z algorytmów wraz z ich teoretycznymi oszacowaniami złożoności.

a) Insertion Sort: Cormen opisuje sortowanie przez wstawianie jako metodę, w której elementy wejściowe są konsumowane po jednym za każdym razem, każdy nowy element jest umieszczany w odpowiedniej pozycji wśród wcześniej posortowanych elementów.

Złożoność czasowa

Najlepszy przypadek:

O(n): osiągany, gdy tablica jest już posortowana, co pozwala algorytmowi na zakończenie sprawdzania każdego nowego elementu natychmiast po stwierdzeniu, że jest na właściwym miejscu.

Średni przypadek:

O(n^2): oczekiwany scenariusz dla losowo rozmieszczonych danych, gdzie prawdopodobieństwo znajdowania elementu blisko jego docelowego miejsca jest niskie.

Najgorszy przypadek:

O(n^2): występuje, gdy elementy są ułożone w odwrotnej kolejności, co wymaga przesuwania każdego elementu do końca już posortowanego fragmentu tablicy.

Złożoność pamięciowa

O(1): Insertion Sort jest sortowaniem na miejscu, co oznacza, że nie wymaga dodatkowej pamięci oprócz tej na wejściową tablicę.

b) **Bubble Sort:** Sortowanie bąbelkowe jest jednym z najprostszych algorytmów sortowania, którego główna idea polega na "bąblowaniu" największego niesortowanego elementu na jego odpowiednie miejsce na końcu listy za każdym przejściem przez tablicę. To osiągane jest przez wielokrotne przechodzenie przez listę, porównywanie i zamienianie miejscami sąsiadujących elementów, jeśli nie są w odpowiedniej kolejności.

Złożoność czasowa

Najlepszy przypadek:

O(n): Podobnie jak w Insertion Sort, najlepszy przypadek występuje, gdy tablica jest już posortowana i algorytm może zakończyć działanie po pierwszym przebiegu, jeśli nie wykonano żadnej zamiany.

Średni przypadek:

O(n^2): dla losowo ułożonych danych, gdzie każdy element prawdopodobnie nie jest blisko swojego docelowego miejsca.

Najgorszy przypadek:

O(n^2): występuje, gdy elementy są ułożone w odwrotnej kolejności, co wymaga maksymalnej liczby zamian dla każdego elementu.

Złożoność pamięciowa

O(1): Bubble Sort również jest sortowaniem na miejscu i nie potrzebuje dodatkowej pamięci oprócz przestrzeni na wejściową tablicę.

c) **Merge Sort:** Sortowanie przez scalanie jest przykładem efektywnego algorytmu sortowania wykorzystującego strategię "dziel i zwyciężaj". Ten algorytm dzieli oryginalną tablicę na coraz mniejsze połowy, aż do uzyskania list jednoelementowych, które są naturalnie posortowane. Następnie te małe listy są scalane w większe, posortowane listy, aż cała tablica będzie posortowana.

Złożoność czasowa

Przypadek średni i najlepszy:

O(n log n): Bez względu na początkowy rozkład danych, Merge Sort zawsze dzieli tablicę na dwie równe części i wykonuje rekurencyjne sortowanie scalające. Dzielenie to wymaga logarytmicznej liczby kroków (proporcjonalnie do liczby elementów n), a w każdym kroku, podczas procesu scalania, każdy element jest przetwarzany. Stąd złożoność wynosi **O(n log n).**

Przypadek najgorszy:

O(n log n): Ta sama złożoność obowiązuje również dla najgorszego przypadku, co czyni Merge Sort szczególnie atrakcyjnym dla danych, które muszą być posortowane w przewidywalnym czasie. Nawet jeśli dane są już posortowane lub są w najgorszym możliwym ułożeniu, czas działania Merge Sort nie ulega zmianie.

Złożoność pamięciowa

O(n): Merge Sort wymaga dodatkowej pamięci na przechowywanie tymczasowych tablic używanych podczas procesu scalania. To sprawia, że jest to algorytm "out-of-place", ponieważ nie sortuje danych in-place jak np. quicksort lub bubble sort. Z każdym rekurencyjnym wywołaniem funkcji sortującej, konieczne jest alokowanie nowej pamięci dla podtablic, co prowadzi do złożoności pamięciowej O(n).

2. Plan eksperymentu

Celem tego badania jest analiza i porównanie wydajności wymienionych algorytmów sortowania w zależności od rozmiaru, typu danych oraz ich początkowego rozkładu. Będziemy badać przypadki tablic całkowicie losowych, posortowanych rosnąco, posortowanych malejąco oraz posortowanych częściowo (33/66%), aby ocenić, jak początkowe ułożenie danych wpływa na efektywność sortowania. Wyniki tych

eksperymentów pozwolą nam zrozumieć, w jakich warunkach każdy z algorytmów może być najbardziej przydatny w praktycznych aplikacjach.

Przy pomiarze czasu w programie została wykorzystana biblioteka "chrono.h", która jest częścią standardowej biblioteki języka C++. Biblioteka ta operuje na czasie procesora, co pozwala na uzyskanie dokładnego czasu wykonywania kodu. Zostaje pobierany czas przed i po uruchomieniu potrzebnej nam funkcji, a zatem liczy się różnica czasu bazując na ilośći okresów prebiegu procesora.

Pomiary zostały wykonane dla następnych rozmiarów tablic: {10 000, 20 000, 40 000, 80 000, 120 000, 160 000, 200 000}. Zakres liczb dla testów [-1000000, 1000000].

Po wykonaniu 100 prób sortowania dla każdego algorytmu wyniki zostały uśrednione i w sprawozdaniu zaprezentowane w formacie tabel (dane dla testów generowane są przez klasę generatorRandom.h).

Fragmenty kodu przedstawiające sposób losowania wartości

Klasa "generatorRandom" służy do generowania wektorów z losowymi danymi dla różnych typów danych, takich jak liczby całkowite i zmiennoprzecinkowe, co umożliwia elastyczność w testowaniu algorytmów sortowania na różnorodnych zestawach danych. Użytkownik może określić rozmiar wektora, zakres wartości oraz sposób sortowania danych początkowych, co obejmuje generowanie danych całkowicie losowych, posortowanych rosnąco, malejąco, oraz częściowo posortowanych na poziomie 33% lub 66%. Dzięki zastosowaniu szablonów, "generatorRandom" jest w stanie obsłużyć dowolny typ danych, co czyni ją uniwersalnym narzędziem wspierającym zaawansowane scenariusze testowania algorytmów sortowania w różnych warunkach początkowych.

```
// Metoda startująca timer; zapisuje obecny czas jako punkt startowy.

yvoid Time::startTimer() {
    start = chrono::high_resolution_clock::now();

}

// Metoda zatrzymująca timer; zapisuje obecny czas jako punkt końcowy.

yvoid Time::stopTimer() {
    end = chrono::high_resolution_clock::now();

}

// Metoda obliczająca i zwracająca czas trwania pomiędzy startem a stopem.

// Wynik jest konwertowany z mikrosekund dla większej dokładności.

ydouble Time::getDuration() const {
    return chrono::duration_cast<chrono::microseconds>( d: end - start).count() / 1000.0;

}
```

Klasa "Time.cpp"

Wynik z klasy "Time", który jest zwracany przez metodę "getDuration" w mikrosekundach, pozwala to ocenić efektywność implementacji i porównać wydajność poszczególnych algorytmów w bardzo precyzyjny sposób.

```
// Sortuje dane metoda przez wstawianie.

pvoid Sorter::insertionSort() {
   for (int i = 1; i < data.size(); i++) {
      double key = data[i];
      int j = i - 1;
      while (j >= 0 && data[j] > key) {
          data[j + 1] = data[j];
          j--;
      }
      data[j + 1] = key;
}
```

Implementacja algorytmu przez wstawianie

W metodzie "insertionSort", algorytm przegląda elementy wektora danych, począwszy od drugiego elementu. Dla każdego elementu, który jest rozważany, algorytm porównuje go z elementami wcześniejszymi i przesuwa te wcześniejsze elementy o jedną pozycję w prawo, aż znajdzie odpowiednie miejsce dla rozważanego elementu. Następnie wstawia element w znalezione miejsce. Proces ten jest powtarzany dla każdego elementu wektora, co prowadzi do ułożenia danych w porządku rosnącym.

```
// Sortuje dane metoda babelkowa.

void Sorter::bubbleSort() {
   bool swapped;

   for (int i = 0; i < data.size() - 1; i++) {
      swapped = false;
      for (int j = 0; j < data.size() - i - 1; j++) {
        if (data[j] > data[j + 1]) {
            swap( &: data[j], &: data[j + 1]);
            swapped = true;
        }
    }
   if (!swapped) break;
}
```

Implementacja algorytmu sortowania bąbelkowego

Metoda "bubbleSort" sortuje wektor danych poprzez wielokrotne przeglądanie listy i zamianę miejscami sąsiednich elementów, jeśli występują w nieprawidłowej kolejności. Pętla zewnętrzna iteruje przez listę od początku do końca, za każdym razem zmniejszając zakres sprawdzania o jeden, ponieważ największe elementy "wypływają" na koniec listy w kolejnych iteracjach. Pętla wewnętrzna porównuje kolejne elementy i zamienia je miejscami, jeśli są w niewłaściwej kolejności. Jeżeli podczas jednej z pełnych iteracji pętli nie wykonano żadnej zamiany, algorytm kończy działanie, co oznacza, że lista jest już posortowana.

```
// Metody do realizacji sortowania przez scalanie.
?void Sorter::merge(vector<double>& left, vector<double>& right, vector<double>& bars) {
    int nL = left.size(), nR = right.size();
    int i = 0, j = 0, k = 0;
}
while (j < nL && k < nR) {
    if (left[j] < right[k]) bars[i++] = left[j++];
    else bars[i++] = right[k++];
}
while (j < nL) bars[i++] = left[j++];
while (k < nR) bars[i++] = right[k++];
}

/void Sorter::mergeSortRecursive(vector<double>& arr) {
    int mid = arr.size() / 2;
    vector<double> left(first arr.begin(), last arr.begin() + mid);
    vector<double> right(first arr.begin() + mid, last arr.end());

    mergeSortRecursive( & left);
    mergeSortRecursive( & right);
    merge( & left, & right, & arr);
}

/void Sorter::mergeSort() {
    mergeSortRecursive( & data);
}
```

Implementacja algorytmu przez scalanie

Metoda "mergeSort" inicjuje proces sortowania przez wywołanie rekurencyjnej funkcji "mergeSortRecursive", która dzieli wektor na coraz mniejsze podzbiory aż do osiągnięcia jednoelementowych segmentów. Następnie, w procesie zwanym scalaniem, metoda "merge" łączy te podzbiory w sposób uporządkowany, efektywnie budując posortowany wektor z powrotem.

Podział wektora na dwie części odbywa się w "mergeSortRecursive", gdzie mid jest środkiem wektora. Funkcja tworzy nowe wektory left i right, które reprezentują lewą i prawą połowę oryginalnego wektora. Po rekursywnym posortowaniu tych połówek, następuje scalenie.

Scalanie, realizowane przez metodę "merge", porównuje elementy z obu połówek i umieszcza mniejszy z nich w wynikowym wektorze bars. Proces ten kontynuowany jest aż do wyczerpania elementów w obu wektorach połówek.

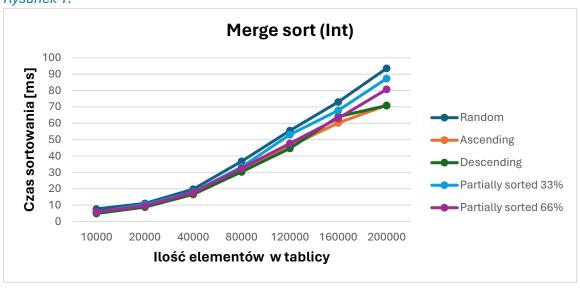
3. Wyniki (w postaci tabel i wykresów)

Merge Sort dla typu danych int

Tabela 1:

Rozmiar Tablicy	Random	Ascending	Descending	Partially sorted 33%	Partially sorted 66%
10000	7,686	5,584	4,924	5,86	5,928
20000	11,112	8,903	8,792	10,17	9,549
40000	19,744	16,472	16,754	18,057	17,712
80000	36,764	30,552	30,366	33,314	32,502
120000	55,522	46,016	44,717	53,084	47,784
160000	73,037	60,265	64,031	67,916	63,108
200000	93,58	70,998	70,826	87,344	80,76



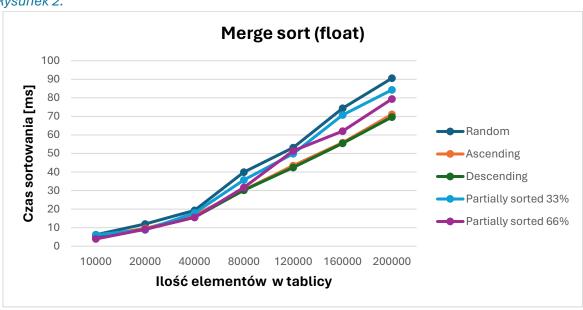


Merge Sort dla typu danych float

Tabela 2:

Rozmiar Tablicy	Random	Ascending	Descending	Partially sorted 33%	Partially sorted 66%
10000	6,151	5,482	5,29	6,183	3,873
20000	11,986	9,525	9,103	8,736	9,057
40000	19,274	16,647	15,581	17,888	15,444
80000	39,904	30,241	30,182	35,713	31,672
120000	53,135	43,426	42,443	49,887	51,41
160000	74,356	55,821	55,464	70,71	62,039
200000	90,575	71,063	69,638	84,327	79,401

Rysunek 2:

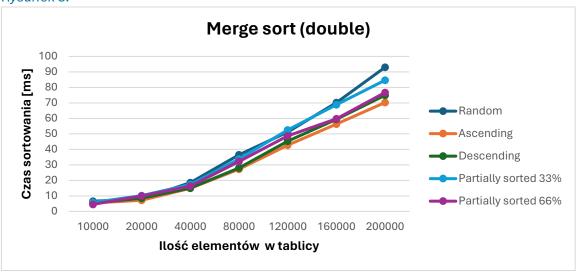


Merge Sort dla typu danych double

Tabela 3:

Rozmiar Tablicy	Random	Ascending	Descending	Partially sorted 33%	Partially sorted 66%
10000	6,564	5,455	5,322	5,762	4,541
20000	8,089	7,195	8,478	10,192	10,016
40000	18,496	14,937	15,047	17,31	16,287
80000	36,466	27,212	27,999	33,834	32,133
120000	51,156	42,662	45,289	52,54	48,74
160000	70,173	56,356	59,491	68,882	59,75
200000	92,984	70,285	74,943	84,647	76,704

Rysunek 3:

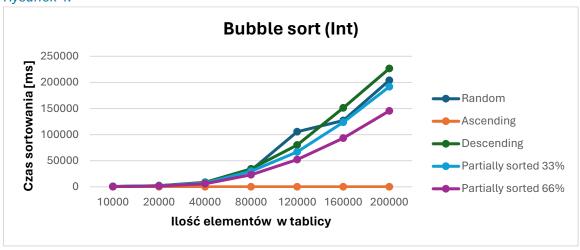


Bubble Sort

Tabela 4:

Rozmiar Tablicy	Random	Ascending	Descending	Partially sorted 33%	Partially sorted 66%
10000	486,395	0,065	528,657	451,037	343,013
20000	1955,485	0,127	2112,233	1776,175	1441,312
40000	8038,938	0,254	8464,424	7338,749	5659,318
80000	31675,208	0,533	34274,353	29655,833	22930,36
120000	105449,091	0,79	80112,455	66946,27	52129,303
160000	126793,92	1,048	151286,611	123683,318	93105,723
200000	203987,522	1,336	226464,515	191721,064	145401,899





Insertion Sort

Tabela 5:

Rozmiar Tablicy	Random	Ascending	Descending	Partially sorted 33%	Partially sorted 66%
10000	99,906	0,09	199,508	89,934	58,049
20000	498,358	0,225	1047,403	477,126	231,685
40000	1801,067	0,294	3880,171	1769,088	1077,541
80000	7618,394	0,43	15674,798	6849,476	4395,082
120000	19432,719	0,943	35270,792	15618,756	10315,601
160000	31842,342	1,609	63604,921	28796,67	18086,848
200000	49905,658	1,984	123913,302	45647,088	28980,548

Rysunek 5:



4. Podsumowanie i wnioski

Efektywność algorytmów w różnych scenariuszach:

Sortowanie przez scalanie wykazuje konsekwentną wydajność niezależnie od porządku i typu danych, co jest zgodne z jego teoretyczną złożonością czasową **O(n log n)**. Jednak wydajność sortowania bąbelkowego znacznie spada wraz ze wzrostem rozmiaru tablicy, co podkreśla jego niepraktyczność dla większych zestawów danych z powodu złożoności **O(n^2)**.

Znaczenie typu danych:

Z danych załączonych w sprawozdaniu nie ma podstaw, aby powiedzieć, że typ danych odgrywa znaczącą rolę w wydajności algorytmów sortowania. **Merge sort** został wybrany w celu oszczędzenia czasu, bo jest szybszy od dwóch innych badanych algorytmów. Chociaż dało nam to możliwość upewnić się w poprawnej implementacji algorytmu w projekcie, ponieważ typ danych nie za bardzo wpływa na jego wydajność.

Co do algorytmów **insertion** i **buble sort**, poza sprawozdaniem przeprowadziłem kilka pomiarów czasu dla tych algorytmów na innych typach danych niż **[Int]**. Pokazało to, że sortowanie liczb całkowitych jest zwykle szybsze w porównaniu z liczbami zmiennoprzecinkowymi i podwójnej precyzji, co jest zgodne z rozumieniem, że mniejsze typy danych wymagają mniejszego obciążenia obliczeniowego.

Stabilność algorytmów:

Stabilność algorytmu, czyli jego zdolność do zachowania względnej kolejności równych elementów, nie jest wyraźnie zaznaczona w danych. Jednak stabilne algorytmy sortowania są kluczowe, gdy pierwotna sekwencja ma pewne istotne znaczenie, na przykład podczas sortowania rekordów według wtórnego klucza.

Wydajność implementacji:

Stosowanie narzędzia <std::chrono::high_resolution_clock> z biblioteki C++ umożliwiło dokładne pomiarowanie czasów sortowania, co jest niezbędne dla wiarygodności przeprowadzonych eksperymentów.

Projekt dał nam wgląd w to, jak optymalizacja kodu może wpływać na wydajność algorytmów, podkreślając wagę pisania efektywnego kodu oraz wykorzystania zaawansowanych funkcji języków programowania do uzyskiwania maksymalnej wydajności.

Potwierdzenie teoretycznych założeń:

Eksperymenty ogólnie potwierdzają teoretyczne oczekiwania dotyczące złożoności różnych algorytmów. Na przykład, sortowanie przez wstawianie pokazuje lepszą wydajność na danych częściowo posortowanych, co jest zgodne z jego najlepszym scenariuszem. Sortowanie przez scalanie konsekwentnie przewyższa sortowanie bąbelkowe dla większych rozmiarów danych, potwierdzając jego złożoność **O(n log n)** wobec złożoności **O(n^2)** sortowania bąbelkowego.

Podsumowanie:

Eksperymenty przeprowadzone w ramach projektu dostarczyły cennych informacji o efektywności różnych algorytmów sortowania, wskazując na ich mocne i słabe strony w zależności od rozmiaru i typu danych, a także ich początkowego uporządkowania. Przyglądając się wynikom, możemy lepiej zrozumieć, w jakich warunkach wybrane algorytmy będą działać najszybciej, a kiedy ich użycie może nie być optymalne. Projekt pokazał także, jak ważne jest stosowanie właściwych narzędzi do pomiaru czasu i optymalizacji kodu, co przekłada się na dokładność i wiarygodność eksperymentalnych ocen wydajności algorytmów. Pozwoliło to na głębsze zrozumienie teoretycznych założeń oraz na ich praktyczną weryfikację. Wiedza ta jest nieoceniona przy projektowaniu nowych rozwiązań oraz optymalizacji istniejących algorytmów sortowania, co może przynieść znaczące korzyści w różnych obszarach informatyki.

5. Literatura

- Cormen T., Leiserson C.E., Rivest R.L., Stein C., Wprowadzenie do algorytmów, WNT
- https://www.geeksforgeeks.org/
- https://pl.wikipedia.org/wiki/Sortowanie_b%C4%85belkowe
- https://en.wikipedia.org/wiki/Merge_sort
- https://en.wikipedia.org/wiki/Insertion_sort

6. Kod źródłowy

```
Main.cpp:
Sorter sorter; // Instancja klasy Sorter do zarządzania danymi i sortowaniem
void menu sorting() {
        displayMenu("--- SORTING MENU ---");
        opt = _getche();
cout << endl:
```

```
sorter.load(generatorRandom<int>::generate(size, min,
max, mode));
                sorter.bubbleSort();
```

```
timer.startTimer();
       cout << endl;
Sorter.cpp:
#include <vector>
void Sorter::display() const {
```

```
void Sorter::bubbleSort() {
   for (int i = 0; i < data.size() - 1; i++) {</pre>
void Sorter::merge(vector<double>& left, vector<double>& right,
void Sorter::mergeSortRecursive(vector<double>& arr) {
       mergeSortRecursive(left);
       mergeSortRecursive(right);
```

```
roid Sorter::mergeSort() {
   mergeSortRecursive(data);
void Sorter::loadFromFile(const string& filename) {
   file.close();
void Sorter::testSort(const string& sortType) {
   Time timer;
   timer.startTimer(); // Rozpoczęcie pomiaru czasu
   if (sortType == "insertion") {
Sorter.h:
#ifndef SORTER H
#define SORTER H
```

```
void insertionSort();
    void bubbleSort();
    void mergeSort();
    void merge(vector<double>& left, vector<double>& right, vector<double>&
    void mergeSortRecursive(vector<double>& arr);
    void loadFromFile(const string &filename);
    void display() const;
generatorRandom.h:
   static std::vector<T> generate(int size, T min, T max, int mode = 0);
    std::vector<T> data(size);
    std::random device rd;
```

```
#endif //AIZO GENERATORRANDOM H
Time.cpp:
double Time::getDuration() const {
Time.h:
#define TIME H
```