Analiza Algorytmów Sortowania

Ksawery Józefowski
Politechnika Wrocławska

November 24, 2024

Contents

1 Wstep

W ramach Listy 1 zaimplementowano i przeanalizowano trzy klasyczne algorytmy sortowania: *Insertion Sort*, *Merge Sort* oraz *Heap Sort*. Każdy z algorytmów został zmodyfikowany: *Insertion Sort* wstawia jednocześnie dwa elementy, *Merge Sort* dzieli dane na trzy cześci zamiast dwóch, a *Heap Sort* korzysta z kopców ternarnych. Celem projektu jest analiza wydajności algorytmów pod wzgledem liczby porównań i przypisań dla różnych rozmiarów danych.

2 Fragmenty Kodów

Poniżej przedstawiono najciekawsze fragmenty kodu dla wybranych algorytmów.

2.1 Insertion Sort z Wstawianiem Dwóch Elementów

Klasyczny algorytm *Insertion Sort* działa poprzez iteracyjne wstawianie kolejnych elementów do uporzadkowanej cześci tablicy, przesuwajac wieksze elementy, aby zrobić miejsce na nowy element. Modyfikacja wprowadzona w tym algorytmie polega na jednoczesnym wstawianiu dwóch kolejnych elementów:

- Dwa kolejne elementy z nieuporzadkowanej cześci sa najpierw porównywane, aby ustalić ich kolejność.
- Nastepnie wiekszy z nich jest wstawiany na pozycje bardziej wysunieta w kierunku końca tablicy.
- Proces wstawiania odbywa sie iteracyjnie, gdzie każdy z dwóch elementów porównywany jest z kolejnymi elementami uporzadkowanej cześci tablicy i wstawiany na odpowiednie miejsce.

Dzieki temu podejściu liczba operacji wstawiania może zostać zmniejszona, gdyż jednocześnie przetwarzane sa dwa elementy.

```
void insertionSortDouble(float arr[], int n, unsigned long
    long& comparisons, unsigned long long& assignments) {
    for (int i = 2; i < n; i += 2) {
        float key1 = arr[i];
        float key2 = arr[i - 1];
        assignments += 2;

        if (++comparisons && key1 < key2) {
            std::swap(key1, key2);
            assignments += 3;
        }

        int j = i - 2;

        while (j >= 0 && ++comparisons && arr[j] > key1) {
            arr[j + 2] = arr[j];
            assignments++;
            j--;
        }

        arr[j + 2] = key1;
        assignments++;

        while (j >= 0 && ++comparisons && arr[j] > key2) {
            arr[j + 1] = arr[j];
            assignments++;
            j--;
        }

        arr[j + 1] = key2;
        assignments++;
    }
    if (n % 2 == 0) {
        int lastKey = arr[n - 1];
        assignments++;
        int j = n - 2;

        while (j >= 0 && ++comparisons && arr[j] > lastKey) {
            arr[j + 1] = arr[j];
            assignments++;
            j--;
        }

        arr[j + 1] = lastKey;
        assignments++;
    }
}
```

Listing 1: Insertion Sort z wstawianiem dwóch elementów

2.2 Merge Sort z Podziałem na Trzy Cześci

Algorytm *Merge Sort* w klasycznej wersji dzieli tablice na dwie cześci, które sa rekurencyjnie sortowane, a nastepnie łaczone w uporzadkowany sposób. Modyfikacja polega na dzieleniu tablicy na trzy cześci, co zmienia strukture rekurencyjnych wywołań:

- Tablica jest dzielona na trzy cześci o zbliżonej liczbie elementów.
- Każda z cześci jest sortowana rekurencyjnie za pomoca *Merge Sort*.
- Po zakończeniu sortowania każda z trzech cześci jest łaczona w jedna, uporzadkowana całość.

Zmiana ta powoduje, że podczas łaczenia musimy operować na trzech posortowanych fragmentach, co zwieksza złożoność procesu scalania, ponieważ liczba operacji porównawczych wzrasta. Z drugiej strony zmniejsza sie liczba poziomów rekurencji, co może przyspieszyć algorytm.

```
void merge3way(float arr[], int left, int mid1, int mid2, int right, unsigned long long& comparisons, unsigned long long& assignments) {
    float temp[right - left + 1];
    int i = left, j = mid1 + 1, k = mid2 + 1, idx = 0;

    while (i <= mid1 && j <= mid2 && k <= right) {
        if (arr[i] <= arr[j] && arr[i] <= arr[k]) temp[idx++] = arr[i++];
        else if (arr[j] <= arr[i] && arr[j] <= arr[k]) temp[idx++] = arr[j++];
        else temp[idx++] = arr[k++];
        comparisons += 2;
        assignments++;
    }

    while (i <= mid1 && j <= mid2) {
        temp[idx++] = arr[i] <= arr[j] ? arr[i++] : arr[j++];
        comparisons++;
        assignments++;
    }
    while (j <= mid2 && k <= right) {
        temp[idx++] = arr[j] <= arr[k] ? arr[j++] : arr[k++];
        comparisons++;
        assignments++;
    }
    while (i <= mid1 && k <= right) {
        temp[idx++] = arr[i] <= arr[k] ? arr[i++] : arr[k++];
        comparisons++;
        assignments++;
    }

    while (i <= mid1) {
        temp[idx++] = arr[i++];
        assignments++;
    }
    while (j <= mid2) {
        temp[idx++] = arr[j++];
        assignments++;
    }
    while (k <= right) {
        temp[idx++] = arr[k++];
        assignments++;
    }

    for (int l = 0; l < idx; ++l) {
        arr[left + l] = temp[l];
        assignments++;
    }
```

```
46 }
47
48 void mergeSort3way(float arr[], int left, int right, unsigned
       long long& comparisons, unsigned long long& assignments)
   {
49     if (left < right) {
50         int third = (right - left) / 3;
51         int mid1 = left + third;
52         int mid2 = right - third;
53
54         mergeSort3way(arr, left, mid1, comparisons,
               assignments);
55         mergeSort3way(arr, mid1 + 1, mid2, comparisons,
               assignments);
56         mergeSort3way(arr, mid2 + 1, right, comparisons,
               assignments);
57
58         merge3way(arr, left, mid1, mid2, right, comparisons,
               assignments);
59     }
60 }
```

Listing 2: Merge Sort z podziałem na trzy cześci

2.3 Heap Sort z Kopcem Ternarnym

Algorytm *Heap Sort* opiera sie na koncepcji kopca binarnego – struktury danych, w której każdy wezeł ma co najwyżej dwóch potomków, a klucz rodzica jest wiekszy lub równy kluczom jego potomków (kopiec maksymalny). Klasyczny *Heap Sort* składa sie z dwóch głównych etapów:

- **Budowanie kopca**: Tablica wejściowa jest przekształcana w kopiec maksymalny.
- **Sortowanie**: Najwiekszy element (szczyt kopca) jest zamieniany z ostatnim elementem tablicy, a kopiec jest ponownie porzadkowany, aby zachować właściwość kopca.

W modyfikacji *Heap Sort* zastosowano kopiec ternarny, w którym każdy wezeł może mieć do trzech potomków. Dzieki tej zmianie zmniejsza sie głebokość kopca, co wpływa na zmniejszenie liczby operacji porównawczych przy wyprowadzaniu najwiekszego elementu na szczyt kopca.

Poniżej przedstawiono fragment kodu implementujacego *Heap Sort* z kopcem ternarnym:

```
void ternaryHeapify(float arr[], int n, int i, unsigned long long& comparisons, unsigned long long& assignments) {
    int largest = i;
    int left = 3 * i + 1;
    int middle = 3 * i + 2;
    int right = 3 * i + 3;

    if (left < n && ++comparisons && arr[left] > arr[largest]) largest = left;
    if (middle < n && ++comparisons && arr[middle] > arr[largest]) largest = middle;
    if (right < n && ++comparisons && arr[right] > arr[largest]) largest = right;

    if (largest != i) {
        std::swap(arr[i], arr[largest]);
        assignments += 3;
        ternaryHeapify(arr, n, largest, comparisons, assignments);
    }
}

void ternaryHeapSort(float arr[], int n, unsigned long long& comparisons, unsigned long long& assignments) {
    for (int i = n / 3; i >= 0; --i) ternaryHeapify(arr, n, i, comparisons, assignments);

    for (int i = n - 1; i >= 0; --i) {
        std::swap(arr[0], arr[i]);
        assignments += 3;  // Przypisania dla zamiany
        ternaryHeapify(arr, i, 0, comparisons, assignments);
    }
}
```

Listing 3: Heap Sort z kopcem ternarnym

3 Analiza i Wyniki

Testy przeprowadzono na losowych tablicach o rozmiarach 10, 10000, 100000 elementów. Dla każdego algorytmu zliczono średnia liczbe porównań, przypisań i czas wykonania.

3.1 Tabela Wyników

Rozmiar danych	Algorytm	Porównania	Przypisania	Czas(s)
10	Insertion Sort	16	34	1.8e-07
10	Insertion Sort (Double)	15	27	9e-08
10	Merge Sort	23	68	2.41e-06
10	Merge Sort (Three-Way)	22	52	3.1e-07
10	Heap Sort	23	78	3.3e-07
10	Heap Sort (Ternary)	40	69	9.3e-07
10000	Insertion Sort	25377684	25397682	0.0454163
10000	Insertion Sort (Double)	12743307	12755796	0.0266696
10000	Merge Sort	120423	267232	0.00202331
10000	Merge Sort (Three-Way)	149314	173914	0.00093419
10000	Heap Sort	166316	372396	0.00142246
10000	Heap Sort (Ternary)	226539	250245	0.00134903
100000	Insertion Sort	2495403337	2495603335	4.485
100000	Insertion Sort (Double)	1246284306	1246409672	2.25296
100000	Merge Sort	1536679	3337856	0.0197074
100000	Merge Sort (Three-Way)	1911782	2175230	0.0120563
100000	Heap Sort	2163214	4725123	0.0175235
100000	Heap Sort (Ternary)	2897570	3136902	0.0172788

Table 1: Porównanie liczby operacji dla różnych rozmiarów danych

3.2 Wykresy Wyników

Poniżej przedstawiono wykresy liczby porównań, przypisań i czasów dla Merge i Heap sorta w zależności od rozmiaru danych.

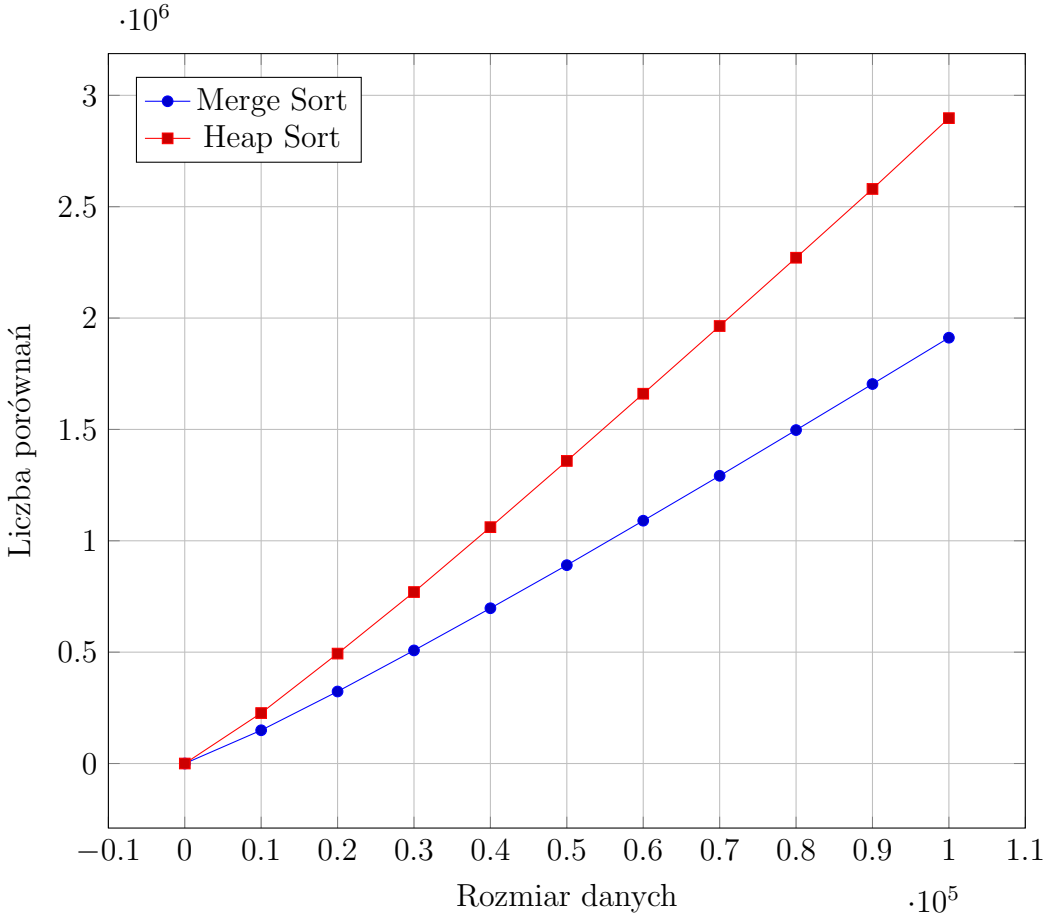

Figure 1: Liczba porównań w zależności od rozmiaru danych

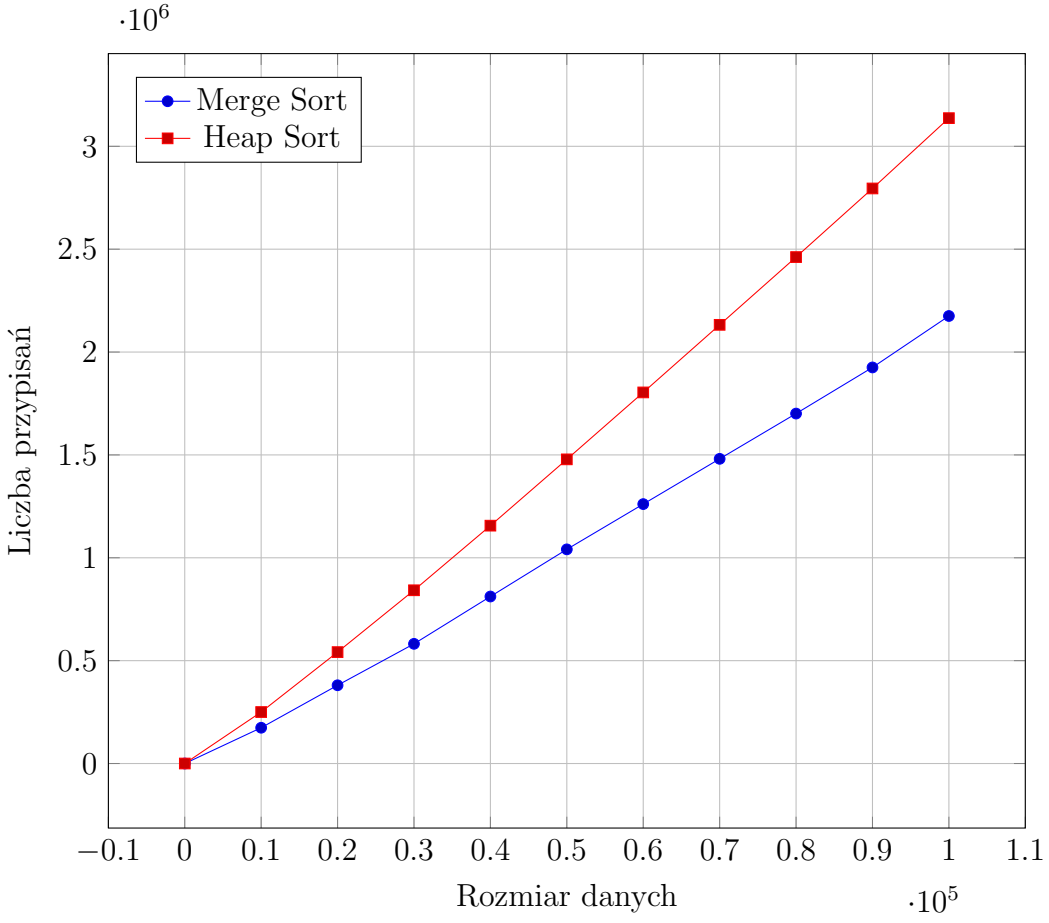

Figure 2: Liczba przypisań w zależności od rozmiaru danych

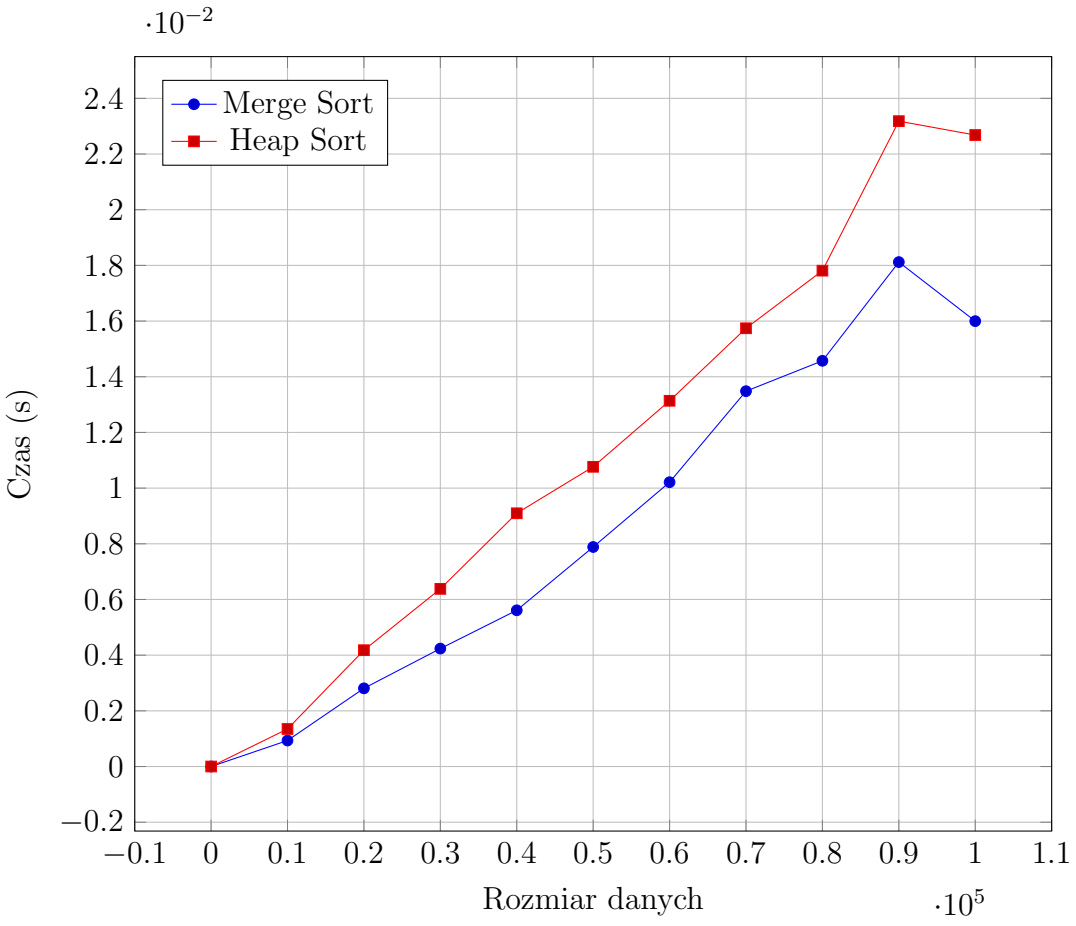

Figure 3: Czas wykonania w zależności od rozmiaru danych

4 Wnioski

Przeprowadzona analiza wykazała, że:

- *Insertion Sort* z wstawianiem dwóch elementów zmniejsza liczbe operacji dla małych zestawów danych, lecz liczba operacji, które wykonuje powoduje, że jest mniej efektywny niż Merge i Heap sort.
- *Merge Sort* z podziałem na trzy cześci zwieksza złożoność algorytmu, co skutkuje wieksza liczba porównań, ale zmniejsza liczbe przypisań w porównaniu do tradycyjnego *Merge Sort*.
- *Heap Sort* z kopcem ternarnym zmniejsza głebokość kopca, co skraca liczbe operacji porównawczych przy dużych zbiorach danych, jednak zwieksza liczbe przypisań.

Z tego wzgledu, klasyczna wersja i zmodyfikowana wersja *Merge Sort* okazały sie najbardziej efektywne przy dużych danych, zachowujac korzystna równowage miedzy liczba porównań a liczba przypisań.