**Analiza eksperymentalna nowych algorytmów sortowania w miejscu**

**Plan prezentacji**

1. Wstęp
2. Motywacja tworzenia algorytmów hybrydowych
3. Metodologia
4. Wady algorytmów podstawowych
5. Rodzina algorytmów Quick Sort
6. QuickMerge Sort
7. Rodzina algorytmów QuickMerge Sort
8. Intro Sort
9. Rodzina algorytmów Intro Sort
10. Rodzina algorytmów QuickMergeIntro Sort
11. Propozycje dalszego rozwoju aplikacji
12. Zakończenie
13. **Wstęp**

W mojej pracy badałem hybrydowe algorytmy sortujące, będące połączeniem wielu technik sortujących.

Algorytmy były analizowane w dwóch grupach

* 1. **Algorytmy deterministyczne** – podczas partycjonowania piwot jest wybierany w sposób deterministyczny
  2. **Algorytmy randomizowane** – piwot jest wybierany w sposób losowy

1. **Motywacja tworzenia algorytmów hybrydowych**

Zasadniczo głównym powodem tworzenia hybrydowych algorytmów sortujących jest poprawa wydajności algorytmów podstawowych. Łączymy algorytmy, aby efektywnie wykorzystać najlepsze części poszczególnych algorytmów oraz niwelować wady każdego z nich. Czasami wprowadzając nieznaczną modyfikację, jak na przykład wybór piwota jako mediana z trzech elementów zyskujemy dużą optymalizację

* 1. **Poprawa wydajności** – minimalna liczba wykonywanych operacji
  2. **Optymalizacja przypadku pesymistycznego** – logarytmiczno-liniowe działanie algorytmu, niezależnie od porządku danych wejściowych
  3. **Działanie w miejscu** – brak potrzeby alokacji dodatkowej pamięci, czas dostępu do pamięci, alokacji i dealokacji może być kosztowny,
  4. **Minimalizacja liczby porównań** – w praktycznych zastosowaniach sortujemy złożone struktury, dla których pojedyncza operacja porównania jest znacznie kosztowniejsza niż np. operacja przypisania

1. **Metodologia**

Ponieważ rzeczywisty czas trwania algorytmu uzależniony jest od maszyny na której przeprowadzane są testy, za podstawowe kryterium porównawcze przyjęto łączną liczbę operacji atomowych wykonywanych na danych testowych. Dodatkowo, ponieważ liczba poszczególnych operacji w dużej mierze zależy od typu sortowanych danych, w trakcie analizy wprowadzono pojęcie współczynnika kosztu, który określa, ile razy operacja porównania jest czasowo dłuższa od operacji przypisania. Ma to sens, jeśli np. sortujemy złożone struktury, oraz operujemy na wskaźnikach do tych struktur. Wtedy przypisanie to podmiana wskaźników, zaś porównanie to dereferencja wszystkich składowych struktury.

1. **Wady algorytmów podstawowych**

Zanim przejdę do analizy algorytmów hybrydowych, chciałbym po krótce przedstawić problemy z jakimi musimy się zmierzyć stosując standardowe algorytmy sortujące.

* 1. **Wady algorytmu Quick Sort**
     1. **O(n2) czas działania dla przypadku pesymistycznego** – dla uporządkowanych danych wejściowych nie sprawdza się technika dziel i zwyciężaj wykorzystywana w algorytmie Quick Sort, głębokość drzewa wywołań rekurencyjnych jest asymptotycznie równa długości tablicy, sam algorytm Quick Sort działa w czasie kwadratowym
     2. **Duża liczba porównań** – algorytm wykonuje prawie 2 razy więcej operacji porównania w stosunku do operacji przypisania, problem w przypadku sortowania złożonych struktur danych
  2. **Wady algorytmu Merge Sort**
     1. **Konieczność posiadania dodatkowej pamięci o wielkości O(n)**– Merge Sort potrzebuje dodatkowego bufora, istnieją implementacje dla których potrzebujemy n/2, nie wszystkie systemy posiadają zasoby aby móc to wykonać
     2. **Dodatkowy nakład czasowy związany z alokacją oraz zwalnianiem pamięci** – operacje alokacji i zwalniania miejsca na stercie są kosztowne

1. **Rodzina algorytmów Quick Sort**

Tak jak wcześniej wspomniałem, czasami stosunkowo niewielka zmiana może znacząco poprawić wydajność algorytmów sortujących. Moją analizę rozpocząłem od badania wydajność różnych wersji algorytmu Quick Sort pod kątem optymalizacji dla przypadku pesymistycznego, z podziałem na metodę partycjonowania oraz strategię wyboru piwota.

* 1. **Partycjonowanie Hoare, losowy piwot** - najwydajniejszą z badanych metod jest partycjonowanie metodą Hoare przy wyborze piwota jako losowy element, dla uporządkowanych danych, czyli przypadku pesymistycznego w przypadku klasycznego Quick Sort, zmodyfikowany algorytm działa w czasie liniowo-logarytmicznym
  2. **Partycjonowanie Hoare, mediana z trzech** – w przypadku algorytmów deterministycznych, najwydajniejsze jest partycjonowanie metodą Hoare oraz wybór piwota jako mediana z trzech elementów, np. pierwszy, środkowy oraz ostatni element tablicy.

1. **QuickMerge Sort**

Kolejnym badanym algorytmem był QuickMerge Sort.

* 1. **Sebastian Wild –** w swojej pracy zatytułowanej **QuickXsort – A Fast Sorting Scheme in Theory and Practice** autor przedstawił koncepcję połączenia

QuickMerge Sort to algorytm Hybrydowy, będący połączeniem algorytmów Quick Sort oraz Merge Sort.

Algorytm rozwiązuje problem z alokacją pamięci dla typowego algorytmu Merge Sort.

Celem tego algorytmu jest minimalizacja wykonywanych operacji porównania, przy równoczesnym zachowaniu działania w miejscu. Algorytm QuickMerge Sort to algorytm rekurencyjny składający się z trzech kroków

1. **Partycjonowanie** – na początku partycjonowana jest tablica wejściowa, przy użyciu metod deterministycznych lub niedeterministycznych
2. **Sortowanie mniejszej części algorytmem Merge Sort** – przy czym bufor stanowi druga część tablicy
3. **Sortowanie drugiej części** – rekurencyjne wywołanie algorytmu dla pozostałej podtablicy

Ważnym elementem tego algorytmu jest modyfikacja algorytmu scalającego podczas sortowania Merge Sort.

Ponieważ bufor do scalania stanowi druga część sortowanej tablicy, musimy zagwarantować że żaden z elementów bufora nie zostanie nadpisany. Aby tego uniknąć zamiast przenosić elementy, zamieniamy je miejscami stosując metodę swap.

1. **Rodzina algorytmów QuickMerge Sort**

Ponieważ składową częścią tego algorytmu jest partycjonowanie ciągu wejściowego, algorytm ten również był analizowany z podziałem na techniki deterministyczne oraz randomizowane.

Z przeprowadzonej analizy wynika, że w przypadku sortowania złożonych struktur danych algorytm QuickMerge Sort jest znacznie wydajniejszy niż klasyczny algorytm Quick Sort. Najwydajniejszą kombinację otrzymano łącząc algorytm QuickMerge Sort z partycjonowaniem Lomuto przy wyborze piwota jako pseudo-mediana z trzech elementów lub mediana z trzech.

1. **Intro Sort**

Kolejnym badanym przeze mnie algorytmem był algorytm Intro Sort, czyli sortowanie introspektywne. Jest to algorytm hybrydowy będący połączeniem algorytmów Quick Sort, Insertion Sort oraz Heap Sort.

Działanie tego algorytmu najlepiej jest to widoczne w pseudokodzie. Algorytm ten dostosowuje się w zależności od rozmiaru sortowanej tablicy oraz głębokości drzewa wywołań rekurencyjnych.

* 1. **Insertion Sort** – jeśli długość tablicy jest mniejsza niż stała wartość wyznaczona eksperymentalnie, stosowany jest algorytm Insertion Sort, w mojej pracy wyznaczyłem tę stałą jako **20**
  2. **Heap Sort** – jeśli drzewo wywołań rekurencyjnych jest głębsze niż **2log(n)**, to rekurencja jest przerywana i stosowany jest algorytm Heap Sort. Dzięki temu, gdy na wejściu są uporządkowanedane, algorytm nie działa z czasem O(n2) lecz z czasem liniowo-logarytmicznym.
  3. **Quick Sort** – w pozostałych przypadkach stosowany jest algorytm Quick Sort

1. **Rodzina algorytmów Intro Sort**

Algorytm był analizowany pod kątem liczby wykonanych operacji w porównaniu do algorytmu Quick Sort.

W każdym z badanych przypadków algorytm Intro Sort okazał się wydajniejszy od algorytmu Quick Sort.

Najefektywniejszym algorytmem sortującym jest Intro Sort z partycjonowaniem metodą Hoare oraz wyborem piwota jako pseudo-mediana z dziewięciu elementów lub mediana z trzech.

1. **Rodzina algorytmów QuickMergeIntro Sort**

Ostatnim badanym przez mnie algorytmem był algorytm QuickMergeIntro Sort. Algorytm ten to próba połączenia algorytmu Intro Sort oraz QuickMerge Sort, aby jeszcze bardziej zoptymalizować liczbę wykonywanych operacji przez te algorytmu. Jest on bardzo podobny do Intro Sort, z podmianą korowego algorytmu sortującego z Quick Sort na Quick Merge Sort. Przeprowadzona analiza sugeruje, że w przypadku sortowania złożonych struktur danych, algorytm ten jest nieznaczenie wydajniejszy od QuickMerge Sort oraz znacznie wydajniejszy od Intro Sort. Wiąże się to z minimalizacją liczby porównań – dla złożonych struktur danych operacja porównania jest bardzo kosztowna, więc nawet niewielka optymalizacja może mieć znaczenie.

1. **Propozycje dalszego rozwoju aplikacji**

Efektem ubocznym mojej pracy było przygotowanie silnika testującego, który automatyzował przeprowadzane testy oraz ich wizualizację. W przyszłości aplikacja ta mogła by być rozwijana równolegle na dwóch płaszczyznach.

* 1. **Algorytmiczna** – wykorzystanie algorytmów wielopiwotowych. W mojej pracy wszystkie algorytmy partycjonujące dzieliły tablicę wejściową na dwie części. Być może bardziej złożony podział, z wykorzystaniem np. algorytmu Dual Pivot Quick Sort autorstwa Vladimira Yaroslavskiego.
  2. **Implementacyjna** – przygotowanie aplikacji webowej do wizualizacji algorytmów w czasie rzeczywistym. Mogło by to być szczególnie pomocne podczas nauki tych algorytmów.

1. **Zakończenie**

Dziękuję za uwagę.