

Projektowanie Algorytmów i Metody Sztucznej Inteligencji

Projekt 2

UWAGA:

Należy wybrać **trzy** algorytmy sortowania (według tabeli). Zaimplementować (**samodzielnie** zgodnie z informacjami omawianymi na wykładzie) wybrane algorytmy oraz przeprowadzić analizę ich efektywności na podanym zbiorze danych.

Ocena	Algorytmy
3	Bąbelkowe, przez wstawianie, przez kopcowanie, przez scalanie
4	przez kopcowanie, przez scalanie, Shella, quicksort
5	przez scalanie, quicksort, introspektywne, kubełkowe

Testy efektywności

Dla danych w pliku projekt2_dane.csv należy wykonać eksperymenty z sortowaniem danych względem rankingu filmów. Załączony plik jest okrojoną bazą filmów “IMDb Largest Review Dataset” z kaggle.com. Plik zawiera tylko tytuł oraz ranking. Proszę o wykonanie następujących zadań:

1. Przefiltrowanie danych i usunięcie pustych wpisów w polu ranking (jeśli występują). Proszę zmierzyć i podać w sprawozdaniu czas przeszukiwania. Czy był on zgodny z oczekiwaną złożonością przeszukiwania dla wybranej struktury danych? Proszę odpowiedzieć w sprawozdaniu.

2. Przygotować strukturę danych zawierającą odpowiednio: **10 000, 100 000, 500 000, 1 000 000, maksymalną ilość danych z pliku.**

3. Przeprowadzić analizę efektywności sortowania na danych z §2 z wykorzystaniem zaimplementowanych algorytmów.

4. Dodatkowo dla **każdego zestawu danych** proszę podać w tabeli **czas sortowania, średnią wartość oraz medianę rankingu.**

Wskazówki:

Należy przeprowadzić wstępną weryfikację poprawności sortowania zaimplementowanych algorytmów !!! Dla małych tablic weryfikacja może być przeprowadzona wizualnie. Dla większych tablic może być przydatna procedura pomocnicza sprawdzająca poprawność uporządkowania elementów w tablicy.

Sprawozdanie

Sprawozdanie powinno zawierać:

- krótkie wprowadzenie,
- ~~opis badanych algorytmów~~ z omówieniem ich złożoności obliczeniowej (dla przypadku średniego i najgorszego)
- omówienie przebiegu eksperymentów i przedstawienie uzyskanych wyników (w postaci tabel i wykresów),
- podsumowanie i wnioski (w przypadku niezgodności uzyskanych wyników z przewidywanymi spróbować wyjaśnić przyczyny),
- Bibliografia (materiały wykorzystane do wykonania ćwiczenia, w tym strony internetowe). Bibliografia powinna być wykonana zgodnie z obowiązującymi zasadami cytowania prac (przykład **tutaj**).

Krótki opis algorytmów

Sortowanie bąbelkowe (bubble sort)

Porównywane są dwa sąsiednie elementy tablicy, na początku ostatni i przedostatni. Jeśli ich kolejność jest niewłaściwa, to zamieniane są one miejscami. Porównanie powtarzane jest dla pozostałych elementów. Po pierwszym przebiegu element najmniejszy znajdzie się na początku tablicy (w części uporządkowanej). W kolejnych krokach procedura jest powtarzana ale tylko dla nieuporządkowanej części tablicy. W ulepszonej wersji algorytmu sortowanie powinno zostać zakończone w sytuacji, gdy po wykonaniu całego przebiegu nie zostanie wykonana ani jedna zamiana.

http://pl.wikipedia.org/wiki/Sortowanie_b%C4%85belkowe

Sortowanie przez wstawianie (insertion sort)

Zakładamy, że początkowa część tablicy jest już posortowana (na początku będzie to pierwszy element tablicy). Pierwszy element z drugiej (nieposortowanej) części tablicy przenoszony jest do zmiennej pomocniczej. Porównywana jest wartość tej zmiennej z kolejnymi elementami pierwszej (posortowanej) części tablicy (zaczynając od ostatniego). Dopóki elementy te są większe, to przesuwane są o jedną pozycję w prawo. Po zakończeniu przesuwania element ze zmiennej pomocniczej wstawiany jest w wolne miejsce. Procedura jest powtarzana dopóki pozostają jeszcze jakieś elementy w drugiej, nieuporządkowanej części tablicy.

http://pl.wikipedia.org/wiki/Sortowanie_przez_wstawianie

Sortowanie przez kopcowanie (heap sort)

Z sortowanej tablicy tworzony jest kopiec binarny. W korzeniu kopca znajdzie się element największy (jeśli przyjmujemy sortowanie w porządku niemalejącym). Element ten zamieniany jest z ostatnim elementem kopca. Elementy przenoszone na koniec tworzą część posortowaną i w kolejnych krokach nie wchodzi już w skład kopca. Pozostały kopiec (bez elementów już posortowanych) jest naprawiany (przywracanie własności kopca). Procedura jest powtarzana dopóki wszystkie elementy nie zostaną posortowane.

http://pl.wikipedia.org/wiki/Sortowanie_przez_kopcowanie

Sortowanie przez scalanie (merge sort)

Sortowana tablica dzielona jest rekurencyjnie na dwie podtablice aż do uzyskania tablic jednoelementowych. Następnie podtablice te są scalane w odpowiedni sposób, dający w rezultacie tablicę posortowaną. Wykorzystana jest tu metoda podziału problemu na mniejsze, łatwiejsze do rozwiązania zadania („dziel i rządź”).

http://pl.wikipedia.org/wiki/Sortowanie_przez_scalanie

Sortowanie Shella (Shellsort)

Sortowana tablica dzielona jest na kilka podtablic utworzonych z elementów oddalonych o ustaloną liczbę pozycji. Tablice te są oddzielnie sortowane (zwykle metodą sortowania przez wstawianie lub bąbelkowego – algorytmy te dobrze sprawdzają się dla danych częściowo posortowanych). W kolejnych krokach wykonywane są podziały z coraz mniejszym odstępem (dające coraz mniejszą liczbę podtablic) aż do uzyskania tylko jednej tablicy. Kluczowym problemem jest dobór odpowiedniej sekwencji odstępów przy podziałach tablicy (można użyć zasady zaproponowanej przez D. Knutha).

http://pl.wikipedia.org/wiki/Sortowanie_Shella

Sortowanie szybkie (quicksort)

Na początku wybierany jest tzw. element osiowy. Następnie tablica dzielona jest na dwie podtablice. Pierwsza z nich zawiera elementy mniejsze od elementu osiowego, druga elementy większe lub równe, element osiowy znajdzie się między nimi. Proces dzielenia powtarzany jest aż do uzyskania tablic jednoelementowych, nie wymagających sortowania. Właściwe sortowanie jest tu jakby ukryte w procesie przygotowania do sortowania. Wybór elementu osiowego wpływa na równomierność podziału na podtablice (najprostszy wariant – wybór pierwszego elementu tablicy – nie sprawdza się w przypadku, gdy tablica jest już prawie uporządkowana).

http://pl.wikipedia.org/wiki/Sortowanie_szybkie

Sortowanie introspektywne (introspective sort lub introsort)

Jest to metoda hybrydowa, będąca połączeniem sortowania szybkiego i sortowania przez kopcowanie. Sortowanie introspektywne pozwala uniknąć najgorszego przypadku dla sortowania szybkiego (nierównomierny podział tablicy w przypadku, gdy jako element osiowy zostanie wybrany element najmniejszy lub największy).

http://pl.wikipedia.org/wiki/Sortowanie_szybkie

Dla wygody podane zostały linki do Wikipedii. Nie należy ich jednak traktować jako jedyne źródła informacji.

Bibliografia

1. Podręcznik z wykładu.
2. Cormen T., Leiserson C.E., Rivest R.L., Stein C., Wprowadzenie do algorytmów, WNT
3. Drozdek A., C++. Algorytmy i struktury danych, Helion