

Fakultet tehničkih nauka  
Univerzitet u Novom Sadu

Računarski sistemi visokih performansi  
**Komparativna analiza algoritama za sortiranje**

Autor:

Vanja Stepanoski

Indeks:

E2 10/2021

Sadržaj

[Uvod 3](#_Toc96900640)

[Algoritmi za sortiranje 4](#_Toc96900641)

[Paralelno računarstvo i zašto je ono bitno 6](#_Toc96900642)

[Paralelizacija algoritama za sortiranje 7](#_Toc96900643)

[Quick sort 7](#_Toc96900644)

[Sekvencijalno rešenje 8](#_Toc96900645)

[Paralelizovano rešenje 8](#_Toc96900646)

[Merge sort 9](#_Toc96900647)

[Sekvencijalno rešenje 10](#_Toc96900648)

[Paralelizovano rešenje 11](#_Toc96900649)

[Counting sort 11](#_Toc96900650)

[Sekvencijalno i paralelno rešenje 11](#_Toc96900651)

[Komparativna analiza rezultata 12](#_Toc96900652)

[Dobijeni rezultati: 13](#_Toc96900653)

[Zaključak 15](#_Toc96900654)

[Literatura 16](#_Toc96900655)

[Spisak slika 17](#_Toc96900656)

# Uvod

Ideja ovog rada jeste da čitaocu kroz primenu nad algoritmima za sortiranje prikaže šta je to paralelizacija i zašto je ona značajna.

U radu će biti reči o tome šta su algoritmi za sortiranje, kao i tome koliko je široka njihova primena. Zatim će biti malo reči i o samoj paralelizaciji, te će se u nastavku rada govoriti o tri algoritma za sortiranje (Quick sort, Merge sort i Counting sort). Ova tri algoritma će biti predstavljena i objašnjena, zajedno sa implementacijom istih, dok će pri samom kraju algoritmi biti testirani i upoređeni nad nizovima različite veličine.

Prilikom testiranja biće korišćen OpenMP.

# Algoritmi za sortiranje

Iako je ideja veoma jednostavna – treba poređati elemente, problem sortiranja je zbog kompleksnosti pronalaska efikasnog rešenja, od samog početka računarskih nauka zauzimao prilično veliki udeo u istraživanjima. Prvi algoritmi za sortiranje datiraju još iz 1950-ih godina, a među autorima istih bila je i Beti Holberton (jedna od šest programera koji su radili na ENIAC-u). Analize Babl (eng. „Bubble“) sorta potiču još iz 1956. godine, štaviše većina optimalnih algoritama je poznata sa sredine 20-og veka. Naravno, još uvek dolazi do otkrivanja novih algoritama za sortiranje, pa tako imamo Timsort algoritam koji je objavljen 2002. godine, ili Library sort algoritamkoji je objavljen 2006. godine [1].

Zapravo, u računarskim naukama, algoritmi za sortiranje predstavljaju jedne od fundamentalnih algoritama, čija je uloga da vrše preraspodelu elemenata određenog skupa podataka tako da nakon izvršavanja algoritma elementi budu poređani u specifičnom poretku. U praksi se najčešće koristi numerički poredak (poredak brojeva po veličini), a odmah nakon njega je i leksikografski poredak (poredak slova, ili karaktera, po abecednom redosledu). Poredak može biti rastući (..., 1, 2, ..., 43, 51, ...) ili opadajući (..., 51, 43, ..., 2, 1, ...).

Efikasno sortiranje je veoma važan, neki bi rekli i ključan, korak u procesu optimizacije ostalih algoritama (npr. algoritam za pretragu) koji pri svom radu, kao jedan od ulaznih parametara, zahtevaju sortiranu listu elemenata. Sortiranje je takođe veoma zastupljeno kada je u pitanju kanonikalizacija podataka radi kreiranja prikaza koji je razumljiviji ljudskom oku.

Da bi se algoritam smatrao algoritmom za sortiranje, podaci koji se dobijaju na izlazu moraju da zadovolje dva kriterijuma:

* Izlazni podaci moraju biti u monotonom poretku (svaki element mora biti veći od prethodnog kada je u pitanju rastući, odnosno svaki element mora biti manji od prethodnog kada je u pitanju opadajući poredak)
* Izlazni podaci moraju biti permutacija (reraspodela, zadržavanje postojećih elemenata, bez ubacivanja novih) ulaznih podataka

Za optimalnu efikasnost najbolje bi bilo u algoritmu za sortiranje koristiti strukture podataka koje dozvoljavaju nasumičan pristup (eng. „random access“) elementima, u odnosu na one strukture podataka koje dozvoljavaju isključivo sekvencijalni pristup.

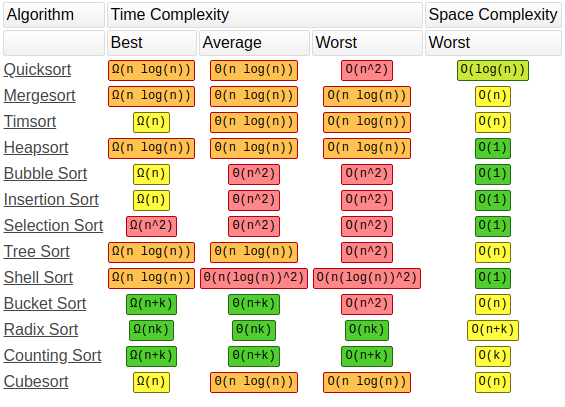
Trenutno postoji na stotine različitih algoritama za sortiranje, svaki sa svojim specifičnim karakteristikama. Oni se mogu razlikovati prema zauzeću memorije – prostorna kompleksnost (eng. „space complexity“) i prema vremenu izvršavanja – vremenska kompleksnost (eng. „time complexity“).

Ove dve vrste kompleksnosti predstavljaju se asimptotskim notacijama (eng. „asimptotic notations“ – više o ovim notacijama se može pronaći u [3]), pretežno korišćenjem simbola O, Θ, Ω („Big-O“, „Theta“ i „Omega“ notacija, respektivno), dok se u zagradama nakon simbola navodi izraz u odnosu na broj *n*, gde broj *n* predstavljabroj koji označava broj elemenata u strukturi podataka (npr. O(n), O(log(n), ...).

Većina algoritama za sortiranje spada u jednu, od sledeće dve, kategorije:

* Logaritamska kompleksnost  
  Kompleksnost proporcijalna binarnom logaritmu (npr. sa osnovom 2) broja *n*. Primer algoritma za sortiranje sa logaritamskom kompleksnošću bio bi Kvik (eng. „Quick“) sort, sa vremenskom i prostornom kompleksnošću:
* Kvadratna kompleksnost  
  Kompleksnost proporcijalna kvadratu broja *n*. Primer algoritma za sortiranje sa kvadratnom kompleksnošču bio bi Babl sort, sa vremenskom kompleksnošću:

Vremenska i prostorna kompleksnost takođe dalje mogu podeljene u tri zasebna slučaja: najbolji slučaj, prosečni slučaj o najgori slučaj (eng. „best case“, „average case“, „worst case“, respektivno) [2].



Slika 1 - Prikaz algoritama za sortiranje sa njihovim kompleksnostima [4]

# Paralelno računarstvo i zašto je ono bitno

Jednostavno rečeno, paralelno računarstvo predstavlja korišćenje više procesora u cilju obavljanja nekog posla. Za razliku od sekvencijalnog računarstva, prilikom primene paralelne arhitekture potrebno je podeliti celokupan posao na manje celine, te identifikovati koje je od manjih celina moguće izvršiti paralelno, te ih pustiti da se paralelno i izvršavaju. Paralelno računarstvo je veoma zgodno kada je u pitanju modelovanje i simulacija svakodnevnih pojava.

Bez primene paralelizacije bi obavljanje većine digitalnog posla bilo prilično nezgodno, pa čak i neizvodljivo ako se u obzir uzme moderni („brzi“) način života. Kada bi naši „pametni“ telefoni ili računari obavljali samo jednu operaciju u trenutnku, svaki posao bi trajao značajno duže nego što danas traje. Da bi se dočarala razlika koju uvodi paralelizacija, dovoljno je spomenuti „pametne“ telefone iz 2010. godine kao što su iPhone 4 i Motorola Droid koji su koristili sekvencijalne procesore, te im je samo za otvaralje e-mail poruke bilo potrebno 30, ili više, sekundi – što je danas nezamislivo dugo.

Prednost paralelnog programiranja je u tome da računari mogu izvršavati kod efikasnije, a neki od benefita koje ono donosi su i [5]:

* Paralelno modelovanje i simulacija pojava iz stvarnog sveta (vreme, saobraćaj, finansije, zdravlje, ...)
* Ušteda vremena
* Ušteda novca
* Rešavanje komleksnijih i većih problema

# Paralelizacija algoritama za sortiranje

Kao što je navedeno, algoritmi za sortiranje predstavljeju jedne od fundamentalnih algoritama, koji svoju primenu nalaze u mnogim aplikacijama. Sama činjenica da je primena algoritama za sortiranje ovoliko široka stavlja do znanja da promena brzine sortiranja može direktno uticati i na brzinu izvršavanja mnogih aplikacija, a paralelizacija nam omogućava da ovo ubrzanje i postignemo.

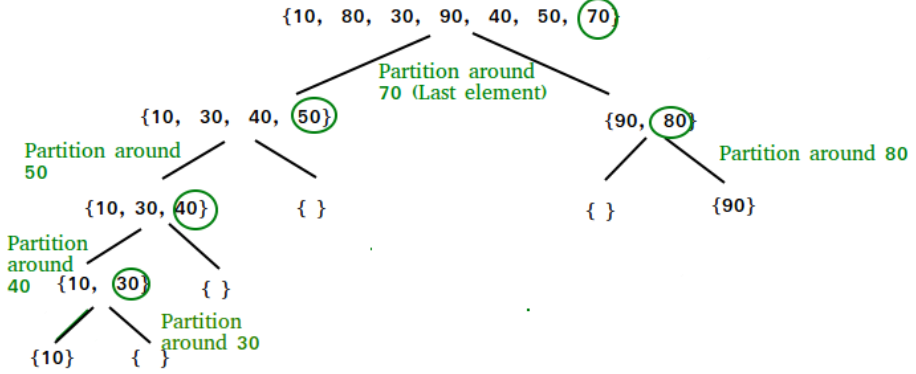
U nastavku rada će biti više reči o Quick, Merge i Counting sortu, te njihovim sekvencijalnim i paralelnim rešenjima.

## Quick sort

Quick sort spada u grupu „podeli pa vladaj (eng. „Divide and Conquer“)[[1]](#footnote-1)“ algoritama za sortiranje. Algoritam funkcioniše tako što odredi element koji će biti „pivot“, i u odnosu na taj pivot element vrši particionisanje niza - deli zadati niz u dva podniza. Postoji više različitih načina kako će Quick sort pronaći pivot element:

* Uvek se uzima prvi element kao pivot
* Uvek se uzima poslednji element kao pivot
* Uvek se uzima nasumični element
* Uvek se uzima medijana niza kao pivot

Ključni process u okviru Quick sort algoritmu je upravo particionisanje niza. Particionisanje niza se vrši tako što funkcija dobije niz i pivot element, te pivot element postavi na odgovarajuću poziciju u odnosu na ostale elemente niza – sve elemente niza koji su manji od pivota prebaci pre njega (levo od pivota), a sve elemente niza koji su veći od pivota prebaci nakon njega (desno od pivota), u slučaju rastućeg poretka, odnosno obrnuto u slučaju opadajućeg poretka.

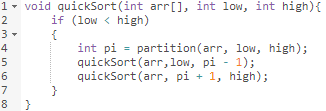


Slika 2 - Prikaz načina rada Quick sort algoritma [8]

Nakon odrađenog particionisanja, funkcija za particionisanje će vratiti novu poziciju pivot elementa, te će ovaj process biti ponovljen po jednom za oba podniza – podniz levo od pivota i podniz desno od pivota.

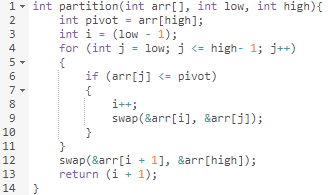
### Sekvencijalno rešenje

U nastavku će biti prikazani isečci implementacije Quick sort algoritma, praćeni kratkim opisima važnih delova.



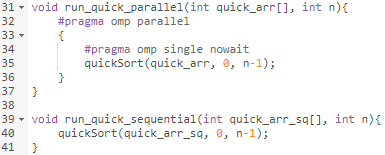
Slika 3 - Prikaz quickSort funkcije

Sa slike XX se može videti princip funkcionisanja Quick sorta, o kome je bilo reči u prethodnom pasusu. Pri ulasku u funkciju quickSort se vrši ispitivanje da li je promenljiva low manja od promenljive high (u slučaju da nije, u podnizu se nalazi samo jedan element, te nema potrebe za dodatnom rekurzijom i može se nastaviti dalje), nakon što je potvrđeno da je promenljiva low manja – tj. da u nizu ima više od jednog elementa, vrši se particionisanje niza i nastavlja se rekurzivno pozivanje funkcije quickSort za oba podniza.



Slika 4 - Prikaz partition funkcije

U funkciji za particionisanje niza se kao pivot element uzima poslednji element, te se nakon toga vrši premeštanje elemenata u odnosu na pivot. Nakon što je premeštanje završeno, funkcija vraća novu poziciju pivot elementa.

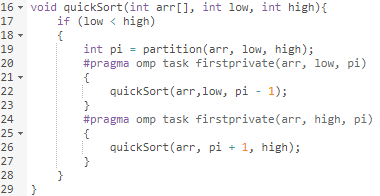


Slika 5 - Prikaz poziva quickSort funkcije

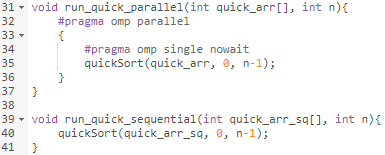
U sekvencijalnom rešenju se funkcija quickSort poziva normalno, sa prosleđenim nizom koji je potrebno sortirati, pozicijom početnog elementa niza, i pozicijom krajnjeg elementa niza.

### Paralelizovano rešenje

Za razliku od sekvencijalnog rešenja, kod paralelnog rešenja se za svaki rekurzivno poziv, koji sledi nakon particionisanja niza, kreira novi omp task i oni se puštaju da rade paralelno. Prilikom paralelizacije, funkcija za particionisanje je ostala nepromenjena.



Slika 6 - Prikaz quickSort funkcije - paralelno

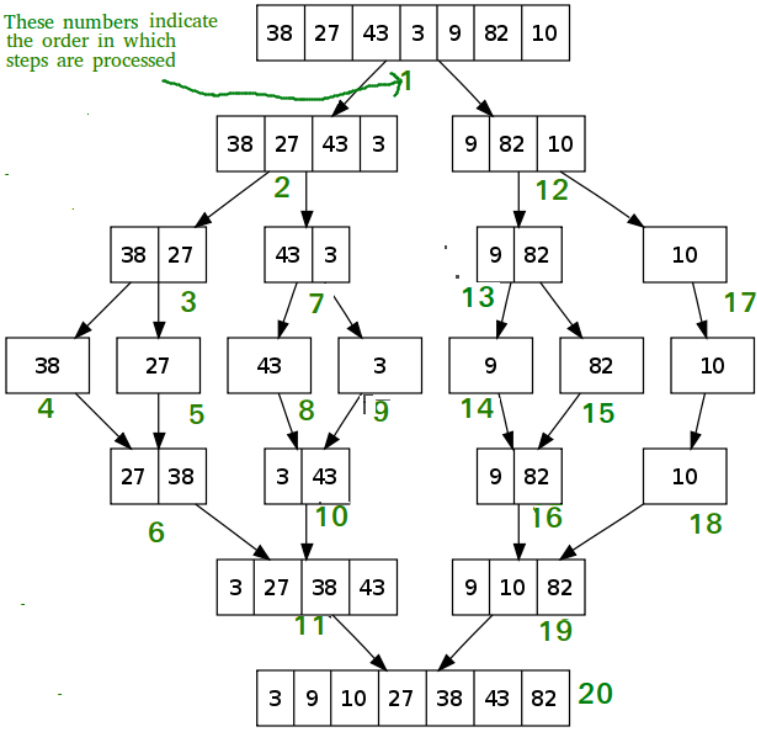


Slika 7 - Prikaz poziva quickSort funkcije - paralelno

U da bi se rešenje izvršilo paralelno, potrebno je početnu quickSort funkciju pozvati unutar paralelnog bloka, u suprotnom neće doći do paralelizacije, već će se sve izvršiti sekvencijalno.

## Merge sort

Kao i Quick sort, i Merge sort spada u grupu „podeli pa vladaj“ algoritama za sortiranje. Ovaj algoritam deli početni niz na dva podniza, te poziva sam sebe za svaki podniz, nakon toga vrši spajanje ovih podnizova, tako što svaki element postavlja sortiran u odnosu na poredak koji se traži.

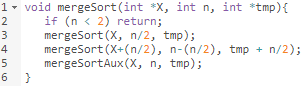


Slika 8 - Prikaz načina rada Merge sort algoritma [10]

Sa prikazanog dijagrama se može uočiti način na koji će niz [38, 27, 43, 3, 9, 82, 10] biti sortiran kroz 20 koraka (sekvencijalno rešenje).

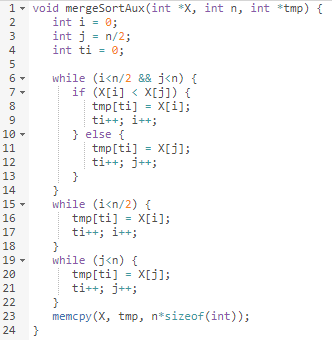
### Sekvencijalno rešenje

U nastavku će biti prikazani isečci implementacije Merge sort algoritma, praćeni kratkim opisima važnih delova.



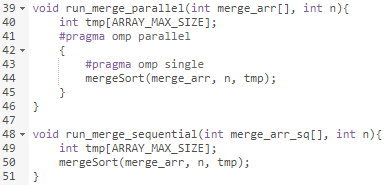
Slika 9 - Prikaz mergeSort funkcije

Funkcija mergeSort vrši podelu niza i rekurzivne pozive sve dok u nizu ne ostane samo jedan element, nakon toga počinje vraćanje i izvršavanje glavne funkcije – mergeSortAux.



Slika 10 - Prikaz mergeSortAux funkcije

Funkcija mergeSortAux vrši spajanje dva podniza, tako da novodobijeni niz bude sortiran u traženom poretku.

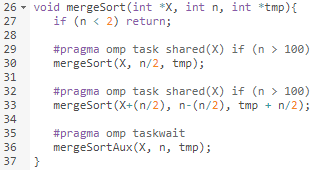


Slika 11 - Prikaz poziva mergeSort funkcije

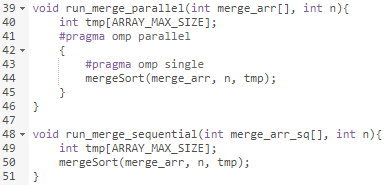
Prilikom poziva funkcije mergeSort, potrebno je proslediti niz koji se sortira, broj elemenata u njemu, kao i pomoćni niz koji će koristiti funkcija mergeSortAux.

### Paralelizovano rešenje

U slučaju paralelnog rešenja, za svaki od rekurzivnih poziva metode mergeSort se kreira poseban task koji će se izvršavati u paraleli. Nakon što je glavni niz podeljen na male podnizove, pristupa se spajanju u funkciji mergeSortAux. Funkcija mergeSortAux se označava kao taskwait, te predstavlja tačku sinhronizacije za taskove koji se izvršavaju unutar funkcije mergeSort. Metoda mergeSortAux je prilikom paralelizacije ostala nepromenjena.



Slika 12 - Prikaz mergeSort funkcije - paralelno



Slika 13 - Prikaz poziva mergeSort funkcije - paralelno

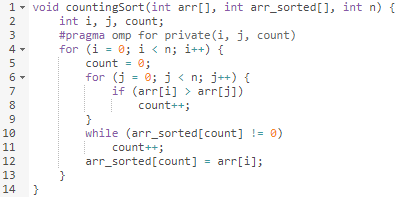
U da bi se rešenje izvršilo paralelno, kao što je slučaj i kod Quick sorta, potrebno je početnu quickSort funkciju pozvati unutar paralelnog bloka, u suprotnom neće doći do paralelizacije, već će se sve izvršiti sekvencijalno.

## Counting sort

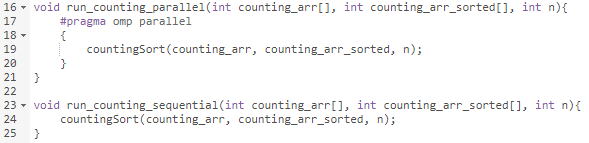
Counting sort predstavlja algoritam za sortiranje baziran na korišćenju ključeva koji označavaju pozicije i broj elemenata koji se nalaze počev od te pozicije. Funkcioniše tako što broji pojavljivanje svakog elementa unutar niza, te nakon toga primenjuje određenu artitmetiku da bi odredio pozicije na kojima se elementi nalaze.

### Sekvencijalno i paralelno rešenje

Prilikom pokušaja implementacije sekvencijalnog i paralelnog rešenja došlo je do delimičnog uspeha, naime ovakva implementacija Counting sort algoritma ne daje svaki put ispravno sortiran niz, već dolazi do grešaka prilikom sortiranja.



Slika 14 - Prikaz countingSort funkcije



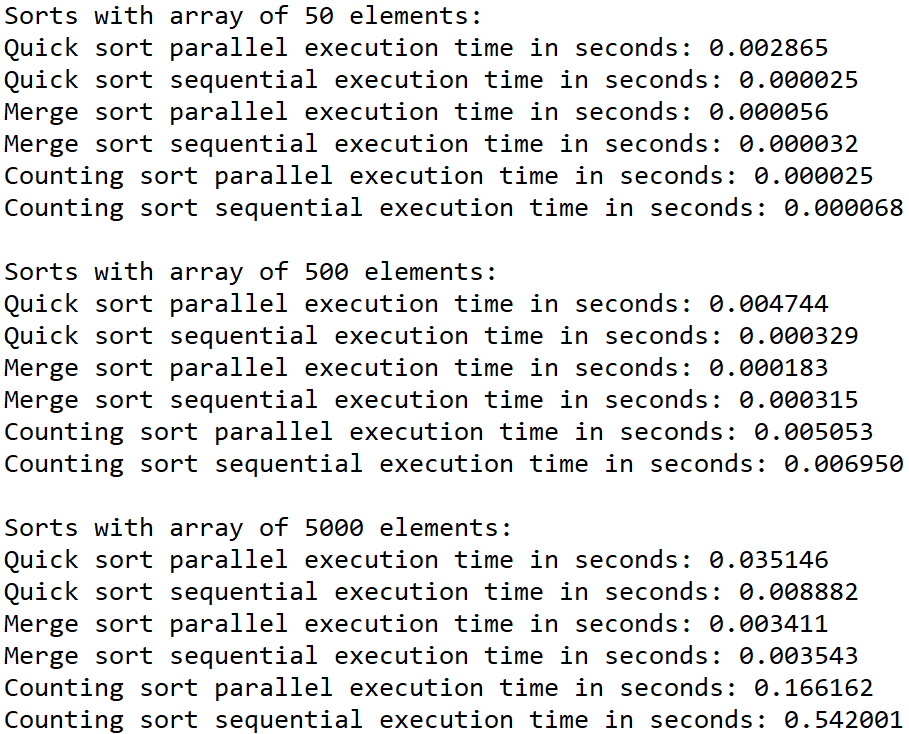
Slika 15 - Prikaz poziva countingSort funkcije

## Komparativna analiza rezultata

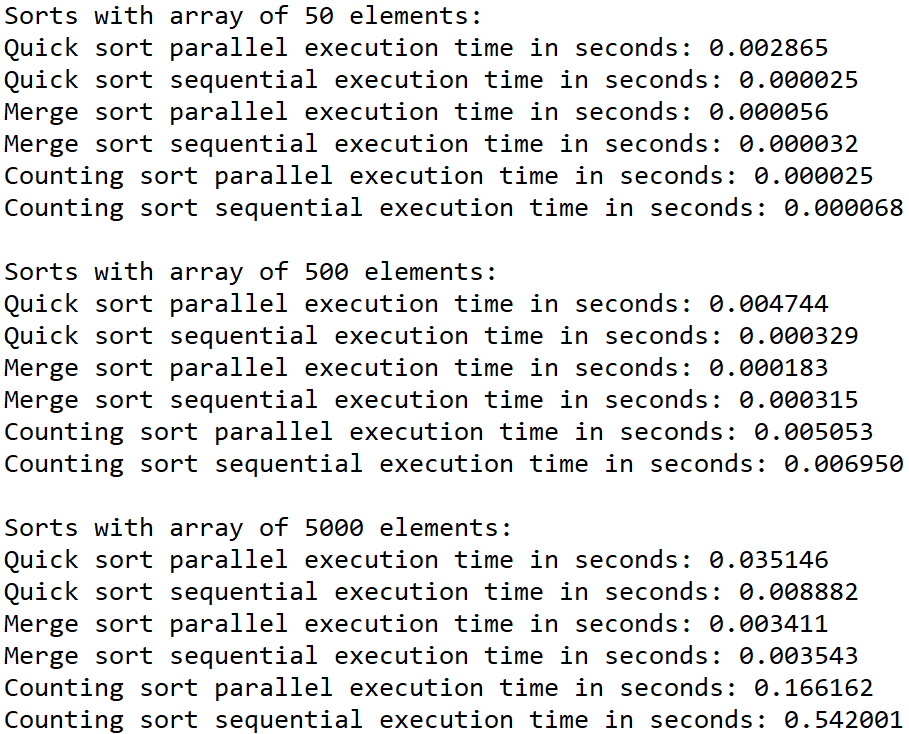
Prilikom pokretanja i testiranja gorenavedenih implementacija korišćen je računar koji ima Intel Core i7-7600HQ 2.60GHz procesor sa 4 fizička, odnosno 8 logičkih jezgara i 16GB ram memorije. Svaki od algoritama je dobijao identičan nesortirani niz sa N elemenata, gde je svaki element nasumično izabran broj u rasponu 1-50. Za testiranje je korišćeno N za svaki broj iz [ 50, 500, 5000, 50000, 500000, 1000000, 2000000, 2500000]. Redosled pokretanja algoritama je pri svakom pokretanju bio sledeći:

1. Quick sort – paralelno
2. Quick sort – sekvencijalno
3. Merge sort – paralelno
4. Merge sort – sekvencijalno
5. Counting sort – paralelno
6. Counting sort – sekvencijalno

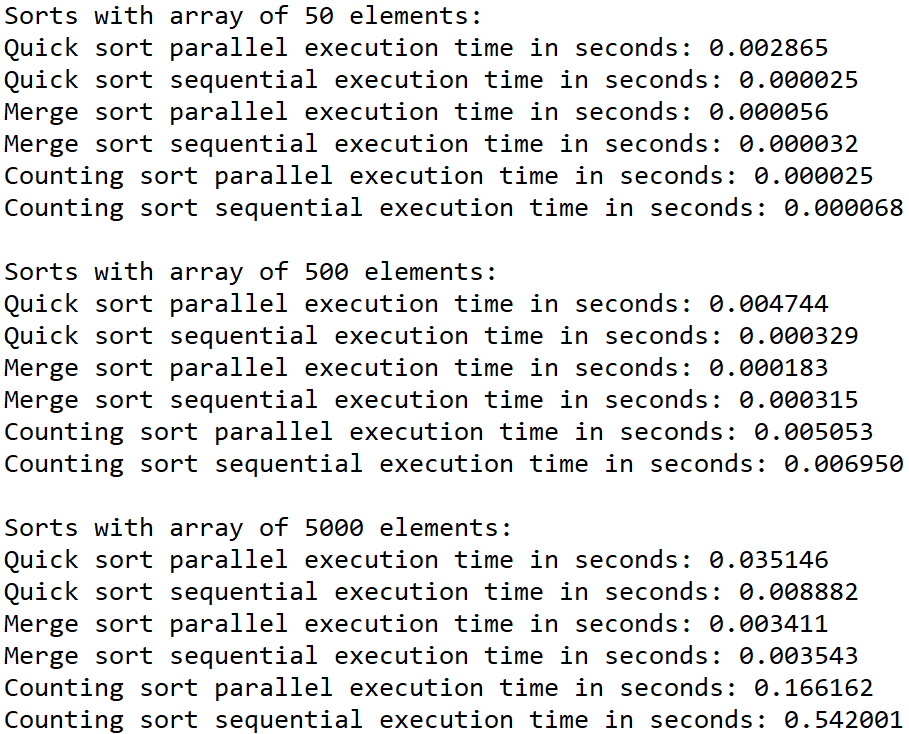
### Dobijeni rezultati:



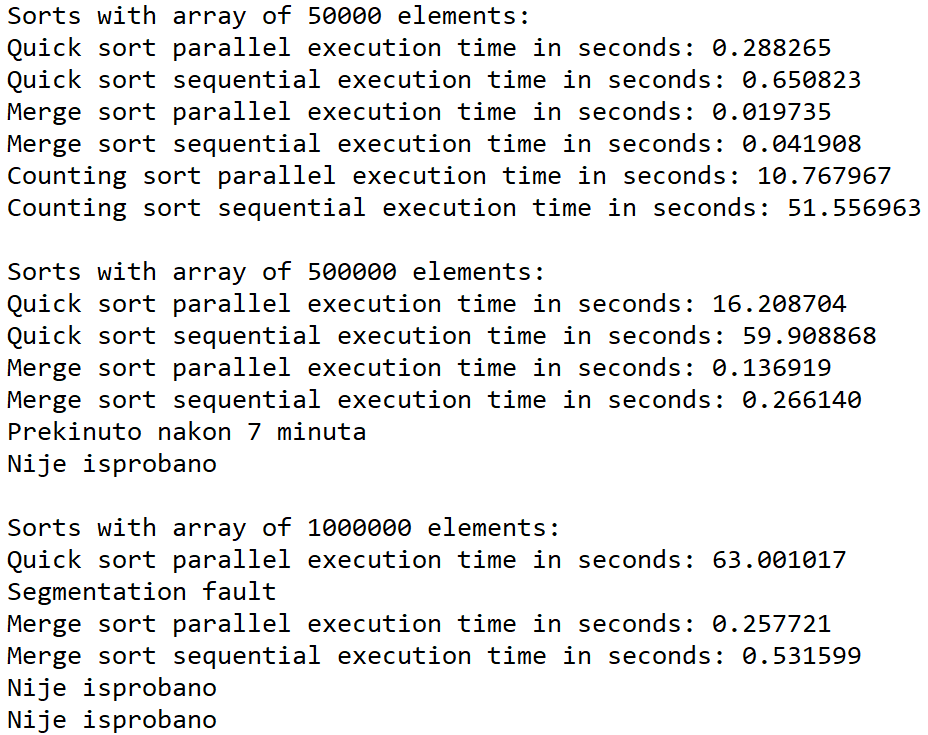
Slika 16 - Test sa N = 50



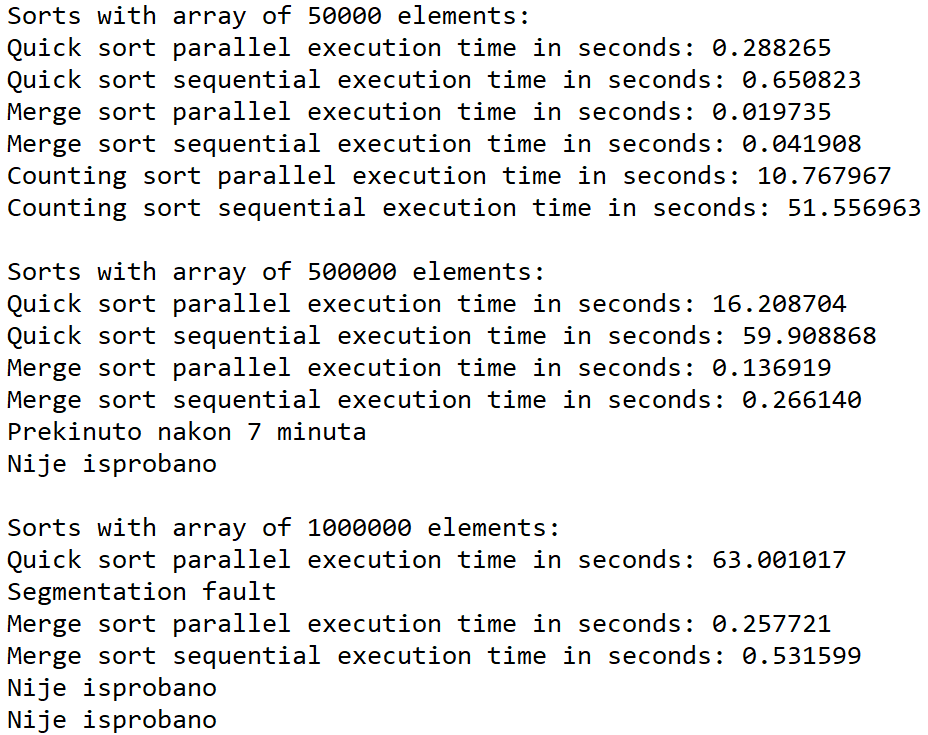
Slika 17 - Test sa N = 500



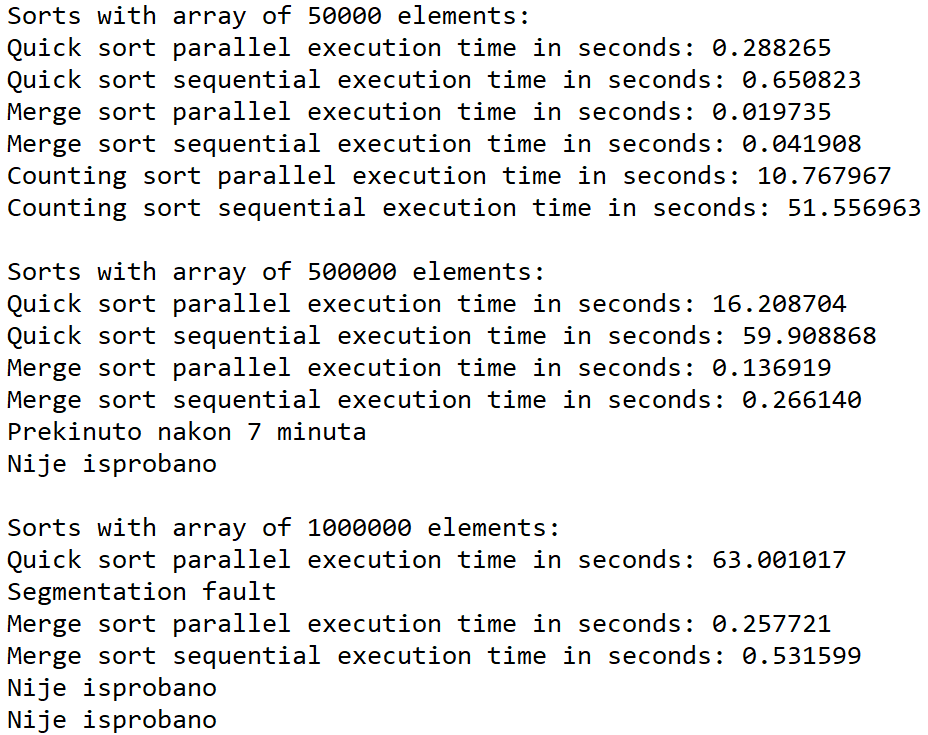
Slika 18 - Test sa N = 5000



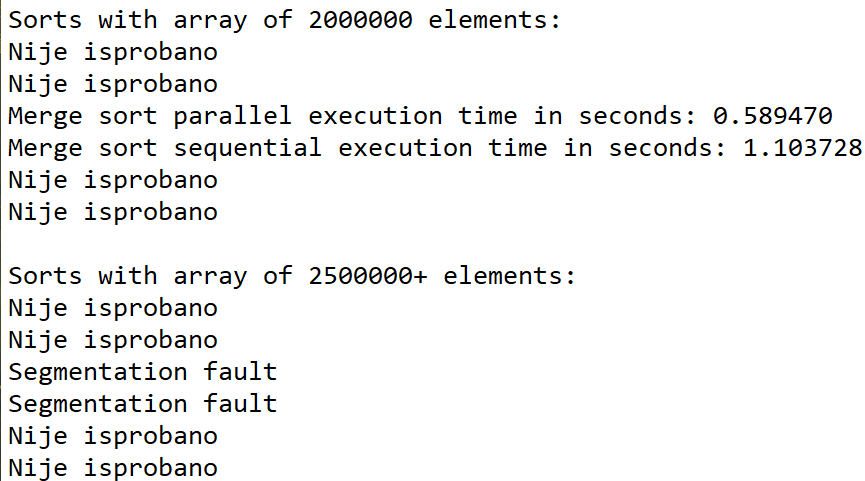
Slika 19 - Test sa N = 50000



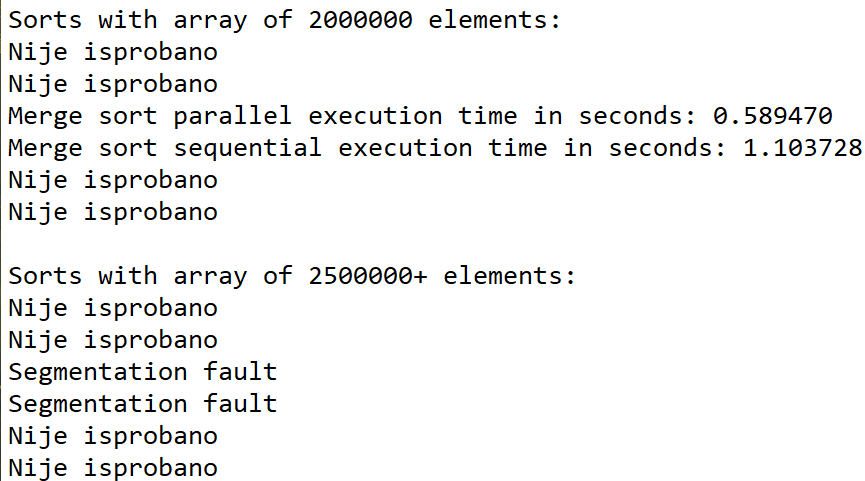
Slika 20 - Test sa N = 500000



Slika 21 - Test sa N = 1000000



Slika 22 - Test sa N = 2000000



Slika 23 - Test sa N = 2500000

# Zaključak

Prilikom izrade ovog rada, zaključeno je da prilikom sortiranja nizova malih dimenzija nema potrebe za primenom paralelizacije, jer samo kreiranje taskova oduzima previše vremena i izvršavanje paralelnog programa bude sporije od izvršavanja sekvencijalnog programa (videti sliku 16). Dok je u slučajevima sortiranja nizova velikih dimenzija poželjno primeniti paralelizaciju jer se postiže ubrzanje od oko 50%.

Implementacije koje su date u radu nisu pogodne za sve vrste nizova, naime kada su u pitanju veliki nizovi date implementacije nemaju efekta (implementacija paralelnog Quick sorta pri radu sa nizovima od milion elemenata ima vreme izvršavanja od oko 60 sekundi, dok kada su u pitanju nizovi sa preko 2.5 miliona elemenata ni jedan od sortova ne radi), te je potrebno istražiti i druge implementacije ili algoritme.

# Literatura

1. Wikipedia, Sorting algorithm – <https://en.wikipedia.org/wiki/Sorting_algorithm>
2. – <https://sortvisualizer.com/>
3. Programiz, Asymptotic notations– <https://www.programiz.com/dsa/asymptotic-notations>
4. Leonardo Galler and Matteo Kimura, Sorting algorithms – <https://lamfo-unb.github.io/2019/04/21/Sorting-algorithms/>
5. HP, Parallel computing and its modern uses – <https://www.hp.com/us-en/shop/tech-takes/parallel-computing-and-its-modern-uses>
6. Ricardo Rocha and Fernando Silva, Parallel computing – <https://www.dcc.fc.up.pt/~ricroc/aulas/1516/cp/apontamentos/slides_sorting.pdf>
7. Geeks for Geeks, Divide and Conquer– <https://www.geeksforgeeks.org/divide-and-conquer-algorithm-introduction/>
8. Geeks for Geeks, Quick sort– <https://www.geeksforgeeks.org/quick-sort/>
9. Geeks for Geeks, Merge sort– <https://www.geeksforgeeks.org/merge-sort/>
10. Geeks for Geeks, Counting sort – <https://www.geeksforgeeks.org/counting-sort/>

# Spisak slika

[Slika 1 - Prikaz algoritama za sortiranje sa njihovim kompleksnostima [4] 5](#_Toc96899256)

[Slika 2 - Prikaz načina rada Quick sort algoritma [8] 7](#_Toc96899257)

[Slika 3 - Prikaz quickSort funkcije 8](#_Toc96899258)

[Slika 4 - Prikaz partition funkcije 8](#_Toc96899259)

[Slika 5 - Prikaz poziva quickSort funkcije 8](#_Toc96899260)

[Slika 6 - Prikaz quickSort funkcije - paralelno 9](#_Toc96899261)

[Slika 7 - Prikaz poziva quickSort funkcije - paralelno 9](#_Toc96899262)

[Slika 8 - Prikaz načina rada Merge sort algoritma [10] 9](#_Toc96899263)

[Slika 9 - Prikaz mergeSort funkcije 10](#_Toc96899264)

[Slika 10 - Prikaz mergeSortAux funkcije 10](#_Toc96899265)

[Slika 11 - Prikaz poziva mergeSort funkcije 10](#_Toc96899266)

[Slika 12 - Prikaz mergeSort funkcije - paralelno 11](#_Toc96899267)

[Slika 13 - Prikaz poziva mergeSort funkcije - paralelno 11](#_Toc96899268)

[Slika 14 - Prikaz countingSort funkcije 12](#_Toc96899269)

[Slika 15 - Prikaz poziva countingSort funkcije 12](#_Toc96899270)

[Slika 16 - Test sa N = 50 13](#_Toc96899271)

[Slika 17 - Test sa N = 500 13](#_Toc96899272)

[Slika 18 - Test sa N = 5000 13](#_Toc96899273)

[Slika 19 - Test sa N = 50000 13](#_Toc96899274)

[Slika 20 - Test sa N = 500000 14](#_Toc96899275)

[Slika 21 - Test sa N = 1000000 14](#_Toc96899276)

[Slika 22 - Test sa N = 2000000 14](#_Toc96899277)

[Slika 23 - Test sa N = 2500000 14](#_Toc96899278)

1. Više o Divide and Conquer algoritmima možete pronaći na [7]. [↑](#footnote-ref-1)