SVEUČILIŠTE U ZAGREBU

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

PROJEKT IZ BIOINFORMATIKE

Ukkonenov Algoritam

Mislav Marković  
Filip Matijević  
Marko Raguž

Voditelj: Mislav Marković

Zagreb, siječanj, 2019.

**Sadržaj**

[1. Uvod 4](#_Toc535226067)

[2. Opis algoritma 5](#_Toc535226068)

[3. Rezultati testiranja 6](#_Toc535226069)

[4. Zaključak 7](#_Toc535226070)

[5. Literatura 8](#_Toc535226071)

# Uvod

Da bismo objasnili kako radi Ukkonenov algoritam potrebno je prvo objasniti što je to sufiksno stablo. Sufiksno stablo je je komprimirani trie koji sadrži sve sufikse zadatnog teksta gdje su ključevi sufiksi a vrijednosti su pozicije u originalnom tekstu.

Naivna implementacija sufiksnog stabla zahtjeva O(n^2) ili čak O(n^3) vremensku složenost.

Ukkonenov Algoritam je online (slovo po slovo) algoritam koji gradi sufiksno stablo u linearnom vremenu, predložen od Eska Ukkonena prema kojem je dobio i ime.

# Opis algoritma

Struktura podataka, tj. Sufiksno stablo, koje na kraju trebamo stvoriti iz danog teksta je poput search trie-a ali umjesto jednog znaka po bridu kao oznaka brida koristit će se par brojeva [od, do]. To su pokazivači na originalni text te tako svaki brid predstavlja String proizvoljne duljine, ali zauzima samo O(1) memorije jer se spremaju samo 2 pokazivača.

Objasnit ćemo algoritam ukratko kroz nekoliko pravila. Za početak potrebne su dodatne varijable.

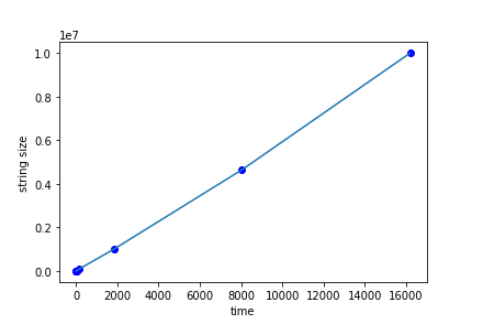
1. Aktivna točka - trojka( aktivni\_čvor, aktivni\_brid, aktivna\_dužina). Govori nam gdje moramo početi sa umetanjem novog sufiksa.
2. Ostatak – Govori nam koliko suffiksa moramo umetnuti izravno. Npr ako je tekst abcdefg, a ostatak je 3 onda moramo procesirati zadnja 3 sufiksa “efg”, “fg”, te “g”.

Dodati ćemo i koncept unutrašnjeg čvora, to su svi čvorovi osim korjena i listova.

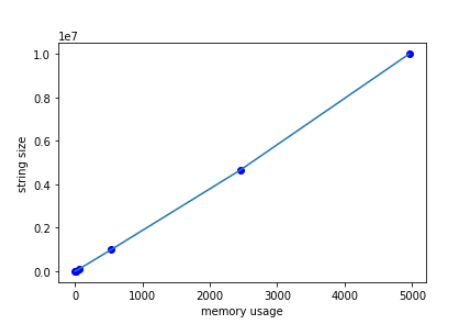
...TODO...

# Rezultati testiranja

Algoritam smo testirali na tekstovima različite duljine, tj. 100, 10k, 100k, 1mil, te 10mil znakova. Vremenska i memorijska složenost je prikazana na grafovima.



Vremenska složenost



Memorijska složenost

# Zaključak

Testiranjem smo uočili da je i vremenska i memorijska složenost linearna, što je puno bolje od naivne implementacije sufiksnog stabla( O(n^2)/ O(n^3) vremenska složenost). Ukkonenov algoritam je također on-line algoritam, u svakom trenutku ima sufiksno stablo spremno od trenutno pročitanog teksta, što može biti korisno ukoliko nemamo potpuni tekst u početku.

# Literatura

1.) <https://www.cs.helsinki.fi/u/ukkonen/SuffixT1.pdf>

2.) <https://en.wikipedia.org/wiki/Ukkonen%27s_algorithm>

3.) <https://stackoverflow.com/questions/9452701/ukkonens-suffix-tree-algorithm-in-plain-english>

4.) <https://en.wikipedia.org/wiki/Suffix_tree>

5.) <https://en.wikipedia.org/wiki/Trie>