

PROJEKTOVANJE I ANALIZA ALGORITAMA

Laboratorijska vežba 1
Ristovski Nikola 19347
Tehnologija : C++

Korišćeni resursi

Za izvršenje koristio sam file-ove od 100, 1000, 10000 i 100000 reči. Tekstualne file-ove sam dobio korišćenjem online generatora Lorem Ipsum, koji je generisao latinske smislene reči odgovarajućih dužina. Hex file-ove sam generisao tako što sam napravio svoju skriptu u C++ (nalazi se u ZIP fajlu kao referenca) koja je formirala hex reči random dužine nakon čega je takođe random postavljan blanko znak, tako da sam na kraju dobio mogućnost da text split-ujem na reči. U svrhu splita sam na internetu pronašao rešenje preko regex klase. Za merenje vremena koristio sam chrono klasu i njene funkcije.

Lorem Ipsum : <https://www.lipsum.com/>

KMP SEARCH

Izvršenje programa

Rezultati u konzoli su prikazani na slikama.

Izvršenje KMP SEARCH
algoritma na HEX file-ove
različitih dužina.

```
[ HEX od 100 reci ]  
  
Za izvršenje je bilo potrebno: 0.112 ms.  
Za izvršenje je bilo potrebno: 0.113 ms.  
Za izvršenje je bilo potrebno: 0.121 ms.  
Za izvršenje je bilo potrebno: 0.121 ms.  
  
[ HEX od 1.000 reci ]  
  
Za izvršenje je bilo potrebno: 1.476 ms.  
Za izvršenje je bilo potrebno: 1.136 ms.  
Za izvršenje je bilo potrebno: 1.142 ms.  
Za izvršenje je bilo potrebno: 1.207 ms.  
  
[ HEX od 10.000 reci ]  
  
Za izvršenje je bilo potrebno: 12.028 ms.  
Za izvršenje je bilo potrebno: 12.05 ms.  
Za izvršenje je bilo potrebno: 11.67 ms.  
Za izvršenje je bilo potrebno: 14.843 ms.  
  
[ HEX od 100.000 reci ]  
  
Za izvršenje je bilo potrebno: 107.613 ms.  
Za izvršenje je bilo potrebno: 125.004 ms.  
Za izvršenje je bilo potrebno: 117.737 ms.  
Za izvršenje je bilo potrebno: 128.462 ms.
```

```
[ Text od 100 reci ]  
  
Za izvršenje je bilo potrebno: 0.047 ms.  
Za izvršenje je bilo potrebno: 0.046 ms.  
Za izvršenje je bilo potrebno: 0.049 ms.  
Za izvršenje je bilo potrebno: 0.052 ms.  
  
[ Text od 1.000 reci ]  
  
Za izvršenje je bilo potrebno: 0.468 ms.  
Za izvršenje je bilo potrebno: 0.432 ms.  
Za izvršenje je bilo potrebno: 0.455 ms.  
Za izvršenje je bilo potrebno: 0.466 ms.  
  
[ Text od 10.000 reci ]  
  
Za izvršenje je bilo potrebno: 4.481 ms.  
Za izvršenje je bilo potrebno: 4.475 ms.  
Za izvršenje je bilo potrebno: 4.692 ms.  
Za izvršenje je bilo potrebno: 4.972 ms.  
  
[ Text od 100.000 reci ]  
  
Za izvršenje je bilo potrebno: 46.746 ms.  
Za izvršenje je bilo potrebno: 47.18 ms.  
Za izvršenje je bilo potrebno: 51.313 ms.  
Za izvršenje je bilo potrebno: 70.525 ms.
```

Izvršenje KMP SEARCH
algoritma na TEXT file-ove
različitih dužina.

Zaključci

Jasno uočavamo da se sa povećanjem broja reči u text-u u kome se pretražuje dati pattern povećava i vreme potrebno za izvršenje pomenute pretrage. Ovde uočavamo da povećanje potrebnog vremena srazmerno odgovara povećanju broja reči u text-u, pa tako sa povećanjem broja reči sa 100 na 1000, što je povećanje reda 10, i vreme za izvršenje se povećava desetostruko, sa 0.112ms na 1.476ms (uzimajući pattern-e dužine 5, mada isto pratimo i u ostalim slučajevima dužine pattern-a). Ovo uočavamo i u text i u hex file-ovima, i potpuno je ono što očekujemo.

Kada razmatramo slučaj iste dužine texta za pretragu a varijabilne dužine pattern-a (tj. podstringa koji se u istom traži), uočavamo neznatnu, ali приметljivu razliku u performansama. Pretraga kraćih pattern-a uglavnom traje kraće od pretrage dužih pattern-a. To je i jasno s obzirom da duži pattern zahteva više poredjenja, upravo zato što je duži. Kao i malopre, i ovo se odnosi na oba text i hex file-ove.

VAŽNO: U ovom izvršenju, prikazanom na slikama, nisam štampao pronađene podstringove. To je zato što štampa na ekran zahteva određeno nezanemarljivo vreme, a namera je bila porediti performanse samog KMP SEARCH-a, u samom kodu je moguće dekomentarisati funkciju za štampu, pa se odatle mogu porediti performanse sa štampom.

Kada poredimo izvršenja text sa hex file-ovima, uočavamo da je приметljivo više vremena potrebno izvršenju hex file-ova. To je zato što su se generisale hex reči velikih dužina usled random generatorske funkcije, pa je malo više poredjenja prisutno.

LEVENSTEIN SEARCH

Izvršenje

Izvršenje TEXT file-ova (rezultat iz konzole):

```
<< LEVENSTEIN SEARCH TEXT FILE >>

[ Text od 100 reci ]

[ Pattern duzine oko 5 ]
Word: mosuere | pattern: posuere | distance: 1
Word: possere | pattern: posuere | distance: 1
Word: mosuere | pattern: posuere | distance: 1
Word: Posuere | pattern: posuere | distance: 1
Za izvršenje je bilo potrebno: 23.794 ms.

[ Pattern duzine oko 10 ]
Za izvršenje je bilo potrebno: 11.267 ms.

[ Pattern duzine oko 20 ]
Za izvršenje je bilo potrebno: 8.892 ms.

[ Text od 1.000 reci ]

[ Pattern duzine oko 5 ]
Za izvršenje je bilo potrebno: 97.309 ms.

[ Pattern duzine oko 10 ]
Word: belerisque | pattern: scelerisque | distance: 2
Za izvršenje je bilo potrebno: 87.436 ms.

[ Pattern duzine oko 20 ]
Word: incomprehensibili | pattern: incomprehensibilis | distance: 1
Za izvršenje je bilo potrebno: 90.496 ms.

[ Text od 10.000 reci ]

[ Pattern duzine oko 5 ]
Word: possere | pattern: posuere | distance: 1
Za izvršenje je bilo potrebno: 821.279 ms.

[ Pattern duzine oko 10 ]
Word: selerisque | pattern: scelerisque | distance: 1
Word: sceverisque | pattern: scelerisque | distance: 1
Za izvršenje je bilo potrebno: 846.723 ms.

[ Pattern duzine oko 20 ]
Word: comprehensibilis | pattern: incomprehensibilis | distance: 2
Word: incomprehensibilum | pattern: incomprehensibilis | distance: 2
Za izvršenje je bilo potrebno: 896.357 ms.

[ Text od 100.000 reci ]

[ Pattern duzine oko 5 ]
Word: nosuere | pattern: posuere | distance: 1
Za izvršenje je bilo potrebno: 6833.82 ms.

[ Pattern duzine oko 10 ]
Word: severisque | pattern: scelerisque | distance: 2
Word: gaelerisque | pattern: scelerisque | distance: 2
Za izvršenje je bilo potrebno: 5018.04 ms.

[ Pattern duzine oko 20 ]
Word: incomprehensibilite | pattern: incomprehensibilis | distance: 2
Word: incomprehensibili | pattern: incomprehensibilis | distance: 1
Za izvršenje je bilo potrebno: 5380.15 ms.
```

Izvršenje HEX file-ova (rezultat iz konzole):

<< LEVENSTEIN SEARCH HEX FILE >>

[Text od 100 reci]

[Pattern duzine oko 5]

Word: e644fb | pattern: e648fb | distance: 1

Za izvršenje je bilo potrebno: 7.261 ms.

[Pattern duzine oko 10]

Za izvršenje je bilo potrebno: 6.733 ms.

[Pattern duzine oko 20]

Word: 2b206d973864bfad2c25 | pattern: 2b206d973863bfdd2c25 | distance:

Za izvršenje je bilo potrebno: 8.762 ms.

[Text od 1.000 reci]

[Pattern duzine oko 5]

Za izvršenje je bilo potrebno: 60.246 ms.

[Pattern duzine oko 10]

Za izvršenje je bilo potrebno: 63.251 ms.

[Pattern duzine oko 20]

Word: 2b206d973863bd1d2c2 | pattern: 2b206d973863bfdd2c25 | distance: 3

Word: 1b206d73863bfdd2c2 | pattern: 2b206d973863bfdd2c25 | distance: 3

Za izvršenje je bilo potrebno: 81.099 ms.

[Text od 10.000 reci]

[Pattern duzine oko 5]

Word: e648ff | pattern: e648fb | distance: 1

Za izvršenje je bilo potrebno: 652.407 ms.

[Pattern duzine oko 10]

Word: 3a8faab3111 | pattern: 3a8faab39c1 | distance: 2

Word: 2a8faab39a1 | pattern: 3a8faab39c1 | distance: 2

Za izvršenje je bilo potrebno: 682.903 ms.

[Pattern duzine oko 20]

Word: 2b206d973863bfdd25 | pattern: 2b206d973863bfdd2c25 | distance: 2

Za izvršenje je bilo potrebno: 799.872 ms.

[Text od 100.000 reci]

[Pattern duzine oko 5]

Word: e647fb | pattern: e648fb | distance: 1

Word: e648ab | pattern: e648fb | distance: 1

Za izvršenje je bilo potrebno: 6463.29 ms.

[Pattern duzine oko 10]

Word: 2a8faab39c1 | pattern: 3a8faab39c1 | distance: 1

Word: 3a8falb39c1 | pattern: 3a8faab39c1 | distance: 1

Word: 3a8faab39c2 | pattern: 3a8faab39c1 | distance: 1

Za izvršenje je bilo potrebno: 6743.16 ms.

[Pattern duzine oko 20]

Word: 2b656d973863bfdd2c25 | pattern: 2b206d973863bfdd2c25 | distance: 2

Za izvršenje je bilo potrebno: 7687.67 ms.

Zaključci

Ovde primećujemo slične osobine kao i kod KMP algoritma. Pre svega, ako razmatramo prelaz sa file-a od 100 na 1000 reči, videćemo da se potrebno vreme povećava 8 do 10 puta. To jasno vidimo u oba tipa file-ova (hex i text).

U okviru jednog text-a a promenom dužine pattern-a, opet primećujemo da se povećanjem dužine pattern-a povećava i potrebno vreme. To je i logično jer je potrebno formirati matricu većih dimenzija, što povećava broj potrebnih operacija. Prisutno je u oba tipa file-ova.

Na kraju, između TEXT i HEX file-ova opet možemo uočiti povišeno zahtevano vreme kod HEX file-a, a opet iz istog razloga kao i kod KMP algoritma, jer su reči u HEX file-ovima u proseku duže zbog randomiziranog generisanja.

VAŽNO: Jasno uočavamo da kada je substring pronađen on se štampa na konzoli. Zato je znatno više vremena potrebno za to izvršenje (štampa dodaje potrebno vreme) u odnosu na situaciju kada substring nije pronađen (uočimo plavo obojene vrednosti vremena gore u izvršenju text file-ova). Ako zakomentarišemo štampu, ova pojava nestaje i uočićemo ono što i očekujemo, da je za kraće patter-ne vreme takođe kraće.

VAŽNO: Takođe sam u kodu formatirao prikaz pattern-a i nađenog substring-a u vidu matrice, ta štampa zahteva još više vremena pa je pogodno koristiti tu štampu samo kada želimo da vizuelno uporedimo ove dve reči. U kodu možete zakomentarisati i raskomentarisati po želji.

Ristovski Nikola, Elektronski fakultet Niš, 2024.