

# Projektovanje i analiza algoritama

Jovan Bogdanović 19059

## 1. zadatak, Knut-Moris-Prat i Levenstein

- Za smislen tekst ascii karaktera sam koristio knjigu Hobbit na engleskom jeziku da ne bih imao eventualnih problema problema sa latiničnim slovima, a hex tekst sam sam generisao pomoću klase HexGenerator. Fajlovi se nalaze u folderu projekta na putanji ...\\19059Csharp\\bin\\Debug kao i fajlovi u kome se nalaze izlazi programa koje sam nazvao izvestajKMP i izvestajLevenstein respektivno.
- Sama aplikacija ne štampa apsolutno ništa na konzolu sem grešaka koje se možda dese u try catch delu kada se fajlovi za čitanje ne otvore dobro.
- Rezultati se mogu videti i u ovom dokumentu ali detaljniji izveštaj programa se vidi bolje u pomenutim izlaznim fajlovima jer program za KMP ispisuje svaku poziciju na kojoj se nalazi poklapanje a Levenstein ispisuje svaku reč koja je na udaljenosti manjoj od 20%, a pošto su podstringovi namerno izabrati tako da ih sigurno ima puno u tekstu, izostavljeni su ti rezultati iz ovog dokumenta.
- Vremena za pretraživanje stringa pomoću KMP-a veličine raste sa veličinom teksta u kome se pretražuje što je očekivano, ni jedan prosledjeni string nema povoljne prefiske i sufikse tako da ne može da se očekuje neko poboljšanje u odnosu na neke prostije algoritme iako je to poenta KMP pretrage koju realno ne iskorišćavamo.
- Izvršenje levenstein algoritam-a se znatno usporava kada povećamo broj reči što je takođe očekivano ali to nema konkretno veze sa sprovođenjem samog algoritma jer algoritam kao takav radi poređenje jednog stringa relativno kratke dužine sa našim prosledjenim stringom.
- Test je rađen na 2 različita procesora: AMD FX-6100 (star 12 godine) i AMD Ryzen 5 7520U (star godinu dana). Rezultati se značajno razlikuju, stariji procesor radnje izvršava i do 3 puta sporije u odnosu na noviji. Priloženi rezultati su sa novijeg procesora.

## Rezultati programa za KMP:

Ascii tekst dužine **100** reci ima 523 karaktera

podstring dužine **5** (green) – 1 polapanje za vreme **0.2776** ms

podstring dužine **10** (tubeshaped) – 1 polapanje za vreme **0.009** ms

podstring dužine **20** (Not a nasty, dirty, ) – 1 polapanje za vreme **0.0075** ms

podstring dužine **50** (Not a nasty, dirty, wet hole, filled with the ends) – 1 polapanje za vreme **0.015** ms

Ascii tekst dužine **1000** reci ima 5587 karaktera

podstring dužine **5** (green) – 3 polapanja za vreme **0.0524** ms

podstring dužine **10** (tubeshaped) – 1 polapanje za vreme **0.0573** ms

podstring dužine **20** (Not a nasty, dirty, ) – 1 polapanje za vreme **0.0488** ms

podstring dužine **50** (Not a nasty, dirty, wet hole, filled with the ends) – 1 polapanje za vreme **0.0513** ms

Ascii tekst dužine **10000** reci ima 53639 karaktera

podstring dužine **5** (Thorin) – 39 polapanja za vreme **0.4637** ms

podstring dužine **10** (tubeshaped) – 1 polapanje za vreme **0.4797** ms

podstring dužine **20** (burglary—especially ) – 1 polapanje za vreme **0.4498** ms

podstring dužine **50** (This news alters them much for the better. So far ) – 1 polapanje za vreme **0.5401** ms

Ascii tekst dužine **100000** reci ima 53639 karaktera

podstring dužine **5** (Bilbo) – 603 polapanja za vreme **5.1895** ms

podstring dužine **10** (determined) – 4 polapanja za vreme **5.0482** ms

podstring dužine **20** (burglary—especially ) – 1 polapanje za vreme **4.6852** ms

podstring dužine **50** (This news alters them much for the better. So far ) – 1 polapanje za vreme **4.4506** ms

Hex tekst dužine **100** reci ima 53639 karaktera

podstring dužine **5** (CBA25) – 1 polapanje za vreme **0.0222** ms

podstring dužine **10** (9B004F6AEA) – 1 polapanje za vreme **0.0243** ms

podstring dužine **20** (10D2678DF0EDA39F1789) – 1 polapanje za vreme **0.0251** ms

podstring dužine **50** (688ACCC3B6A4DF326B 876529E4126 73564C2CAFO 4A2762C) – 1 polapanje za vreme **0.0195** ms

Hex tekst dužine **1000** reci ima 13050 karaktera

podstring dužine **5** (59711) – 1 polapanje za vreme **0.1249** ms

podstring dužine **10** (4FAF97CCFE) – 1 polapanje za vreme **0.1232** ms

podstring dužine **20** (B83A2492050A9560DAFB) – 1 polapanje za vreme **0.1226** ms

podstring dužine **50** (B64E624F2CE004C4770 CD661EBD7E578EBB30 FA26613E3 B) – 1 polapanje za vreme **0.1314** ms

Hex tekst dužine **10000** reci ima 129748 karaktera

podstring dužine **5** (4F047) – 1 polapanje za vreme **1.2065** ms

podstring dužine **10** (E3D7F3EBF8) – 1 polapanje za vreme **1.1889** ms

podstring dužine **20** (B79080C1E6E5D84AB63F) – 1 polapanje za vreme **1.2214** ms

podstring dužine **50** (2C508B55D70A9F4162 1EA01D7 86C5732CA427D296C42 B4A) – 1 polapanje za vreme **1.1945** ms

Hex tekst dužine **100000** reci ima 1300569 karaktera

podstring dužine **5** (C3C3F) – 5 polapanja za vreme **12.0278** ms

podstring dužine **10** (66943CB78E) – 1 polapanje za vreme **12.0715** ms

podstring dužine **20** (A3FDDF4B86A09D09E54F) – 1 polapanje za vreme **13.5843** ms

podstring dužine **50** (F72C8096935793DDB7 127038DC 17F933B4721A2F78 83436) – 1 polapanje za vreme **11.9947** ms

## Rezultati programa za levenstein (poklapanje do 20%):

Ascii tekst dužine **100** reci

podstring dužine **5** (round) – **2** poklapanja

podstring dužine **10** (tubeshaped) – **1** poklapanje

podstring dužine **15** (panelled-walls) – **1** poklapanje

Ascii tekst dužine **1000** reci

podstring dužine **5** (lived) – **3** poklapanja

podstring dužine **10** (tubeshaped) – **1** poklapanje

podstring dužine **15** (panelled-walls) – **1** poklapanje

Ascii tekst dužine **10000** reci

podstring dužine **6** (Thorin) – **34** poklapanja

podstring dužine **10** (determined) – **1** poklapanje

podstring dužine **20** (burglary—especially,) – **1** poklapanje

Ascii tekst dužine **100000** reci

podstring dužine **5** (Bilbo) – **545** poklapanja

podstring dužine **10** (determined) – **4** poklapanja

podstring dužine **20** (burglary—especially,) – **1** poklapanje

Hex tekst dužine **100** reci

podstring dužine **5** (CBA25) – **1** poklapanje

podstring dužine **10** (9B004F6AEA) – **1** poklapanje

podstring dužine **20** (10D2678DF0EDA39F1789) – **1** poklapanje

Hex tekst dužine **1000** reci

podstring dužine **5** (59711) – **1** poklapanje

podstring dužine **10** (4FAF97CCFE) – **1** poklapanje

podstring dužine **20** (B83A2492050A9560DAFB) – **1** poklapanje

Hex tekst dužine **10000** reci

podstring dužine **5** (4F047) – **1** poklapanje

podstring dužine **10** (E3D7F3EBF8) – **1** poklapanje

podstring dužine **20** (B79080C1E6E5D84AB63F) – **1** poklapanje

Hex tekst dužine **100000** reci

podstring dužine **5** (C3C3F) – **2** poklapanja

podstring dužine **10** (66943CB78E) – **1** poklapanje

podstring dužine **20** (A3FDDF4B86A09D09E54F) – **1** poklapanje