

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I
RAČUNALSTVA

PROJEKT IZ BIOINFORMATIKE

**Računanje najduljeg zajedničkog
prefixa temeljeno na BWT**

Autori

ZVONIMIR JURELINAC, TOMISLAV ŽIVEC, TONKO ČUPIĆ

Voditelj

doc.dr.sc MIRJANA DOMAZET- LOŠO

Zagreb, prosinac 2017.

Contents

1	Uvod	2
----------	-------------	----------

1 Uvod

Bioinformatika je grana znanosti koja usko povezuje biologiju i računarstvo, a ubrzano se razvijala zadnja dva desetljeća. Pojeftinjenje i sve veća dostupnost tehnologije sekvenciranja rezultirale su stvaranjem velikih skupova bioloških podataka. Često se kao zadatak u bioinformatici nameće analiza sekvence genoma. Pošto su te sekvence predugačke za uobičajenu pohranu i analizu, potrebno je koristiti posebna sufiksna polja i polja najdužeg zajedničkog prefiksa. Cilj projekta je bio implementirati algoritme 1 i 2 iz rada Beller et al. (2013), koristeći gotovu knjižnicu za izgradnju sufiksnog polja. Zatim je implementacija testirana na 2 načina: sa postojećom knjižnicom stabla valića i sa našom implementacijom stabla valića i njegove funkcije rang. Rješenje je uspoređeno s rezultatima iz studentskog rada Mrčela et al. (2016). Kao ulazni niz koristili smo genom bakterije E. Coli