

Zagreb, 2 March 2018

MASTER THESIS ASSIGNMENT No. 1572

Student: **Neven Miculinić (0036477241)**
Study: Computing
Profile: Computer Science

Title: **End-to-End Deep Learning Model for Base Calling of MinION Nanopore Reads**

Description:

In the MinION device, single-stranded DNA fragments move through nanopores, which causes drops in the electric current. The electric current is measured at each pore several thousand times per second. Each event is described by the mean and variance of the current and by event duration. This sequence of events is then translated into a DNA sequence by a base caller. Develop a base-caller for MinION nanopore sequencing platform using a deep learning architecture such as convolutional neural networks and recurrent neural networks. Instead of events, use current waveform at the input. Compare the accuracy with the state-of-the-art basecallers. For testing purposes use publicly available datasets and Graphmap or Minimap 2 tools for aligning called reads on reference genomes. Implement method using TensorFlow or similar library. The code should be documented and hosted on a publicly available Github repository.

Issue date: 16 March 2018
Submission date: 29 June 2018

Mentor:



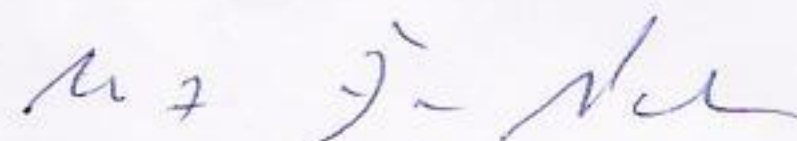
Associate Professor Mile Šikić, PhD

Committee Secretary:



Assistant Professor Tomislav Hrkać, PhD

Committee Chair:



Full Professor Siniša Srbljić, PhD

Zagreb, 2. ožujka 2018.

DIPLOMSKI ZADATAK br. 1572

Pristupnik: **Neven Miculinić (0036477241)**
Studij: Računarstvo
Profil: Računarska znanost

Zadatak: **Model dubokog učenja s kraja na kraj za određivanje očitanih baza dobivenih uređajem za sekvenciranje MinION**

Opis zadatka:

Unutar uređaja MinION, fragmenti jednostruke DNA prolaze kroz nanopore, što uzrokuje promjene u električnoj struji. Struja proizvedena na svakoj nanopori mjeri se nekoliko tisuća puta u sekundi. Svaki događaj opisan je srednjom vrijednosti i varijancom struje te svojim trajanjem. Postupak kojim se takav slijed događaja prevodi u niz nukleotida naziva se određivanje očitanih baza. Razviti alat za prozivanje baza za uređaj za sekvenciranje MinION koristeći modele dubokog učenja kao što su konvolucijske i povratne neuronske mreže. Umjesto događaja na ulazu koristi valni oblik struje. Usporediti dobivenu točnost s postojećim rješenjima. U svrhu testiranja koristiti javno dostupne skupove podataka i alate GraphMap ili Minimap 2 za poravnanje očitavanja na referentni genom. Alat implementirati koristeći programsku biblioteku TensorFlow (ili neku sličnu). Programski kod treba biti dokumentiran i javno dostupan preko repozitorija GitHub.

Zadatak uručen pristupniku: 16. ožujka 2018.

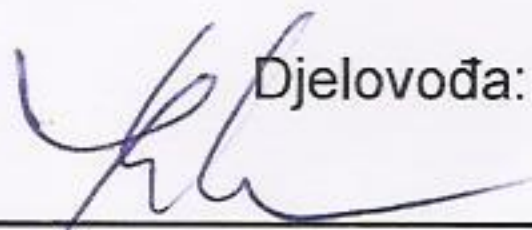
Rok za predaju rada: 29. lipnja 2018.

Mentor:



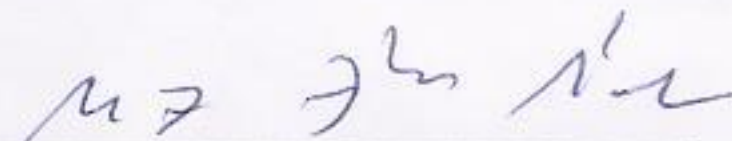
Izv. prof. dr. sc. Mile Šikić

Djelovođa:



Doc. dr. sc. Tomislav Hrkać

Predsjednik odbora za
diplomski rad profila:



Prof. dr. sc. Siniša Srbljić