



# УНИВЕРЗИТЕТ У БЕОГРАДУ - ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ

Булевар краља Александра 73, 11000 Београд, Србија

Тел. 011/324-8464, Факс: 011/324-8681

## КОМИСИЈИ ЗА СТУДИЈЕ II СТЕПЕНА ЕЛЕКТРОТЕХНИЧКОГ ФАКУЛТЕТА У БЕОГРАДУ

Комисија за студије II степена, Електротехничког факултета у Београду, на својој седници одржаној 16.05.2017. године именовала нас је у Комисију за преглед и оцену мастер рада дипл. инж. Марије Станојевић под насловом „Одређивање сличности између научних радова применом метода машинског учења“. Након прегледа материјала Комисија подноси следећи

### ИЗВЕШТАЈ

#### 1. Биографски подаци кандидата

Марија Станојевић је рођена 11.08.1991. године у Нишу. Завршила је основну школу "Доситеј Обрадовић" у Ћићевцу као вуковац и ђак генерације. Завршила је Гимназију „Светозар Марковић“ у Нишу, специјално одељење за ученике обдарене за математику, као вуковац и ђак генерације. Током школовања освојила је више награда на државним такмичењима из физике, хемије и математике. У средњој школи, 2008. године, основала је Фестивал „Наук није баук“ који је водила до средине 2011. године, а који траје већ 9 година и други је најуспешнији фестивал науке у Србији са интернационалним карактером.

Електротехнички факултет Универзитета у Београду уписала је 2010. године. Дипломирала је на одсеку за Софтверско инжењерство 2016. године са просечном оценом 8,51. Дипломски рад одбранила је у септембру 2016. године са оценом 10. У периоду 2010-2011. била је члан студентске организације ИСТЕК и у периоду 2010-2016. била је члан студентске организације БЕСТ у којој је била активна и на интернационалном нивоу. У периоду 2012-2014. године студирала је, у оквиру студентске размене, на Универзитету у Марибору, Факултет за електротехнику, рачунарство и информатику. Током студија на Електротехничком факултету у Београду била је демонстратор на неколико предмета у периоду октобар 2011 – јануар 2012. и октобар 2014 – јануар 2015. Од новембра 2014. до јануара 2017. године радила је као софтверски инжењер. Дипломске академске – мастер студије на Електротехничком факултету у Београду, на Модулу за сигнале и системе уписала је у октобру 2016. године. Положила је све испите са просечном оценом 9,4. Докторске студије уписала је на Темпле универзитету у Филаделфији, Пенсилванија, јануара 2017. године, са тренутном просечном оценом А (најбоља оцена).

#### 2. Опис мастер рада

Мастер рад кандидаткиње Марије Станојевић се бави описом методологија машинског учења за обраду природног текста и кластеризацију и одређивањем сличности радова на основу ових технологија. Предмет рада је кластеризација научних радова прикупљених са вебсајта *arxiv.org* у групе, коришћењем метода дубоког учења за креирање језичког модела и употребом алгоритама за кластеризацију текстова на основу њихове репрезентације у одговарајућим језичким моделима. Кандидаткиња је упоредила различите методологије описујући њихове карактеристике и на основу тога урадила тестирање за теоретски најбоље и најбрже технике над независним скупом података који није коришћен за тестирање ових метода раније. Рад је имплементиран у програмским језицима *python* и *matlab*. Дата тема је обрађена у седам поглавља у оквиру којих се налазе увод, закључак као и списак коришћене литературе.

Уводно поглавље садржи опис проблема, циљ рада преглед структуре рада. Описан је значај квалитетног алгоритма за ненадгледано груписање научних радова и одређивање сличности међу радовима.

Друго поглавље представља преглед техника за претпроцесирање језика. У оквиру поглавља су презентоване различите методе за рад са природним или програмским језиком које се појављују у литератури. За сваки од поменутих алгоритама, описана је идеја, теоретска основа, начини за примену и имплементације у постојећим библиотекама.

Треће поглавље фокусирано је на алгоритме за процесирање језика. Кандидаткиња је дала историјски приказ ових техника, с нагласком на најновије методе. За сваку методу дат је теоретски опис, идеја, предности и мане у односу на друге алгоритме, временска сложеност и имплементација у различитим програмским језицима. Такође, у овом поглављу приказане су структуре за дубоко учење на којима су се базирали горе наведени алгоритми.

Четврто поглавље обухвата опис техника за кластеризацију, при чему су приказане методе добро познате у литератури, али и новија решења базирана на дубоком учењу. За те методе приказана је идеја, теоретска основа, предности, мане и временска сложеност.

У петом поглављу имплементирана је кластеризација научних радова са вебсајта *arxiv.org* користећи најбоље технике за креирање језичког модела и често коришћене технике за кластеризацију. Кандидаткиња је анализирала рад алгоритама у одређивању сличности између речи, наслова и апстракта радова. Такође, упоређиване су перформансе у кластеризацији докумената на основу њихове репрезентације у језичким моделима.

Шесто поглавље је закључак у коме је извршена рекапитулација рада. Сумирана су искуства која је кандидаткиња током израде ове тезе стекла, те су предложени даљи кораци за анализу коришћених методологија и унапређење кластеризације научних радова.

У седмом поглављу дат је преглед коришћене литературе при изради тезе.

### 3. Оцена рада и закључак

Кандидаткиња Марија Станојевић се у свом мастер раду бавила кластеризацијом текстова на основу метода машинског учења, с акцентом на методама дубоког учења. Успешно и систематично је образложена идеја креирања језичког модела различитим алгоритмима и кластеризације репрезентација научних радова. Кандидаткиња је показала самосталност и систематичност у свом раду.

На основу изложеног Комисија има задовољство да предложи Комисији за студије другог степена Електротехничког факултета у Београду да прихвати мастер рад под насловом „Одређивање сличности између научних радова применом метода машинског учења“ и да његовом аутору, кандидаткињи Марији Станојевић, дипл.инж. одобри усмену одбрану.

Београд, 08. 09. 2017. године

Чланови комисије:

  
Проф. др Жељко Ђуровић

  
Проф. др Бранко Ковачевић