



# 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 именовало нас је у Комисију за преглед и оцену мастер рада дипл. инж. Александра Милинковића, бр. индекса 3288/17, под насловом “Утицај техника морфолошких нормализација на одређивање семантичке сличности кратких текстова на српском језику”. Након прегледа материјала Комисија подноси следећи

### ИЗВЕШТАЈ

#### 1. Биографски подаци кандидата

Александар Милинковић рођен је 15.05.1988. у Београду. Основну и средњу школу завршио је у Београду, а на Електротехнички факултет се уписао 2007. године, где је 2013. дипломирао на Одсеку за рачунарску технику и информатику са просечном оценом 7,73 и оценом 10 на дипломском раду. У периоду од 2014. до 2015. радио је у Лондону за фирму Maxeler. По повратку се запослио у италијанској фирмам Evolvea, где је радио до 2017. Тренутно је запослен у фирмам Fidelity Information Services (FIS) као софтвер инжењер. Мастер студије на Електротехничком факултету уписао је у октобру 2017. године на модулу Рачунарска техника и информатика. Положио је све испите са просечном оценом 8,20.

#### 2. Опис мастер рада

Мастер рад обухвата 50 страна, а подељен је у 5 поглавља и садржи 6 слика, 3 табеле и 64 цитиране референце.

У првом поглављу, уводу, дат је кратак опис проблема, наведене су неке постојеће методе његовог решавања и указано на њихове предности и недостатке. Такође, у овом делу описана је и структура самог рада.

У другом поглављу представљени су неки алгоритми за генерирање вектора који описују везу речи и контекста у којем се те речи налазе. Наведена теорија односи се на речи које се директно добијају из улазног корпуса. Приказане су и две методе за убрзавање потребних израчунавања. Дефинисани су параметри који се могу подешавати ради добијања максималне тачности код употребље наведених алгоритама.

У трећем поглављу описане су две врсте морфолошких нормализација улазног корпуса (стемери и лематизатори), заједно са укупно шест алгоритама који припадају овим врстама, и дефинисане су величине које су коришћене у анализи утицаја нормализација на одређивање семантичке сличности.

У четвртом поглављу описан је начин прикупљања и обраде текста за добијање корпуса речи српског језика. Описан је и програмски пакет за одређивање семантичке сличности кратких текстова, са освртом на употребљене алгоритме. Затим су приказани резултати варијација хиперпараметара алгоритма за одређивање векторске презентације речи, и на крају, приказан је утицај морфолошких нормализација на тачност одређивања семантичке сличности.

У петом поглављу, закључку, наводе се постигнути циљеви овог рада и указује се на могућности за евентуално проширивање и унапређивање реализованог решења.

### 3. Анализа рада са кључним резултатима

Мастер рад дипл. инж. Александра Милинковића припада области обраде природног језика, прецизније, области одређивања семантичке сличности, која представља концепт додељивања одређене метрике скуповима израза или докумената засноване на сличности њиховог значења. Семантичка сличност кратких текстова је посебно значајна, јер су кратки текстови данас у широкој употреби у виду упита и резултата претраживача, наслова и сажетака вести, коментара на друштвеним мрежама, итд.

Коришћењем STSSFramework-а, у раду је испитивана тачност процене семантичке сличности код употребе морфолошких нормализација на корпусу српских парафраза, а у оквиру векторског простора који је добијен из корпуса речи целокупног .rs домена.

Основни доприноси рада су: 1) data set припрема: филтрирање корпуса српског језика величине око 3 GB (уклањање специјалних знакова и интерпункције, таговање бројева и URL-ова, итд.) и његова нормализација; 2) анализа алгоритма за добијање вектора обележја и одређивање његових оптималних хиперпараметара за постизање максималне тачности; 3) испитивање утицаја стемера и лематизатора на тачност одређивања семантичке сличности, уз векторизацију употребом претходно одређених оптималних хиперпараметара.

### 4. Закључак и предлог

Да би се добили резултати помоћу којих бимашински семантичка сличност била упоредива са људском проценом, потребно је дефинисати обележја и затим их у оквиру одговарајуће метрике користити за утврђивање степена сличности. Кандидат Александар Милинковић је у свом раду успешно дефинисао та обележја, употребио их за добијање векторског простора на српском језику, а затим, на корпусу парафраза, анализирао како технике нормализације утичу на семантичке сличности кратких текстова на српском језику.

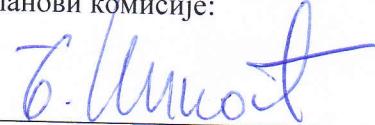
Алгоритми за векторизацију и нормализацију, заједно са великим скуповима улазних података, веома су временски и просторно захтевни. Кандидат је испитао перформансе ових алгоритама када се они извршавају коришћењем уобичајених рачунарских ресурса.

Кандидат је исказао самосталност и систематичност у своме поступку као и иновативне елементе у решавању проблематике овог рада.

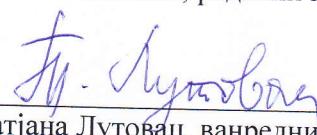
На основу изложеног, Комисија предлаже Комисији за студије II степена Електротехничког факултета у Београду да рад дипл. инж. Александра Милинковића прихвати као мастер рад и кандидату одобри јавну усмену одбрану.

Београд, 20.08.2018 године

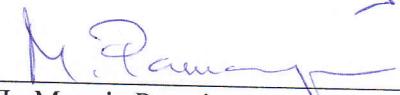
Чланови комисије:

  
Б. Николић

Др Бошко Николић, редовни професор

  
Т. Лутовац

Др Татјана Лутовац, ванредни професор

  
М. Рашајски

Др Марија Рашајски, ванредни професор