

УНИВЕРЗИТЕТ У НИШУ  
ЕЛЕКТРОНСКИ ФАКУЛТЕТ

Александра Медведева 14 · Поштански фах 73  
18000 Ниш · Србија  
Телефон 018 529 105 · Телефакс 018 588 399  
E-mail: efinfo@elfak.ni.ac.rs; http://www.elfak.ni.ac.rs  
Текући рачун: 840-1721666-89; ПИБ: 100232259



UNIVERSITY OF NIŠ  
FACULTY OF ELECTRONIC ENGINEERING

Aleksandra Medvedeva 14 · P.O. Box 73  
18000 Niš - Serbia  
Phone +381 18 529 105 · Fax +381 18 588 399  
E-mail: efinfo@elfak.ni.ac.rs  
http://www.elfak.ni.ac.rs

ДЕКАН

31.05.2019. године

О Б А В Е Ш Т Е Њ Е  
НАСТАВНИЦИМА И САРАДНИЦИМА ЕЛЕКТРОНСКОГ ФАКУЛТЕТА

Докторска дисертација кандидата мастер инж. Аделе Љајић под насловом „Обрада негације у кратким неформалним текстовима у циљу побољшања класификације сентимента” и Извештај Комисије за оцену и одбрану докторске дисертације доступни су на увид јавности у електронској верзији на званичној интернет страници Факултета и налазе се у штампаном облику у Библиотеци Електронског факултета у Нишу и могу се погледати до **30.06.2019. године**.

Примедбе на наведени извештај достављају се декану Електронског факултета у Нишу у напред наведеном року.

Председник Наставно-научног већа  
ЕЛЕКТРОНСКОГ ФАКУЛТЕТА У НИШУ

Декан  
*Проф. др Драган Манчић*

## ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

### ПОДАЦИ О КАНДИДАТУ

Презиме, име једног  
родитеља и име      Љајић, Бесим, Адела  
Датум и место рођења      28.08.1982.године, Нови Пазар

### Основне студије

Универзитет      Универзитет у Београду  
Факултет      Факултет организационих наука  
Студијски програм      Информациони системи  
Звање      Дипломирани инжењер организационих наука – одсек за информационе системе  
Година уписа      2001  
Година завршетка      2007  
Просечна оцена      8,40

### Мастер студије, магистарске студије

Универзитет      Универзитет у Београду  
Факултет      Факултет организационих наука  
Студијски програм      Информациони системи и технологије  
Звање      Мастер инжењер организационих наука  
Година уписа      2007  
Година завршетка      2009  
Просечна оцена      10  
Научна област  
Наслов завршног рада      SMS мобилни сервиси као подршка образовању

### Докторске студије

Универзитет      Универзитет у Нишу  
Факултет      Електронски факултет  
Студијски програм      Електротехника и рачунарство (модул: Рачунарство и информатика)  
Година уписа      2012  
Остварен број ЕСПБ бодова      Укупно 140 ЕСПБ бодова и то: 60 ЕСПБ бодова који су стечени полагањем испита из наставних предмета и 80 ЕСПБ бодова стечених истраживачким радом  
Просечна оцена      10.00

### НАСЛОВ ТЕМЕ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Наслов теме докторске  
дисертације      Обрада негације у кратким неформалним текстовима у циљу побољшања  
класификације сентимента  
Име и презиме ментора,  
звање      Сузана Стојковић, ванредни професор  
Број и датум добијања  
сагласности за тему  
докторске дисертације      8/20-01-009/18-021 od 14.11.2018.

### ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Број страна      109  
Број поглавља      10  
Број слика (шема, графикона)      7  
Број табела      25  
Број прилога      1



**ПРИКАЗ НАУЧНИХ И СТРУЧНИХ РАДОВА КАНДИДАТА  
који садрже резултате истраживања у оквиру докторске дисертације**

P. бр.	Аутор-и, наслов, часопис, година, број волумена, странице	Категорија
1	Ljajić, A., Marovac, U., “Improving Sentiment Analysis for Twitter Data by Handling Negation Rules in the Serbian Language”, <i>Computer Science and Information Systems</i> , Vol. 16, No. 1, 289-311. (2019), <a href="https://doi.org/10.2298/CSIS180122013L">https://doi.org/10.2298/CSIS180122013L</a>	M23
2	Ljajić Adela, Marovac Ulfeta, Stanković Milena, “Comparison of the Influence of Different Normalization Methods on Tweet Sentiment Analysis in the Serbian language”, <i>Facta Universitatis</i> , Series: Mathematics and Informatics, 33(5), (2018), 683-696.	M51
3	Ulfeta Marovac, Aldina Avdić, Adela Ljajić, AnidaFazlagć, PRIMENA CROWDSOURCING-A U DETEKCIJI PSIHOLOŠKIH PROBLEMA, <i>Book of proceedings</i> , 5th International conference on Knowledge management and informatics, Kopaonik, 08.-09. January, 2019 , pp 390- 398.	M33
4	Ulfeta Marovac, Adela Ljajić, AldinaAvdić, AnidaFazlagć, Automation of psychological testing of stressful situations in the Serbian language, 9th International Conference on Information Society and Techology, Kopaonik, Serbia on Mar 10-13, 2019.	M33
5	Marovac Ulfeta, Ljajić Adela, Kajan Ejub, Avdić Aldina, Towards the Lexical Resources for Sentiment-Reach Informal Texts-The Serbian Language Case, 5th International Conference CONTEMPORARY PROBLEMS OF MATHEMATICS, MECHANICS AND INFORMATICS (CPMMI 2018), State University of Novi Pazar, (2018).	M34
6	Ljajić Adela, Stanković Milena, Marovac Ulfeta, “Detection of Negation in the Serbian language”, <i>Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics</i> , WIMS’18, Novi Sad, Serbia, June 25-27 2018, ACM, ISBN 978-1-4503-5489-9/18/06, doi: 10.1145/3227609.3227660.	M33
7	Ljajić Adela, Marovac Ulfeta, Avdić Aldina, Processing of Negation in Sentiment Analysis for the Serbian Language, <i>Proceedings of 4th International Conference on Electrical, Electronics and Computing Engineering</i> , IcETRAN 2017, Kladovo, Serbia, June 05-08, ISBN 978-86-7466-692-0, pp. RTI2.6.1-5.	M33
8	Ljajić Adela, Marovac Ulfeta, Avdić Aldina, “Sentiment Analysis of Twitter for the Serbian Language”, 7th International Conference on Information Society and Technology ICIST 2017, (2017), pp 86-89.	M33
9	Crnišaniin Adela, Ljajić Ertan, Spalević Petar, Arsić Branko, Vučković Darko, “Sentiment analysis of textual comments in field of sport”, <i>Proceedings of the Twenty-fourth International Electrotechnical and Computer Science Conference ERK</i> , pp. 35 - 38, issn: 1581-4572, Словенија, 21. - 23. Sep. 2015.	M33
10	Arsić Branko, Spalević Petar, Bojić Ljubiša, Crnišaniin Adela, “Social Networks in Logistics System Decision-Making”, <i>Proceedings of the 2nd Logistics International Conference LOGIC 2015</i> , University of Belgrade, Faculty of Transport and Traffic Engineering, vol. , no. , pp. 166 - 171, issn: 978-86-7395-339-7, udc: , doi: , Србија, 21. - 23. May, 2015	M33
11	Crnišaniin Adela, Spalević Petar, Rančić Dejan, Micalović Marko, “Primena metode logistističke regresije za određivanje sličnosti tekstualnih podataka”, XIII međunarodni naučno-stručni simpozijum INFOTEH 2014, Elektrotehnički fakultet, Istočno Sarajevo, vol. 13, no. , pp. 784 - 788, issn: 978-99955-763-3-2, Jahorina, Bosna i Hercegovina, 19. Mar - 21. Dec, 2014.	M33
12	Marovac Ulfeta, Crnišaniin Adela, Pljasković Aldina, Kajan Ejub, “Similarity Search in Text Data for Serbian Language”, <i>Proceedings of ICEST (2013)</i> , pp. 607 – 610, ISBN: 978-9989-786-90-7, (2013).	M33
13	Crnišaniin Adela, Pljasković Aldina, Marovac Ulfeta, Kajan Ejub, “One Solution of Searching Text Documents in Serbian Language”, <i>Proceedings of ICIST (2013)</i> , pp. 28-33, ISBN: 978-86-85525-12-4, (2013).	M33



## ИСПУЊЕНОСТ УСЛОВА ЗА ОДБРАНУ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Кандидат испуњава услове за оцену и одбрану докторске дисертације који су предвиђени Законом о високом образовању, Статутом Универзитета и Статутом Факултета.

ДА НЕ

У извештају Комисије за оцену испуњености критеријума за покретање поступака за пријаву докторске дисертације, покретање поступка за оцену и одбрану докторске дисертације и изборе у звања наставника на Електронском факултету у Нишу, у решењу бр. **07/03-013/19-003, од 13.05.2019. године**, наводи се да кандидат Адела Љајић, мастер инжењер организационих наука, **ИСПУЊАВА** предвиђене критеријуме за покретање поступка за оцену и одбрану докторске дисертације.

## ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Кратак опис појединих делова дисертације (до 500 речи)

Рад је организован у десет поглавља. Прво поглавље је уводно и у њему је дата мотивација за истраживање, описан је предмет истраживања и истакнути су циљеви дисертације.

У другом поглављу дат је кратак увод у област анализе података, са освртом на типове података које методе за анализу података користе. Такође се посебно даје преглед истраживања из области анализе текста, са освртом на процесирање природних језика, са обзиром да се истраживање представљено у дисертацији односи на текстове на српском језику

У трећем поглављу је описано која су генерална знања о језику потребна за његову обраду помоћу рачунара.

Четврто поглавље садржи дефиниције анализе сентимента, задатке анализе сентимента и проблем и изазове ове области. Најпре је дат опис метода за класификацију текста по сентименту и преглед метрика за оцену квалитета метода класификације. У наставку је извршен преглед публикованих истраживања и резултата која се односе на анализу сентимента у текстовима на српском језику. Поглавље се завршава прегледом техника које су коришћене за обраду негације код анализе сентимента у текстовима на другим језицима (енглески, шпански).

У петом поглављу је описана негација у српском језику са лингвистичке тачке гледишта. Дат је преглед врста негација и подела на одговарајуће типове који су даље обрађивани.

Коришћена метода је описана у шестом поглављу: мотивација за примену методе, скуп података, потребна нормализација, утицај негације на речи којима се изражава сентимент, обрада правила негације, структура и функционисање методе. Поредбене методе су такође описане на крају шестог поглавља.

Коришћени језички ресурси су описану у седмом поглављу. Посебан нагласак у поглављу је на речнику сентимената где је урађена и његова валидација различитим начинима нормализације и утицај ове нормализације на одређивање сентимента твитова. Осим речника сентимената, описани су и речник сигнала негације, речник одречних квантификатора, речник стоп речи и речник појачавача.

Тестирање и резултати примене методе за класификацију твитова су приказани у осмом поглављу. Тестирана су два приступа: класификација методом која се заснива на речнику сентимената и класификација методом машинског учења. Пре примене методе машинског учења, тестирана је примена додатних атрибута трансформацијом текста у ветор речи.

Девето поглавље је закључак док десето даје преглед праваца могућег даљег истраживања са предлозима за даље унапређење и проширење методе.

Рад се завршава списком референци којих укупно има 81, као и прилогом са коришћеним скраћеницама.

## ВРЕДНОВАЊЕ РЕЗУЛТАТА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Ниво остваривања постављених циљева из пријаве докторске дисертације (до 200 речи)

Истраживања која су представљена у овој докторској дисертацији обухватају анализу сентимента твитова на српском језику са посебним нагласком на обраду правила синтаксичке негације. Показано је да се обрадом правила негације и њиховом интеграцијом у метод класификације твитова по сентименту може побољшати тачност класификације сентимента кратких текстова – у овом случају твитова.

У дисертацији су коришћене методе класификације базиране на речнику сентимената и базиране на машинском учењу. Постављени циљ је био побољшање методе класификације твитова пре свега обрадом правила негације у српском језику. Показано је да се обрадом ових правила повећава тачност класификације у односу на методе које не обрађују негацију или је обрађују хеуристички.

Како би се подаци припремили за класификацију, коришћена је нормализација која је специфична за



овакав тип текста (твит). Од лексичких ресурса су коришћени речник речи којима се изражава сентимент, речник стоп речи, речник специфичних врста речи за које постоје правила употребе у реченичним конструкцијама која су обрађена код процесирања негације.

Додатна побољшања су добијена креирањем атрибута текста који носе критичну количину информација. Ови атрибути су добијени техникама селекције и редукције атрибута. Овако креирана метода представља оригиналну методу за обраду негације код одређивања сентимента твитова на српском језику.

Један од циљева је био да се обезбеди скуп обележених твитова припремљених за надгледано машинско учење. Такође су и остали језички ресурси коришћени у току израде ове дисертације оригинално креирани. План је да ови ресурси буду јавно доступни за даља истраживање. Такође, планира се и њихово проширење у сарадњи са заинтересованим странама.

Вредновање значаја и научног доприноса резултата дисертације (*до 200 речи*)

У тези је дефинисана методологија за анализу сентимента кратких неформалних текстова на српском језику која укључује специфична правила обраде синтаксичке негације. Може се истаћи да су то прва истраживања која се односе на обраду негација код класификације текстова по сентименту у српском језику. Резултати методе су приказани за два случаја: када се користи приступ заснован на речнику сентимента и приступ који користи машинско учење

Посебно су значајна истраживања која се односе на примену техника машинског учења. Коришћено је више метода машинског учења и експериментално је утврђено да метода полиномијалне логистичке регресије даје најбоље резултате.

Применом разичитих филтера за селектоване атрибута, утврђено је да униграми и униграми у комбинацији са биграмама највише доприносе квалитету класификације. Утврђено је да је најбоље рачунати само појаву речи (без ТФ, ИДФ, бројања броја појава речи...) – што је повезано са типом корпуса.

Значајан утицај код примене методе, и у приступу заснованом на речнику и у приступу коришћењем машинског учења, имају језички ресурси. Поред обавезног речника сентимента, коришћени су и речник сигнала негације, речник одричних квантификатора, речник интензификатора и неутрализатора. Наведени језички ресурси су креирани у току израде дисертације и резултат су рада на њој. Од екстерних језичких ресурса је коришћен само морфолошки речник.

Тестирани су различити начини нормализације и речника сентимента и корпуса твитова: нормализација стемом, нормализација морфолошким речником и нормализација одсецањем на различите дужине н-грама. Утврђено је да нормализације применом стемера даје најбоље резултате класификације по сентименту.

За детаљнију обраду правила негације креиран је корпус твитова који су ручно обележени од стране три особе. Корпус је за потребе анализе негације и детектовања правила синтаксичке негације у фази експеримена коришћен на три начина: у првом случају је коришћен скуп свих твитова, у другом су коришћени само твитови који садрже бар једну негацију (у твиту је присутан бар један негатив-сигнал негације) и у трећем случају је коришћен подскуп твитова који садрже бар једно правило негације које је обрађено. Тестирање методе на овако различитим скуповима је омогућило бољи увид на то колико се увођењем дефинисаних правила негације утиче на квалитет класификације по сентименту.

Значај добијених резултата је већи ако се узме у обзир природа текста од којих се састоји корпус који је коришћен, Твитови су кратки текстови, код којих је садржај (па и сентимент ако постоји) доста сажето изражен, садрже доста неформално написаних речи, скраћеница, садрже и граматичке грешке у писању и речи су често написане без дијакритичких знакова (нпр. “неце више” уместо “неће више”) и слично. Све ово, као и комплексност обраде текста на морфолошки богатом српском језику, отежава процес обраде оваквог типа текста. Додатно, сложеност правила негације, отежава и начин њеног детектовања и интеграције у методу за класификацију твитова по сентименту. Ако се узме све ово у обзир, добијени резултати дају задовољавајућу тачност класификације овако сложених текстова по сентименту. Што је још значајније, побољшање које се добија обрадом одређеног скупа правила синтаксичке негације у односу на традиционалне начине обраде негације значајно доприноси квалитету класификације.



Оцена самосталности научног рада кандидата (до 100 речи)

Кандидат је у току истраживања показао изузетан напредак у научно-истраживачком раду и до најзначајнијих резултата представљених у научним публикацијама и самој дисертацији дошао је самосталним аналитичким сагледавањем проблема, анализом актуелних трендова у области, формирањем корпуса кратких коментара (твитова) на српском језику, постављањем методологије за обраду негација у српском језику, извођењем експеримента за класификацију сентимента, самосталним развојем методологије евалуације, као и анализом и правилном интерпретацијом експерименталних резултата.

ЗАКЉУЧАК (до 100 речи)

Докторска дисертација под насловом *Обрада негације у кратким неформалним текстовима у циљу побољшања класификације сентимента*, кандидата Аделе Љајић, представља значајан допринос у области машинске анализе текстова на српском језику у циљу утврђивања поларитета сентимента који се изражава, са посебним освртом на третирање негација којима се у неким случајевима мења поларитет а у другим појачава поларитет сентимента у тексту.

Комисија закључује да је докторска дисертације под насловом *Обрада негације у кратким неформалним текстовима у циљу побољшања класификације сентимента* научно заснована и предлаже Наставно-научном већу Електронског факултета и Научно-стручном већу за техничко-технолошке науке Универзитета у Нишу да прихвати дисертацију и одобри њену јавну одбрану.




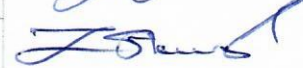

КОМИСИЈА

Број одлуке ННВ о именовану Комисије

8/20-01-004/19-030

Датум именовања Комисије

20.05.2019

Р. бр.	Име и презиме, звање	Потпис
1.	Проф. др Сузана Стојковић, ванредни професор Електротехничко и рачунарско инжењерство – (ужа: Рачунарство и информатика) Електронски факултет, Универзитет у Нишу	ментор, председник 
2.	Проф. др Милена Станковић, редовни професор у пензији Електротехничко и рачунарско инжењерство – (ужа: Рачунарство и информатика) Електронски факултет, Универзитет у Нишу	члан 
3.	Проф. др Драган Јанковић, редовни професор Електротехничко и рачунарско инжењерство – (ужа: Рачунарство и информатика) Електронски факултет, Универзитет у Нишу	члан 
4.	Проф. др Леонид Стоименов, редовни професор Електротехничко и рачунарско инжењерство – (ужа: Рачунарство и информатика) Електронски факултет, Универзитет у Нишу	члан 
5.	Проф. др Ејуб Кајан, ванредни професор Електротехничко и рачунарско инжењерство (ужа: Рачунарство) Државни универзитет у Новом Пазару	члан 

Датум и место:

29.05.2019. године, Ниш

РЕПУБЛИКА СРБИЈА  
УНИВЕРЗИТЕТ У НИШУ  
ЕЛЕКТРОНСКИ ФАКУЛТЕТ У НИШУ

Бр. 07/03-013/19-007

31.05.2019. год.

Ниш, ул. Александра Медведова бр. 14