

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Реферат о урађеној докторској дисертацији кандидата **Вука Батановића**

Одлуком Наставно-научног већа Електротехничког факултета Универзитета у Београду бр. 5045/11-3 од 18.09.2020. године, именовани смо за чланове Комисије за преглед, оцену и одбрану докторске дисертације кандидата Вука Батановића под насловом

**Методологија решавања семантичких проблема у обради кратких текстова
написаних на природним језицима са ограниченим ресурсима**

После прегледа достављене Дисертације и других пратећих материјала и разговора са Кандидатом, Комисија је сачинила следећи

РЕФЕРАТ

1. УВОД

1.1. Хронологија одобравања и израде дисертације

Кандидат Вук Батановић уписао је докторске академске студије на Електротехничком факултету Универзитета у Београду школске 2011/2012. године, на модулу Софтверско инжењерство студијског програма Електротехника и рачунарство. Током студија положио је све испите са просечном оценом 10,00 и одрадио све обавезе везане за студијски истраживачки рад предвиђене планом и програмом.

Предлог теме докторске дисертације под насловом „Методологија решавања семантичких проблема у обради кратких текстова написаних на природним језицима са ограниченим ресурсима“ кандидат је поднео 07.04.2016. године. 12.04.2016. године Комисија за студије трећег степена разматрала је предлог теме за израду докторске дисертације и упутила Наставно–научном већу на усвајање предлог Комисије о оцени подобности теме и кандидата. Наставно–научно веће Електротехничког факултета Универзитета у Београду је на седници одржаној 21.04.2016. године, именovalo Комисију за оцену услова и прихватање теме докторске дисертације (одлука бр. 5045/11-1 од 28.04.2016.) у саставу:

1. др Драган Бојић, ванредни професор, Универзитет у Београду – Електротехнички факултет
2. др Зоран Шеварац, доцент, Универзитет у Београду – Факултет организационих наука
3. др Горан Квашчев, доцент, Универзитет у Београду – Електротехнички факултет

За менторе докторске дисертације предложени су др Бошко Николић, редовни професор, Универзитет у Београду – Електротехнички факултет, и др Милош Цветановић, доцент, Универзитет у Београду – Електротехнички факултет.

Кандидат је предложеној тему докторске дисертације усмено одбранио пред Комисијом за оцену услова и прихватање теме докторске дисертације на јавној одбрани одржаној 17.05.2016. године. Наставно-научно веће Електротехничког факултета Универзитета у Београду је усвојило извештај Комисије за оцену услова и прихватање теме докторске дисертације (одлука бр. 5045/11-2 од 05.07.2016. године). Веће научних области техничких наука Универзитета у Београду је на седници одржаној 19.09.2016. године дало сагласност на тему докторске дисертације (одлука бр. 61206-4367/2-16 од 19.09.2016. године).

Кандидат је 27.08.2020. године предао докторску дисертацију на преглед и оцену. Комисија за студије трећег степена потврдила је 01.09.2020. године испуњеност потребних услова за подношење предлога Наставно-научном већу Електротехничког факултета за формирање Комисије за преглед и оцену докторске дисертације.

Наставно-научно веће Електротехничког факултета Универзитета у Београду је на седници одржаној 08.09.2020. године именovalo Комисију за преглед и оцену докторске дисертације (одлука бр. 5045/11-3 од 18.09.2020. године) у саставу:

1. др Бошко Николић, редовни професор, Универзитет у Београду – Електротехнички факултет
2. др Милош Цветановић, ванредни професор, Универзитет у Београду – Електротехнички факултет
3. др Драган Бојић, ванредни професор, Универзитет у Београду – Електротехнички факултет
4. др Зоран Шеварац, ванредни професор, Универзитет у Београду – Факултет организационих наука
5. др Дражен Драшковић, доцент, Универзитет у Београду – Електротехнички факултет

На основу одлуке Наставно-научног већа бр. 545/2 од 13.03.2012. године, студијски програм је започео у пролећном семестру школске 2011/2012. године, па се рок за завршетак докторских академских студија рачуна од почетка тог семестра, сагласно Статуту Универзитета у Београду и Статуту Електротехничког факултета. На основу члана 101. Статута Универзитета у Београду, члана 74. Статута Универзитета у Београду–Електротехничког факултета и захтева студента, одобрено је продужење рока за завршетак студија до истека троструког броја школских година потребних за реализацију уписаног студијског програма.

1.2. Научна област дисертације

Докторска дисертација припада области техничких наука, научној области Електротехника и рачунарство, ужој научној области Софтверско инжењерство, за коју је матичан Електротехнички факултет Универзитета у Београду.

Ментори др Бошко Николић, редовни професор Електротехничког факултета Универзитета у Београду, и др Милош Цветановић, ванредни професор Електротехничког факултета Универзитета у Београду, испуњавају све формалне и законске услове за менторе ове дисертације. Аутори су већег броја радова у истакнутим међународним часописима који су у вези са предметом докторске дисертације. Релевантни радови ментора наведени су приликом пријаве теме докторске дисертације овог кандидата.

1.3. Биографски подаци о кандидату

Вук Батановић је рођен 02.06.1987. у Београду. Основне академске студије на Електротехничком факултету Универзитета у Београду је уписао 2006. године. Дипломирао је на модулу Рачунарска техника и информатика 2010. године, са просечном оценом 9,56. Мастер академске студије на истом модулу Електротехничког факултета је уписао 2010. и завршио 2011. године, са просечном оценом 10. У пролећном семестру школске 2011/2012. године уписао је докторске академске студије на модулу Софтверско инжењерство истог факултета. Испите на докторским студијама је положио са просечном оценом 10. У току студија усавршавао се присуствовањем већем броју летњих школа из обраде природних језика и машинског учења.

Сарађивао је на међународном пројекту *Regional Linguistic Data Initiative (ReLDI)* и словеначком пројекту *CLARIN.SI (Common Language Resources and Technology Infrastructure)*. Од 2018. године ради као истраживач-сарадник у Иновационом центру Електротехничког факултета у Београду. Био је ангажован на пројекту Програма Уједињених нација за развој (UNDP) о аутоматизацији семантичке претраге правних докумената на српском језику. Од школске 2017/2018. године је ангажован у настави на мастер академским студијама Електротехничког факултета, на предметима Обрада природних језика и Проналажење скривеног знања.

Научна интересовања су му усмерена на решавање семантичких проблема из обраде природних језика, пре свега применом метода машинског учења. Нарочито га интересују специфичности обраде кратких текстова и решења примењива и у језицима са ограниченим ресурсима. Објавио је 16 научних публикација, од чега 2 рада у међународним часописима (M20), 2 рада у часописима националног значаја (M50), 7 радова на међународним (M30) и 3 на домаћим конференцијама (M60) и 2 техничка решења (M80).

2. ОПИС ДИСЕРТАЦИЈЕ

2.1. Садржај дисертације

Докторска дисертација кандидата Вука Батановића под насловом „Методологија решавања семантичких проблема у обради кратких текстова написаних на природним језицима са ограниченим ресурсима“ написана је у складу са Упутством за обликовање докторске дисертације Универзитета у Београду. Дисертација је писана на српском језику и латиничном писму и садржи 123 стране, 15 слика, 26 табела и 181 библиографску референцу. Подељена је у следећих седам поглавља:

1. Увод
2. Процес развоја статистичких решења у семантичкој обради природних језика
3. Прикупљање текстуалног садржаја
4. Анотација података
5. Формулисање, обучавање и евалуација модела
6. Преглед методологије развоја статистичких решења семантичких проблема у природним језицима са ограниченим ресурсима
7. Закључак

На почетку дисертације се налазе насловне стране на српском и енглеском језику, страна са информацијама о менторима и члановима комисије, кратки резимеи дисертације на српском и на енглеском језику, захвалнице и садржај. На крају дисертације су дати прилози, списак

коришћене литературе, спискови слика, табела и скраћеница, биографија аутора, изјаве о ауторству, о истовестности штампане и електронске верзије докторског рада и о коришћењу.

2.2. Кратак приказ појединачних поглавља

Уводно поглавље дисертације даје кратак преглед развоја области обраде природних језика и описује неравномерност тог развоја по језицима. Након тога је дата дефиниција језика са ограниченим ресурсима и дискутована је шира неприменљивост решења развијених за велике светске језике попут енглеског на језике са ограниченим ресурсима. Затим је описан главни методолошки циљ дисертације као и проблеми одређивања семантичке сличности и анализе сентимента кратких текстова кроз чије ће решавање методологија бити приказана. Коначно, изнети су и ужи циљеви дисертације и полазне хипотезе које су коришћене у раду.

У другом поглављу је приказан начелни процес развоја статистичких решења у семантичкој обради природних језика, и дискутована је свака од фаза у овом процесу. Затим су детаљно размотрена постојећа статистичка решења која се користе у семантичкој обради језика са ограниченим ресурсима, са акцентом на методе трансферног учења. На крају поглавља је дат упоредни преглед по фазама развоја за решења семантичких проблема која су креирана у оквиру дисертације.

Треће, четврто и пето поглавље описују по једну фазу развоја решења за проблеме одређивања семантичке сличности и анализе сентимента кратких текстова на српском језику. Треће поглавље се односи на фазу прикупљања текстуалног садржаја, четврто на фазу анотације података, а пето на фазу формулисања, обучавања и евалуације модела. У сва три поглавља најпре је приказан развој за проблем одређивања семантичке сличности, а затим за проблем анализе сентимента. Овакав паралелни приказ омогућава јасно уочавање методолошких сличности и разлика између приступа коришћених за решавање посматрана два проблема.

У трећем поглављу најпре је приказана изградња корпуса парова реченица *STS.news.sr* за потребе одређивања семантичке сличности кратких текстова, а затим и прикупљање текстуалног садржаја за израду корпуса *SentiComments.SR*, креираног за потребе анализе сентимента кратких текстова на српском језику. За оба проблема су такође представљени постојећи референтни скупови података на енглеском језику, а дискутовано је и постојање, доступност и квалитет раније направљених скупова овог типа на српском језику.

Четврто поглавље описује анотацију корпуса текстова приказаних у претходном поглављу. Најпре су приказани постојећи системи означавања података за посматране проблеме и анализирана је њихова адекватност и квалитет. За проблем одређивања семантичке сличности усвојен је међународно стандардизован систем означавања података, док је за проблем анализе сентимента развијен потпуно нов систем означавања. За оба проблема дата су коришћена упутства за анотацију, као и опис спровођења анотације и статистичка анализа резултата. У оквиру анализе анотације сентимента текстова, развијена је нова метрика за одређивање економичности употребе упутстава за анотацију која је на овом проблему и примењена. Коначно, добијени аниотирани скупови података, односно системи означавања, упоређени су са ранијим сличним решењима представљеним на другим језицима.

У петом поглављу је приказано формулисање, обучавање и евалуација већег броја статистичких модела за оба посматрана проблема. Основни модели и примена финог подешавања неуралних језичких модела су описани за оба проблема. Поред тога, за проблем одређивања семантичке сличности кратких текстова развијена су и три нова наменска

модела. Сви описани модели су евалуирани коришћењем погодних метрика и њихове перформансе су упоређене и продискутоване.

Шесто поглавље даје преглед методологије развоја статистичких решења семантичких проблема у обради природних језика са ограниченим ресурсима. У овом поглављу су на основу искустава и резултата добијених у изради решења за посматране проблеме одређивања семантичке сличности и анализе сентимента представљене опште методолошке смернице применљиве и на друге семантичке проблеме или на друге природне језике.

У завршном, седмом поглављу, изнети су закључци рада и кључни доприноси. Поред тога, предложени су и могући правци даљих истраживања.

3. ОЦЕНА ДИСЕРТАЦИЈЕ

3.1. Савременост и оригиналност

Обрада природних језика је област на раскршћу вештачке интелигенције и машинског учења, лингвистике, математичког моделовања и рачунарских наука у ширем смислу, која је све актуелнија и развија се убрзаним темпом. Количина садржаја доступних у текстуалном облику је експоненцијално порасла ширењем интернета, а важност обраде кратких текстова је додатно добила на значају ширењем друштвених мрежа и Web 2.0 концепта. Знатна комерцијална применљивост учинила је да проблеми семантичке обраде текстова буду међу најпопуларнијим у актуелним истраживањима, што важи и за конкретне семантичке проблеме разматране у овој дисертацији. Семантичка сличност представља један од фундаменталних проблема у семантичкој обради природних језика, и представља интегрални део других задатака, попут израде сажетака текстова, одговарања на питања, дохватања информација, итд. Са друге стране, због велике потребе за аутоматском детекцијом и обрадом ставова људи, анализа сентимента спада у проблеме обраде природних језика са највећом директном практичном применом.

У оквиру ове дисертације представљена је нова методологија развоја статистичких решења семантичких проблема прилагођена језицима са ограниченим ресурсима. Поред тога, развијени су потпуно нови системи означавања података и процене економичности анотације, израђени су нови наменски модели за одређивање семантичке сличности кратких текстова, и креирани су први референтни јавно доступни анотирани скупови података на српском језику за посматрана два семантичка проблема. Наведени резултати демонстрирају висок степен оригиналности рада кандидата и у потпуности прате савремене токове научних истраживања.

3.2. Осврт на референтну и коришћену литературу

Дисертација садржи врло темељан приказ референтне и актуелне литературе из области обраде природних језика, са 181 библиографском јединицом. Коришћена литература се начелно може поделити на:

- Референце које су везане за уже проблеме одређивања семантичке сличности или анализе сентимента кратких текстова
- Референце које се баве методама трансферног учења, укључујући савремене методе засноване на неуралним мрежама
- Референце које се односе на различите приступе анотацији података
- Референце које приказују раније развијена решења везана за обраду српског језика, укључујући оне који описују методе морфолошке нормализације текстова на српском

У литературу су убројани како кључни референтни радови везани за посматране проблеме обраде природних језика, тако и они најактуелнији, објављени у последњих годину дана на реномираним међународним научним конференцијама и у часописима. Кандидат је у литератури обухватио и референце свог истраживачког рада током докторских студија и писања дисертације које су директно повезане са тематиком дисертације. Списак референци које је кандидат користио показује да је спроведена детаљна и свеобухватна анализа постојеће научне литературе и да су наведене све значајне релевантне библиографске јединице.

3.3. Опис и адекватност примењених научних метода

Научне методе истраживања примењене у овој дисертацији су:

- Процена и поређење постојећих решења семантичких проблема у обради кратких текстова написаних на природним језицима, са акцентом на њихову применљивост на језике са ограниченим ресурсима.
- Ручна анотација података – због ограничене доступности потребних ресурса на српском језику и природе самих проблема који се разматрају није било могуће аутоматски прикупити анотирани податке који су потребни за обучавање и евалуацију модела, те је било неопходно да се прикупљени текстуални подаци ручно анотирају. Ради спровођења анотације установљен је систем критеријума за означавање. На крају анотације експериментално је утврђен квалитет добијених ознака на основу степена сагласности оцена анотатора, а разматрана је и брзина и економичност анотације.
- Софтверска имплементација – сви модели и алати представљени у овој дисертацији су имплементирани у виду софтверских пакета, пре свега у програмским језицима Java и Python. При томе су, по потреби, коришћене и већ постојеће софтверске библиотеке имплементиране у овим или другим програмским језицима.
- Експериментална евалуација – провера квалитета предложених алгоритама и различитих вредности њихових параметара је спроведена експериментално, коришћењем метрика за мерење перформанси које су у литератури најраширеније за посматрану проблематику.

Коришћене научне методе представљају комбинацију теоријских и практичних приступа и у потпуности одговарају тематици и циљевима дисертације.

3.4. Применљивост остварених резултата

Резултати дисертације се могу поделити на оне са широком и оне са ужом применљивошћу. Предложена методологија развоја статистичких решења семантичких проблема је применљива на широк спектар језика са ограниченим ресурсима. Исто важи и за креирани систем означавања сентимента кратких текстова и нове наменске моделе за одређивање семантичке сличности, док се представљена метрика за утврђивање економичности анотације може употребити на произвољном природном језику и у оквиру различитих задатака семантичке анотације. Сви наведени резултати се могу директно применити на даља истраживања у области обраде природних језика, и то на великом броју језика.

Са друге стране, у резултате везане конкретно за српски језик спадају израда првих референтних скупова анотираних података за посматране проблеме, израда и евалуација основних и неуралних модела за решавање ових проблема на српском језику, и компаративна евалуација ефеката техника морфолошке нормализације на српском језику. Ови резултати дисертације су практично примењиви у погледу даљег развоја рачунарске обраде српског језика.

3.5. Оцена достигнутих способности кандидата за самостални научни рад

У изради дисертације кандидат је најпре анализирао постојећа решења у обради природних језика, проценио њихову адекватност за језике са ограниченим ресурсима и идентификовао најпогодније приступе. Након тога је успешно креирао решења за проблеме одређивања семантичке сличности и анализе сентимента кратких текстова на српском језику, почевши од прикупљања текстуалног садржаја, преко анотације података, до формулисања, израде и евалуације статистичких модела. Развој конкретних статистичких решења је спроведен у контексту ширег методолошког оквира за решавање семантичких проблема у језицима са ограниченим ресурсима. Закључено је да добијени резултати дисертације имају како ширу применљивост на знатан број језика са ограниченим ресурсима, тако и приметну вредност у даљем развоју рачунарске обраде српског језика. Кандидат је препознао важност и актуелност теме истраживања и показао креативност и иновативност у научном раду, што је и верификовано већим бројем публикација на релевантним научним конференцијама и у научним часописима. На основу свега наведеног, Комисија сматра да је кандидат показао висок степен способности за самостални научни рад.

4. ОСТВАРЕНИ НАУЧНИ ДОПРИНОС

4.1. Приказ остварених научних доприноса

У оквиру докторске дисертације остварени су следећи научни доприноси:

1. Предложена је методологија статистичког решавања семантичких проблема у обради кратких текстова која је прилагођена природним језицима са ограниченим ресурсима и омогућава лакши и боље планиран развој таквих решења.
2. Идентификовани су постојећи приступи за решавање семантичких проблема у обради кратких текстова који су погодни за примену у језицима са ограниченим ресурсима, уз посебно разматрање трансферног учења, као начелно најадекватнијег приступа у овом контексту.
3. Представљена је нова метрика за утврђивање економичности анотације тј. за мерење економичности употребе упутстава за анотацију у оквиру одређеног система означавања података.
4. Израђен је нов, флексибилан и економичан систем означавања сентимента текстова, који омогућава више нивоа интерпретације за ознаке сентимента и посебно је погодан за примену у језицима са ограниченим ресурсима.
5. Створени су први референтни и јавно доступни аотирани скупови података за проблеме одређивања семантичке сличности и анализе сентимента кратких текстова на српском језику.
6. Креирана су и евалуирана основна референтна решења за одређивање семантичке сличности и анализу сентимента кратких текстова на српском језику, и размотрене су њихове варијанте и унапређења.
7. Спроведена је компаративна евалуација ефеката техника морфолошке нормализације текстова на српском језику – стемовања и лематизације – и различитих решења развијених за ове потребе на проблемима одређивања семантичке сличности и анализе сентимента кратких текстова.
8. Коришћењем идеје тежинског пондерисања сличности речи на основу њихове учесталости и/или врсте речи, развијена су и евалуирана три нова наменска модела за одређивање семантичке сличности кратких текстова, који су адекватни за употребу у језицима са ограниченим ресурсима, и који постижу боље перформансе од основних модела и њихових унапређења.

9. Помоћу креираних аотираних скупова података успешно је спроведено фино прилагођавање и евалуација неуралних језичких модела заснованих на *Transformer* архитектурама на посматраним проблемима на српском језику.

4.2. Критичка анализа резултата истраживања

Имајући у виду предмет истраживања, полазне хипотезе и циљеве дисертације, као и добијене резултате, Комисија констатује да је кандидат успешно одговорио на постављене изазове и да резултати оправдавају почетна очекивања. Сходно њиховој применљивости, доприноси дисертације се могу начелно поделити у две групе – оне који представљају унапређење ширих научних знања, и оне који су везани конкретно за напредак у решавању посматраних семантичких проблема на српском језику.

Главни шири допринос дисертације је представљена методологија развоја статистичких решења семантичких проблема, која се бави не само проблемима израде статистичких модела у језицима са ограниченим ресурсима, већ и питањима прикупљања и аотације података, која су доста запостављена у актуелној научној литератури. Предложена метрика за утврђивање економичности аотације је прва широко и лако применљива метрика овог типа у литератури, јер су сви ранији предлози овог типа били уско везани за одређене проблеме у обради природних језика или су захтевали врло специфичне и тешко добављиве податке везане за процес аотације. Креирани систем означавања сентимента кратких текстова се издваја од раније предложених пре свега по флексибилности ознака коју пружа, јер се једне исте ознаке сентимента могу на различите начине редуковати на мањи скуп ознака, чиме се избегава потреба за поновним спровођењем аотације за сваки ужи задатак у анализи сентимента. Коначно, нови наменски модели за одређивање семантичке сличности текстова постижу боље перформансе од основних решења, при чему не захтевају потпуну синтактичку анализу текста.

Поред наведених доприноса, резултати дисертације представљају знатан помак и у погледу развоја рачунарске обраде српског језика. Наиме, израда првих референтних, јавно доступних аотираних скупова података на српском језику за проблеме одређивања семантичке сличности и анализе сентимента кратких текстова је од великог значаја, јер ће олакшати даљи развој ових области у рачунаској обради српског језика и омогућити каснију упоредну евалуацију старих и нових решења. Такође, резултати евалуације основних решења за ове проблеме на српском језику, као и најсавременијих неуралних језичких модела заснованих на *Transformer* архитектурама, ће у овом контексту бити важна реперна тачка у даљим истраживањима. Поред тога, у дисертацији је спроведена прва компаративна евалуација метода за морфолошку нормализацију текстова на српском језику, чији ће закључци у будућности омогућити лакши одабир оптималних метода за морфолошку нормализацију кратких текстова.

4.3. Верификација научних доприноса

Кандидат Вук Батановић је објавио следеће радове који су у вези са темом докторске дисертације:

Категорија M21a:

1. Bojan Furlan, **Vuk Batanović**, Boško Nikolić, “Semantic similarity of short texts in languages with a deficient natural language processing support,” *Decision Support Systems*, vol. 55, no. 3, pp. 710–719, 2013. DOI: 10.1016/j.dss.2013.02.002, ISSN: 0167-9236, **IF (2012) = 2.201**, rang: 6/79

Kategorija M23:

1. **Vuk Batanović**, Dragan Bojić, "Using Part-of-Speech Tags as Deep-Syntax Indicators in Determining Short-Text Semantic Similarity," *Computer Science and Information Systems*, vol. 12, no. 1, pp. 1–31, 2015. DOI: 10.2298/CSIS131127082B, ISSN: 1820-0214 (Print), 2406-1018 (Online), IF (2015) = 0.623, rang: 83/106

Kategorija M33:

1. **Vuk Batanović**, Miloš Cvetanović, Boško Nikolić, "Fine-grained Semantic Textual Similarity for Serbian," in *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, pp. 1370-1378, 2018
2. **Vuk Batanović**, Boško Nikolić, "Sentiment Classification of Documents in Serbian: The Effects of Morphological Normalization," in *Proceedings of the 24th Telecommunications Forum (TELFOR 2016)*, Belgrade, Serbia, 2016. DOI: 10.1109/TELFOR.2016.7818923, ISBN: 978-1-5090-4087-2 (Print), 978-1-5090-4086-5 (Online)
3. **Vuk Batanović**, Boško Nikolić, Milan Milosavljević, "Reliable Baselines for Sentiment Analysis in Resource-Limited Languages: The Serbian Movie Review Dataset," in *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia, pp. 2688–2696, 2016

Kategorija M52:

1. **Vuk Batanović**, Boško Nikolić, "Sentiment Classification of Documents in Serbian: The Effects of Morphological Normalization and Word Embeddings," *Telfor Journal*, vol. 9, no. 2, pp. 104–109, 2017. DOI: 10.5937/telfor1702104B, ISSN: 1821-3251 (Print), 2334-9905 (Online)
2. **Vuk Batanović**, Dragan Bojić, "Evaluation and Classification of Syntax Usage in Determining Short-Text Semantic Similarity," *Telfor Journal*, vol. 6, no. 1, pp. 64–68, 2014. DOI: 10.5937/telfor1401064B, ISSN: 1821-3251 (Print), 2334-9905 (Online)

Kategorija M63:

1. **Vuk Batanović**, Dragan Bojić, „Evaluacija i klasifikacija korišćenja sintaksnih informacija u određivanju semantičke sličnosti kratkih tekstova“, u *Zborniku radova sa 21. telekomunikacionog foruma (TELFOR 2013)*, Beograd, Srbija, pp. 821–824, 2013. DOI: 10.1109/TELFOR.2013.6716356, ISBN: 978-1-4799-1419-7 (Print), 978-1-4799-1420-3 (Online)
2. **Vuk Batanović**, Bojan Furlan, Boško Nikolić, „Softverski sistem za određivanje semantičke sličnosti kratkih tekstova na srpskom jeziku“, u *Zborniku radova sa 19. telekomunikacionog foruma (TELFOR 2011)*, Beograd, Srbija pp. 1249–1252, 2011. DOI: 10.1109/TELFOR.2011.6143778, ISBN: 978-1-4577-1499-3 (Print), 978-1-4577-1500-6 (Online)

Kategorija M81:

1. **Vuk Batanović**, Bojan Furlan, Boško Nikolić, „Softverski sistem za određivanje semantičke sličnosti kratkog teksta napisanog na srpskom jeziku“, 2013

Kategorija M86:

1. **Vuk Batanović**, Bojan Furlan, Boško Nikolić, „Korpus parafraza srpskog jezika“, 2013

5. ЗАКЉУЧАК И ПРЕДЛОГ

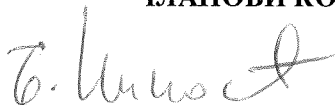
Докторска дисертација кандидата Вука Батановића је у целини написана у складу са образложењем наведеним у пријави теме и садржи све битне елементе који се захтевају Правилником о докторским студијама Електротехничког факултета Универзитета у Београду.

Поред главног методолошког оквира, у склопу дисертације израђен је и анализиран већи број језичких ресурса, алата и модела, представљен је нов систем означавања података и нова метрика за утврђивање економичности анотације. Тиме је остварен значајан научни допринос у развоју статистичких решења семантичких проблема у језицима са ограниченим ресурсима, а такође су у великој мери унапређене могућности решавања посматраних семантичких проблема на српском језику. Стога резултати дисертације укључују како оне са широм применљивошћу на језике са ограниченим ресурсима уопште, тако и оне који су директно применљиви у обради српског језика. Успешним спровођењем истраживања овог обима, чији су резултати верификовани кроз већи број радова у научним часописима и зборницима конференција, кандидат је показао висок степен способности и самосталности у научном раду.

На основу свега изложеног, Комисија предлаже Наставно-научном већу Електротехничког факултета Универзитета у Београду да се докторска дисертација кандидата **Вука Батановића**, мастер инжењера електротехнике и рачунарства, под насловом „**Методологија решавања семантичких проблема у обради кратких текстова написаних на природним језицима са ограниченим ресурсима**“ прихвати, изложи на увид јавности и упуту Већу научних области техничких наука Универзитета у Београду на коначно усвајање и давање одобрења кандидату да приступи усменој одбрани.

у Београду,
20.09.2020. године

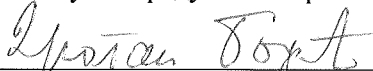
ЧЛАНОВИ КОМИСИЈЕ



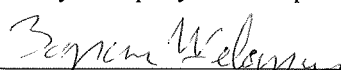
др Бошко Николић, редовни професор
Универзитет у Београду – Електротехнички факултет



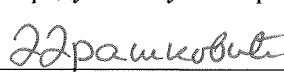
др Милош Цветановић, ванредни професор
Универзитет у Београду – Електротехнички факултет



др Драган Бојић, ванредни професор
Универзитет у Београду – Електротехнички факултет



др Зоран Шеварац, ванредни професор
Универзитет у Београду – Факултет организационих наука



др Дражен Драшковић, доцент
Универзитет у Београду – Електротехнички факултет