

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Извештај о урађеној докторској дисертацији кандидата Николе Пејића, 2020/5029, мастер инжењера електротехнике и рачунарства.

Одлуком бр. 860/27 од 13.05.2025. године, именовани смо за чланове Комисије за преглед, оцену и одбрану докторске дисертације кандидата Николе Пејића 2020/5029 под насловом

Побољшање система за рецензију софтверских измена на основу информација генерисаних из историје рецензија

(енг. *Improving code review systems by generating insights from the review history*)

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

Докторске академске студије на Електротехничком факултету Универзитета у Београду, на модулу за Рачунарску технику и информатику, кандидат Никола Пејић је уписао у октобру 2020. године. Кандидат је положио све испите предвиђене планом и програмом докторских студија са просечном оценом 10. Кандидат је 28.08.2023. године пријавио тему за израду докторске дисертације. За ментора је предложен др Милош Цветановић, ванредни професор Електротехничког факултета Универзитета у Београду.

Катедра за рачунарску технику и информатику, на својој седници одржаној дана 29.08.2023. године, размотрила је пријаву теме докторске дисертације коју је поднео кандидат. Катедра је утврдила да је надлежна за разматрање пријављене теме докторске дисертације, као и да су достављена пријава и њени прилози суштински и формално одговарајући и комплетни.

Комисија за студије трећег степена Електротехничког факултета Универзитета у Београду је на својој седници одржаној дана 05.09.2023. године разматрала пријаву теме за израду докторске дисертације и предлог састава Комисије за оцену научне заснованости теме докторске дисертације. Комисија за студије трећег степена је пријаву теме и предлог састава Комисије за оцену научне заснованости теме докторске дисертације упутила Наставно-научном већу Електротехничког факултета Универзитета у Београду на усвајање.

Наставно-научно веће Електротехничког факултета Универзитета у Београду је, на својој 889. седници одржаној дана 12.09.2023. године, донело одлуку бр. 1290/56 о именовану Комисије за оцену научне заснованости теме докторске дисертације у саставу:

- др Захарије Радивојевић, ванредни професор Универзитета у Београду, Електротехнички факултет
- др Зоран Чича, редовни професор Универзитета у Београду, Електротехнички факултет
- др Сениша Влајић, редовни професор Универзитета у Београду, Факултет организационих наука

За ментора је предложен др Милош Цветановић, ванредни професор Електротехничког факултета Универзитета у Београду.

Јавна усмена одбрана теме докторске дисертације је одржана дана 21.09.2023. године. Комисија за оцену научне заснованости теме докторске дисертације оценила је усмену одбрану као успешну (оцена „задовољно“).

Комисија за студије трећег степена Електротехничког факултета Универзитета у Београду је на својој седници која је одржана дана 03.10.2023. године разматрала записник Комисије за оцену научне заснованости теме докторске дисертације са јавне усмене одбране, који је упутила Наставно-научном већу Електротехничког факултета Универзитета у Београду на усвајање.

Наставно-научно веће Електротехничког факултета Универзитета у Београду је, на својој 890. седници одржаној дана 10.10.2023. године, усвојило извештај Комисије за оцену научне заснованости теме докторске дисертације кандидата, а за ментора је именован др Милош Цветановић, ванредни професор Електротехничког факултета Универзитета у Београду.

Веће научних области техничких наука Универзитета у Београду је, на својој седници одржаној дана 20.11.2023. године, дало сагласност на предложену тему докторске дисертације и именовање ментора (бр. одлуке 61206-4275/2-23).

Кандидат је предао докторску дисертацију на преглед и оцену 28.04.2025. године.

Комисија за студије трећег степена Електротехничког факултета Универзитета у Београду је, на седници одржаној дана 06.05.2025. године, потврдила испуњеност потребних услова за подношење предлога Наставно-научном већу Електротехничког факултета Универзитета у Београду за формирање Комисије за преглед, оцену и одбрану докторске дисертације.

Наставно-научно веће Електротехничког факултета Универзитета у Београду је, на својој 908. седници одржаној дана 13.05.2023. године, именovalo Комисију за преглед, оцену и одбрану докторске дисертације у саставу:

- др Захарије Радивојевић, ванредни професор Универзитета у Београду, Електротехнички факултет
- др Зоран Чича, редовни професор Универзитета у Београду, Електротехнички факултет
- др Сениша Влајић, редовни професор Универзитета у Београду, Факултет организационих наука

1.2. Научна област дисертације

Дисертација припада научној области Електротехника и рачунарство, а ужа научна област дисертације је Рачунарска техника и информатика, док у оквиру уже научне области припада области Софтверског инжењерства. Ментор дисертације је др Милош Цветановић, ванредни професор Електротехничког факултета Универзитета у Београду, који има доприносе у наставном и научном раду у области Рачунарске технике и информатике и Софтверског инжењерства.

1.3. Биографски подаци о кандидату

Никола Пејић рођен је 28. октобра 1996. године у Лесковцу. Електротехнички факултет у Београду уписао је 2014. године. Дипломирао је 2018. године на модулу Рачунарска техника и информатика са просечном оценом 9,69. Дипломски рад под насловом „Креирање статистика екстерних табела лоцираних на кластеру” израдио је под менторством ван. проф. др Милоша Цветановића и одбранио га са оценом 10.

Мастер академске студије уписао је 2018. године на Електротехничком факултету у Београду на модулу Рачунарска техника и информатика, и завршио их је 2019. године са просечном оценом 10,00. Мастер рад под насловом „Процена сличности процедура преведених различитим компајлерима применом неуралних мрежа” израдио је под менторством ван. проф. др Милоша Цветановића и одбранио га је са оценом 10. Докторске академске студије уписао је 2020. године на Електротехничком факултету у Београду. Све испите предвиђене студијским програмом положио је са оценом 10. У току докторских студија објавио је 2 рада у међународним часописима са *SCI* листе.

Од 2019. године ради као софтверски инжењер у развојном центру компаније *Microsoft* у Београду. Током своје каријере био је део *SQL* тима, тима за мешовиту стварност (*Mixed Reality*) и примењене науке (*Applied Sciences Group*), а данас је део *Azure DevOps* тима.

2. ОПИС ДИСЕРТАЦИЈЕ

2.1. Садржај дисертације

Дисертација је написана на српском језику ћириличним писмом и има 152 стране од чега је 130 нумерисано. Дисертација садржи 38 слика, 20 табела, и 15 исечака програмског кода. Дисертација је подељена на 8 делова (7 поглавља и један додатак):

1. Увод
2. Развој софтвера
3. Утицај *COVID-19* пандемије на *GitHub*
4. Претраживање експерата и предлога измена
5. Подсећање рецензената
6. Системи за препоруку рецензената
7. Закључак
- A. Системи за колаборативни развој и рецензију софтвера

Додатно дисертација садржи и насловне стране на српском и енглеском језику, страну са информацијама о ментору и члановима комисије за преглед и оцену, сажетак на српском и енглеском језику, садржај, списак литературе са 111 референци наведених по редоследу појављивања у тексту, списак слика, списак табела, списак исечака програмског кода, биографију аутора и потребне изјаве (о ауторству, о истоветности штампане и електронске верзије докторског рада и о коришћењу).

2.2. Кратак приказ појединачних поглавља

Осим уводног поглавља, дисертација има још 7 поглавља. У уводном поглављу укратко је представљена мотивација која стоји иза дисертације, уз дефинисање почетних хипотеза и научних доприноса истраживања, а затим је дат и преглед структуре дисертације.

Друго поглавље садржи преглед историје развоја и рецензије софтвера од својих зачетака до модерног процеса рецензије софтвера. Представљено је и неколико модерних сервиса за колаборативни развој и рецензије софтвера, као и начини за дохватање података са поменутих платформи.

Треће поглавље садржи истраживање које се бави анализом утицаја *COVID-19* пандемије на развој програма отвореног кода (*open source*) на платформи *GitHub*. Након кратког прегледа развоја пандемије на светском нивоу и претходних истраживања која су се бавила овом темом, представљен је начин за приступање релевантној телеметрији са *GitHub* сервиса користећи *GH Archive* и *BigQuery*. Затим је вршена анализа трендова различитих догађаја из телеметрије како би се утврдило да ли је дошло до значајнијих промена услед пандемије.

Четврто поглавље бави се екстракцијом информација из претходних предлога измена. Представљен је систем који извучи кључне речи из описа претходних измена и одређује њихове тежинске факторе, како би уз њихову помоћ омогућио претраживање инжењера. Систем је имплементиран и учињен доступним групи инжењера из индустрије током периода од десет месеци, а затим је анализирана телеметрија о коришћењу имплементiranог система.

Пето поглавље садржи опис система за подсећање рецензената на рецензије које чекају на њих, уз анализу резултата примене таквог система над скупом инжењера/рецензената једног репозиторијума током периода од пет месеци. Анализирани су резултати две анкете, где се једна бавила испитивањем потребе и преференци корисника пре имплементације система, док се друга бавила испитивањем сентимента корисника након периода коришћења, а такође је анализирана и телеметрија о раду система.

Шесто поглавље бави се системима за препоруку рецензената и даје преглед неколико постојећих приступа. Затим је представљена нова техника за побољшање перформанси постојећих приступа која их фокусира само на релевантне рецензенте, где се релевантност одређује на основу више различитих критеријума. Мерења перформанси вршена су над историјским подацима неколико репозиторијума из индустрије.

Седмо поглавље је закључак који даје преглед главних резултата дисертације, и даје предлоге за потенцијалне правце даљег рада. На самом крају је додаток који у осмом поглављу описује још два популарна система за колаборативни развој и рецензију софтвера.

3. ОЦЕНА ДИСЕРТАЦИЈЕ

3.1. Савременост и оригиналност

Савременост дисертације огледа се у томе да се бави процесом рецензије софтверског кода који је саставни део софтверске индустрије. Дисертација представља неколико приступа екстракцији информација из историје рецензија и њихову примену зарад једноставнијег и бољег искуства софтверских инжењера. Такође је урађена анализа утицаја *COVID-19* пандемије на развој програма отвореног кода на платформи *GitHub* и то током дужег периода него што је то раније учињено. Коначно, представљен је проблем препоруке рецензената софтверских измена, и потешкоће које постојећи приступи имају у случају репозиторијума са великим бројем инжењера. Представљена је нова техника која фокусира системе за

препоруку само на релевантне рецензенте, и испробано је неколико различитих критеријума за филтрирање нерелевантних рецензената. Такође, значај и актуелност ове теме огледа се и у тренутним истраживањима о чему сведоче прикупљене референце у дисертацији.

3.2. Осврт на референтну и коришћену литературу

У дисертацији је наведено 111 референци које су нумерисане према редоследу појављивања у тексту. Референце обухватају базичне као и најновије научне радове који су у вези са темом докторске дисертације, а који су објављени у међународним часописима и зборницима радова међународних конференција, што потврђује значај, релевантност и савременост теме. Такође, међу референцама су наведени и радови које је кандидат објавио као аутор верификујући доприносе дисертације.

3.3. Опис и адекватност примењених научних метода

Истраживање у оквиру докторске дисертације засновано је на теоријским, емпиријским и математичким методама. Ове методе укључују систематично прикупљање, анализу и интерпретацију података, имплементацију система и анализу телеметрије о њиховом коришћењу, спровођење анкета и анализу резултата, као и тумачење резултата научно-стручним методама. Методе истраживања у докторској дисертацији обухватају проучавање процеса развоја и рецензије софтверског кода, као и функционалности тренутно најпопуларнијих платформи за колаборативни развој софтвера. Анализом стационарности и линеарном регресијом испитан је дугорочни утицај *COVID-19* пандемије на развој програма отвореног кода на платформи *GitHub*. Спровођењем упитника и интерпретацијом њихових резултата уз анализу телеметрије система, испитана је валидност представљеног система за претрагу експерата, као и система за подсећање рецензената на њихове рецензије. На крају, систематичном анализом постојећих метода за препоруку рецензената уочен је тренд лошијих перформанси у случају великих репозиторијума са мноштвом инжењера. Представљена је техника за филтрирање мање релевантних инжењера, и применом ове технике над историјским подацима који су прикупљени за неколико репозиторијума из индустрије закључено је да техника може побољшати перформансе постојећих метода.

Примењене методе су у потпуности примерене проблемима који су решавани у дисертацији, што је довело до остварења циљева дисертације.

3.4. Применљивост остварених резултата

Истраживања из докторске дисертације директно су инспирисана тренутним стањем софтверске индустрије и искуством рада у њој. Неколико система је имплементирано и активно коришћено у индустрији, и вршена је анализа реалног коришћења датих система, као и анализа резултата анкета које су спроведене над инжењерима из индустрије, како би се верификовала корист представљених система. Такође, анализирани су историјски подаци из индустрије, и техника за побољшање система за препоруку рецензената би зато била директно применљива у индустрији.

3.5. Оцена достигнутих способности кандидата за самостални научни рад

Током докторских студија кандидат је детаљно истражио и прегледао литературу из области развоја и рецензије софтверског кода, те је показао способност да сагледа тренутно стање области, предности и недостатке постојећих решења. Током истраживања теме докторске

дисертације кандидат је успео да постави хипотезе истраживања и да их на ваљан начин провери и верификује, што је резултирало објављивањем научних радова кандидата у истакнутим међународним часописима. Кандидат је током израде докторске дисертације успешно имплементирао различита постојећа решења и нове системе, спровео неколико анкета, анализирао различиту телеметрију о коришћењу датих система, и изнео закључке зарад тестирања и верификовања полазних хипотеза. Треба поменути да се кандидат од почетка докторских студија бавио овом облашћу, остао истрајан и доследан у истраживању чији је коначан резултат докторска дисертација. На основу свега наведеног, можемо закључити да је кандидат Никола Пејић показао задовољавајући степен способности за самостални научни рад.

4. ОСТВАРЕНИ НАУЧНИ ДОПРИНОС

4.1. Приказ остварених научних доприноса

Главни доприноси дисертације су остварени у следећим правцима:

- Систематичан преглед литературе у области рецензије софтверских измена.
- Анализа трендова догађаја у системима за рецензију софтверских измена на примеру платформе *GitHub*.
- Пројектовање и евалуација система за аутоматско генерисање експертизе инжењера на основу описа њихових претходних измена.
- Пројектовање и евалуација система за подсећање рецензената на измене које захтевају њихову пажњу.
- Систематичан преглед литературе у области препоруке рецензената софтверских измена.
- Имплементација и евалуација метода за препоруку рецензената софтверских измена описаних у доступној литератури.
- Предлог нових техника за побољшање перформанси система за препоруку рецензената софтверског кода.
- Експериментално утврђивање обима очекиваних побољшања применом предложених техника над различитим репозиторијумима.
- Експериментално утврђивање обима очекиваних побољшања применом предложених техника над различитим методама за препоруку рецензената.

4.2. Критичка анализа резултата истраживања

Процес рецензије софтверског кода може имати доста малих и репетитивних корака који се могу или аутоматизовати или поједноставити, чиме би се позитивно утицало на искуство софтверских инжењера. Нарочито у случају великих репозиторијума са мноштвом инжењера (више стотина или хиљада), није могуће познавати све чланове тима и знати ко је експерт за одређену област и ко би био најбољи рецензент за дату измену. Са друге стране, велики репозиторијуми имају богату историју из које је могуће извући мноштво информација. Зато је пожељно искористити историју претходних рецензија и макар извући и презентовати корисне информације корисницима, а уколико је могуће и аутоматизовати одређене делове процеса рецензије.

Дисертација се састојала од неколико истраживања. Анализом утицаја *COVID-19* пандемије на платформу *GitHub* закључено је да је било одређених промена у трендовима догађаја (догађаји који се везују за комуникацију о коду су имали пораст, док су догађаји који се везују за активност заједнице отвореног кода и рад на документацији били у паду). На основу имплементације система за претрагу експерата и анализе телеметрије, закључено је

да одређени број инжењера може имати користи од представљеног система (они који су га активније користили), али да је то био мањи проценат укупног скупа корисника. На основу имплементације система за подсећање рецензента на њихове рецензије, и на основу резултата неколико анкета и прикупљене телеметрије о коришћењу система, закључено је да иако корисници нису претерано интераговали са подсетницима, ипак су сматрали да су они имали или позитиван (58% испитаника) или макар неутралан (42% испитаника) утицај на продуктивност инжењера, и велика већина је желела да настави да добија подсетнике (89% испитаника). Коначно, техника за побољшање перформанси система за препоруку рецензента показала се као релативно успешна – тестирањем над историјским подацима из индустрије и над постојећим системима за препоруку рецензента добијени су резултати по којима је у већини случајева могуће добити побољшање перформанси чак до 40% у погледу прецизности и опозива.

4.3. Верификација научних доприноса

Кандидат Никола Пејић је објавио следеће радове који су у непосредној вези са докторском дисертацијом:

Категорија M22:

1. N. Pejić, Z. Radivojević, M. Cvetanović, Helping Pull Request Reviewer Recommendation Systems to Focus, IEEE ACCESS, Vol. 11, pp. 71013 - 71025, Jul, 2023, <http://doi.org/10.1109/ACCESS.2023.3292056>, Impact Factor: 3.9 (M22)
2. N. Pejić, Z. Radivojević, M. Cvetanović, Analyzing the Impact of COVID-19 on GitHub Event Trends, Sustainability, Vol. 15, No. 19, pp. 1 - 16, Oct, 2023, <http://doi.org/10.3390/su151914622>, Impact Factor: 3.9 (M22)

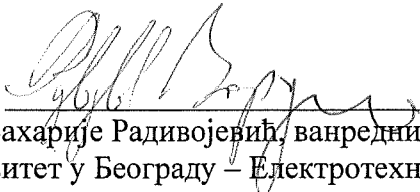
5. ЗАКЉУЧАК И ПРЕДЛОГ

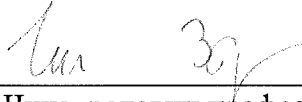
Дисертација истражује неколико приступа за генерисање информација из историје рецензија и њихово коришћење за различите видове аутоматизације. Прво је представљена анализа утицаја *COVID-19* пандемије на телеметрију сервиса *GitHub* из које је закључено да је већина догађаја наставила са претходним трендовима, али да су догађаји који се везују за активност заједнице отвореног кода и рада на документацији били у паду, док су догађаји који се везују за развој кода и комуникацију о коду имали пораст. Затим је описан систем за аутоматско генерисање експертисе инжењера на основу описа њихових претходних измена. Прототип система је имплементиран и учињен доступним групи од око 300 инжењера током периода од десет месеци, и користи од његовог рада су разматрани анализом телеметрије интеракције са системом, на основу чега је закључено да већина корисника није у значајној мери интераговала са системом, уз изузетак неколико појединаца који су га активно користили. Затим је представљен систем за подсећање инжењера на рецензије софтверског кода, на који је (након имплементације) било пријављено око 40 инжењера током периода од пет месеци. Утицај система анализиран је прегледом телеметрије и анкетањем корисника, и закључено је да је систем имао већински позитиван утицај на продуктивност инжењера. Након тога су описани неки од постојећих приступа за препоруку рецензента софтверског кода, и представљена је техника за побољшање перформанси постојећих система. Утицај технике експериментално је утврђен на основу историјских података неколико репозиторијума из индустрије, и закључено је да већина приступа може имати користи од примене предложене технике.


На основу свега наведеног, Комисија констатује да је Кандидат Никола Пејић испунио све формалне и суштинске услове предвиђене Законом о високом образовању, Статутом и Правилником о докторским студијама Универзитета у Београду и Електротехничког факултета Универзитета у Београду. Комисија има задовољство да предложи Наставном-научном већу Електротехничког факултета Универзитета у Београду да се докторска дисертација под насловом „Побољшање система за рецензију софтверских измена на основу информација генерисаних из историје рецензија” кандидата Николе Пејића прихвати, изложи на јавни увид и упути на коначно усвајање Већу научних области техничких наука Универзитета у Београду.

У Београду, 30.06.2025.

ЧЛАНОВИ КОМИСИЈЕ


Др Захарије Радивојевић, ванредни професор
Универзитет у Београду – Електротехнички факултет


др Зоран Чича, редовни професор
Универзитет у Београду – Електротехнички факултет


др Синиша Влајић, редовни професор
Универзитет у Београду – Факултет организационих наука