

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

I ПОДАЦИ О КОМИСИЈИ		
1. Датум и орган који је именовано комисију: Декан Факултета техничких наука у Новом Саду на основу предлога матичне катедре и одлуке Наставно-научног већа Факултета техничких наука; Решење број 012-72 / 07-2014 од 27.05.2021.		
2. Састав комисије у складу са <i>Правилима докторских студија Универзитета у Новом Саду</i> :		
1. Ристић Александар презиме и име	Редовни професор звање	Геоинформатика, 05.11.2019. ужа научна област и датум избора
Факултет техничких наука, Нови Сад установа у којој је запослен-а		председник функција у комисији
2. Стојановић Драган презиме и име	Редовни професор звање	Електротехника и рачунарство, 20.04.2015. ужа научна област и датум избора
Електронски факултет, Ниш установа у којој је запослен-а		члан функција у комисији
3. Јовановић Душан презиме и име	Ванредни професор звање	Геоинформатика, 11.03.2021. ужа научна област и датум избора
Факултет техничких наука, Нови Сад установа у којој је запослен-а		члан функција у комисији
4. Драган Дину презиме и име	Ванредни професор звање	Примењене рачунарске науке и информатика, 01.11.2020. ужа научна област и датум избора
Факултет техничких наука, Нови Сад установа у којој је запослен-а		члан функција у комисији
5. Говедарица Миро презиме и име	Редовни професор звање	Геоинформатика, 26.04.2012. ужа научна област и датум избора
Факултет техничких наука, Нови Сад установа у којој је запослен-а		ментор функција у комисији
II ПОДАЦИ О КАНДИДАТУ		
1. Име, име једног родитеља, презиме: Владимир Жарко Пајић		
2. Датум рођења, општина, држава: 30.11.1979, Зрењанин, Србија		
3. Назив факултета, назив претходно завршеног нивоа студија и стечени стручни/академски назив: Факултет техничких наука, Нови Сад, магистарске студије, магистар техничких наука		
4. Година уписа на докторске студије и назив студијског програма докторских студија: 2014. Рачунарство и аутоматика		

<b>III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b>
<b>Модел управљања великим серијама геопросторних података</b>
<b>ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b>
Навести кратак садржај са знаком броја страница, поглавља, слика, схема, графика и сл.
<p>Докторска дисертација кандидата Владимира Пајића под насловом “Модел управљања великим серијама геопросторних података“ садржи укупно 7 поглавља, и то:</p> <ol style="list-style-type: none"> <li>1. Уводна разматрања и хипотезе</li> <li>2. Стање у области</li> <li>3. Просторно индексирање и криве које попуњавају простор</li> <li>4. Модел система за управљање великим серијама геопросторних података типа облак тачака</li> <li>5. Имплементациони модел и верификација</li> <li>6. Дискусија</li> <li>7. Закључак</li> </ol> <p>Испред основног текста дисертације дата је насловна страна и пратећи уводни материјал који садржи: обавезну кључну документацијску информацију, резиме рада на енглеском и српском језику, садржај рада, листу слика, листу табела. Дисертација је написана на 100 страница чији се текст позива на 77 референци и садржи 9 табела и 43 слике. На крају основног текста је дат списак коришћене литературе. Дисертација садржи 1 прилог.</p>
<b>ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b>
<p>Поглавље 1 - <b>Уводна разматрања и хипотезе</b> описује проблем и представља предмет истраживања - управљање великим серијама геопросторних података. Геопросторни тип података, облак тачака, је изабран као репрезентативан због специфичних карактеристика и комплексности управљања. На основу предмета истраживања је постављена основна хипотеза да употреба Big Data техника омогућује управљање геопросторним подацима у условима (велика количина или брзина пристизања) који превазилазе могућности традиционалних техника, заснованих на фајл систему или релационим базама података. Поред основне постављене су и три посебне хипотезе које се односе на очување просторне блискости у дистрибуираном окружењу, унапређење перформанси применом паралелне обраде података и потврђивање скалабилности модела. Као основни циљ истраживања је постављено дефинисање архитектуре модела за управљање великим серијама геопросторних података типа облак тачака која укључује све релевантне аспекте, као што су структурирање и складиштење података, индексирање и дистрибуција, геопросторне анализе и друго.</p> <p><i>Комисија закључује да су предмет, план истраживања, план реализације, заједно са постављеним циљевима дисертације, представљеним истраживачким хипотезама и истраживачким питањима прецизно и адекватно дефинисани.</i></p> <p>Поглавље 2 - <b>Стање у области</b> у уводном делу даје дефиницију појма Big Data и приказује основне концепте на којима се Big Data базира, попут функционалног програмирања, дистрибуираног фајл система, Map-Reduce програмског модела и NoSQL база података. Након тога су детаљно приказане платформе на којима се заснива модел за управљање великим серијама геопросторних података, Apache Spark и Apache HBase. Затим је дата детаљна анализа досадашњих истраживања, као и постојећих решења из области управљања растерским и векторским подацима применом Big Data техника. Што се тиче облака тачака прво су представљена решења за управљање заснована на фајл систему и релационим базама података. Након тога су детаљно представљена и анализирана досадашња истраживања из области управљања облаком тачака у Big Data парадигми. На крају поглавља кандидат сумира концепте и резултате истраживања и изводи релевантне закључке, пре свега да су истраживања која се тичу теме дисертације још увек релативно малобројна и фокусирана на уске области примене, као и да постоји потреба за истраживањем које обухвата дефинисање свеобухватног модела управљања.</p> <p><i>Комисија закључује да је преглед стања у области постављен јасно и концизно у складу са темом дисертације.</i></p> <p>Поглавље 3 – <b>Просторно индексирање и криве које попуњавају простор</b> даље разрађује традиционалне методе за управљање облаком тачака као геопросторним типом података, са освртом</p>

на њихове предности и мане. Затим се детаљно бави кривама за попуњавање простора (енглески space-filling curves) које су одабране као метода за индексирање због њихове особине пресликавања вишедимензионалних података на једнодимензионалне уз очување просторне блискости. Представљене су две најчешће коришћене криве за попуњавање простора Хилбертова и Z-крива. Због специфичности кривих за попуњавање простора приказана је техника просторног филтрирања која се базира на Z-кривој. Такође су приказани начини за унапређење перформанси таквих упита коришћењем вишеструких кривих.

*Комисија сматра да су технике за индексирање геопросторних података типа облака тачака применом кривих за попуњавање простора представљене на јасан и адекватан начин, уз адекватне илустрације и референце. Садржај овог поглавља неопходан је за свеобухватно разумевање модела за управљање великим серијама геопросторних података типа облак тачака.*

Поглавље 4 – **Модел система за управљање великим серијама геопросторних података типа облак тачака** детаљно описује свеобухватну архитектуру модела која садржи све релевантне елементе: структурирање и складиштење података, геопросторне анализе, визуализацију геопросторних података, управљање временским серијама, дистрибуцију података и контролу квалитета. Свака од наведених области је детаљно анализирана и објашњена. Затим је дефинисан концептуани модел података у смислу логичког модела података заснованог на Z-кривој за попуњавање простора и Мортон коду и физичког модела записа у оквиру дистрибуираног складишта. Типови података су дефинисани применом Spark кориснички дефинисаних типова, а операција над подацима применом SQL упита и кориснички дефинисаних функција. Дефинисане су и имплементирани две врсте операција над типом облака тачака у дистрибуираном окружењу, просторно филтрирање и проналажење најближих суседа. Такође је дефинисана и имплементирана метода за побољшање перформанси упита применом вишеструких Z-кривих.

*Комисија закључује да је модел за управљање великим серијама геопросторних података типа облак тачака представљен адекватно, јасно и детаљно узимајући у обзир све релевантне аспекте. Дефиниција модела података и операција применом концепата описаних у претходим поглављима је описана на систематичан начин те је кандидат је на одговарајући начин представио основне елементе истраживања.*

Поглавље 5 – **Имплементациони модел и верификација** садржи приказ имплементације модела описаног у поглављу 4 применом програмског језика Scala и Spark DataFrame програмског интерфејса. Елементи имплементационог модела су описани применом UML дијаграма. Верификација је извршена применом три скупа података различите величине. Вршена су два типа експеримената, у првом је поређен модел дефинисан у оквиру дисертације са PostgreSQL релационом базом података, а у другом је верификована скалабилност модела повећањем броја рачунара у кластеру. Утврђено је да модел пружа бољу перформансу него релациона база. Такође, са повећањем броја тачака модел скалира приближно линеарно за разлику од релационе базе. Поред тога модел показује повећање перформанси додавањем рачунара у кластер.

*Комисија закључује да је верификација адекватно дефинисана и систематично спроведена и да су постављене хипотезе у потпуности верификоване.*

Поглавље 6 – **Дискусија** поред анализе резултата верификације, даје објашњење за одређене изборе начињене приликом дефинисања модела, као и анализу њихових предности и недостатака. Посебан осврт је дат на анализу области примене резултата изведеног истраживања који су верификовани кроз истраживања која се на њега ослањају.

*Комисија закључује да је дискусија спроведена детаљно и да је обухватила све релевантне елементе.*

Поглавље 7 – **Закључак** сумира резултате истраживања и наводи даље правце истраживања.

*Комисија закључује да закључна разматрања потврђују адекватност и значајност модела за управљање великим серијама геопросторних података, и дају детаљну анализу будућих праваца истраживања.*

Литература - садржи 77 прегледно систематизованих библиографских навода.

*Комисија сматра да је литература адекватно одабрана, правилно коришћена и да одговара теми дисертације.*

*Сходно претходно наведеном, Комисија позитивно оцењује све делове докторске дисертације кандидата Владимира Пајића.*

#### **VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ:**

Таксативно навести називе радова, где и када су објављени. Прво навести најмање један рад објављен или прихваћен за објављивање у складу са *Правилима докторских студија Универзитета у Новом Саду* који је повезан са садржајем докторске дисертације. У случају радова прихваћених за објављивање, таксативно навести називе радова, где и када ће бити објављени и приложити потврду уредника часописа о томе.

- I.1. **V. Pajić**, M. Govedarica, and M. Amovic, Model of pointcloud data management system in big data paradigm, ISPRS International Journal of Geo-Information, vol. 7, p. 265, 2018
- I.2. Jovanović, D.; Milovanov, S.; Ruskovski, I.; Govedarica, M.; Sladić, D.; Radulović, A.; **Pajić, V.**, (2020): Building Virtual 3D City Model for Smart Cities Applications: A Case Study on Campus Area of the University of Novi Sad. ISPRS Int. J. Geo-Inf., 9, p. 476
- I.3. Gordana Jakovljevic, Miro Govedarica, Flor Alvarez-Taboada, and **Vladimir Pajić**. Accuracy assessment of deeplearning based classification of lidar and uav points clouds for dtm creation and flood risk mapping. Geosciences, 9(7),2019. ISSN 2076-3263.
- I.4. M. Amovic, **V. Pajić**, M. Govedarica, S. Vasiljevic, Spatio – temporal types of data in Big data paradigm, IFKAD 2016, 11th International Forum on Knowledge Asset Dynamics, Dresden, Germany, Volume: XI
- I.5. **V. Pajić**, M. Govedarica, Z. Galic, I. Alargic, OBRADA OBLAKA TAČAKA NA APACHE SPARK PLATFORMI POINT CLOUD PROCESSING ON APACHE SPARK, YU INFO 2015 ZBORNIK RADOVA, Beograd, 2015. ISBN: 978-86-85525-15-5

#### **VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА:**

У оквиру дисертације је дефинисан модел за управљање великим серијама геопросторних података, са фокусом на облак тачака. Пре дефинисања самог модела представљени су концепти на којима се заснива рад са великим серијама и количинама геопросторних података. Основу представља Map-Reduce програмски модел, дистрибуирани фајл системи, и Big Data складишта података.

На основу прегледа стања у области утврђени су елементи који су заједнички за сва истраживања. Пре свега, већина решења за индексирање геопросторних и просторно-временских података заснивају на кривама за попуњавање простора. То је само по себи оправдано јер представља најпогоднији метод индексирања у дистрибуираном окружењу. Најчешће се користе Z-крива или Хилбертова крива. Такође, постојећа решења заснована на Big Data парадигми за управљање растерским и векторским подацима су развијена у великој мери и омогућују примену у оквиру продукционих апликација. С друге стране, истраживања о управљању облаком тачака у Big Data парадигми су још увек релативно малобројна, разуђена и покривају одређене уске области примене.

На основу наведеног, у оквиру модела за управљање великим серијама геопросторних података је фокус постављен на управљање облаком тачака због његове комплексности и још увек релативно мање покривености истраживањима. Модел је заснован на новијим Big Data концептима и платформама попут Sparkа те стога превазилази ограничења са којима се суочавају традиционалне методе управљања. Даље, његова предност је то што је имплементиран применом SparkSQL модула и DataFrame API-ја што доноси омогућује бољу оптимизација упита због познавања структуре података, као и коришћење језика домена за опис података и операција.

Основни проблем који морају да реше Big Data платформе за рад са геопросторним или просторно-временским подацима је индексирање у оквиру дистрибуираног система где није могуће једноставно представити вишедимензионалне структуре. Због тога врши редукција димензионалности на једну димензију и делови низова података дистрибуирају кроз систем. Основни начин за свођење вишедимензионалних података на једнодимензионалне су криве за

попуњавање простора, јер омогућују задржавање просторне блискости података након линеаризације.

Као метод за индексирање облака тачака је изабрана Z-крива, због тога што задржава солидну просторну блискост, омогућује једноставно мапирање на октално стабло, и једноставну имплементацију просторних упита. Облак тачака индексиран Z-кривом се дистрибуира у оквиру кластера ради паралелне обраде. Као платформа за обраду је изабран Spark јер пружа одличну перформансу и скалабилност.

Дефинисане су две веома често коришћене операције над облаком тачака, просторно филтрирање и претрага најближих суседа. Такође, су објашњене могућности за унапређење тачности и ефикасности обраде. Коначно, дефинисани модел се може посматрати као први корак ка свеобухватној дистрибуираној бази података облака тачака.

У поређењу са традиционалним начинима управљања облаком тачака, заснованим на фајл систему и једној процесорској јединици, дистрибуирани приступ заснован на Spark-у, је показао бољу скалабилност, паерформансу, робустност и отпорност на грешке.

#### **VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА:**

Експлицитно навести позитивну или негативну оцену начина приказа и тумачења резултата истраживања.

Прегледом докторске дисертације, Комисија закључује да су резултати спроведеног истраживања приказани и тумачени на јасан, систематичан и научно коректан начин. Такође, приказани резултати и закључци су у складу са дефинисаним циљевима, задацима и постављеним полазним хипотезама и истраживачким питањима. Сви закључци донесени на основу добијених резултата су адекватно аргументовани и потврђују дефинисане полазне хипотезе.

Текст дисертације проверен је у софтверу за детекцију плагијаризма „iThenticate“. На основу резултата провере, Комисија је донела закључак да је докторска дисертација оригинално ауторско дело кандидата Владимира Пајића.

У складу са наведеним, Комисија ПОЗИТИВНО оцењује начин на који су резултати истраживања приказани и тумачени.

<p><b>IX КОНАЧНА ОЦЕНА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b></p> <p>Експлицитно навести да ли дисертација јесте или није написана у складу са наведеним образложењем, као и да ли она садржи или не садржи све битне елементе. Дати јасне, прецизне и концизне одговоре на 3. и 4. питање:</p>
<p>1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме?  <b>Да, дисертација је написана у складу са образложењем наведеним у пријави теме.</b></p>
<p>2. Да ли дисертација садржи све битне елементе?  <b>Да, дисертација садржи све битне елементе који се захтевају по Статуту Факултета техничких наука и Универзитета у Новом Саду, као и Закона о високом образовању.</b></p>
<p>3. По чему је дисертација оригиналан допринос науци?</p> <p>Дисертација представља свеобухватан модел за управљање великим серијама геопросторних података и дефинише његову архитектуру применом Big Data концепата и техника. На основу анализе стања у области је утврђено да не постоји истраживање које се бави геопросторним подацима, и посебно облаком тачака, у Big Data парадигми, узимајући у обзир све аспекте коришћења. Такође, архитектура модела приказаног у дисертацији се базира на платформама које обезбеђују максималне перформансе применом концепата унутар меморијске обраде. Следећи значајан допринос је дефинисање типова података и операција применом језика домена, које омогућује коришћење дефинисаног модела без познавања његових интерних карактеристика.</p>
<p>4. Који су недостаци дисертације и какав је њихов утицај на резултат истраживања?  <b>У дисертацији нису уочени недостаци који би утицали на резултате истраживања</b></p>
<p><b>X ПРЕДЛОГ:</b></p> <p>На основу наведеног, комисија предлаже:</p>
<p>а) да се докторска дисертација прихвати, а кандидату одобри одбрана;  б) да се докторска дисертација врати кандидату на дораду (да се допуни односно измени);  в) да се докторска дисертација одбије.</p>

Место и датум: Нови Сад, 07.06.2021.

1. Александар Ристић, редовни професор  
\_\_\_\_\_, председник
2. Драган Стојановић, редовни професор  
\_\_\_\_\_, члан
3. Душан Јовановић, ванредни професор  
\_\_\_\_\_, члан
4. Дину Драган, ванредни професор  
\_\_\_\_\_, члан
5. Миро Говедарица, редовни професор  
\_\_\_\_\_, ментор

**НАПОМЕНА:** Члан комисије који не жели да потпише извештај јер се не слаже са мишљењем већине чланова комисије, дужан је да унесе у извештај образложење односно разлоге због којих не жели да потпише извештај и да исти потпише.