

РЕФЕРАТ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Студијски програм: Интелигентни системи

Комисија у саставу:

Проф. др Владан Девеџић, Факултет организационих наука Универзитета у Београду;

Проф. др Ненад Митић, редовни професор у пензији, Математички факултет
Универзитета у Београду;

Др Александра Марковић, виши научни сарадник, Институт за српски језик САНУ,
именована Одлуком Већа за студије при Универзитету у Београду број
06-4574/IX/563/3-25 од 21. септембра 2026. године, доставља реферат о оцени докторске
дисертације

„Машинско учење и велики језички модели у развоју семантичких мрежа и
њиховој примени на аутоматско разумевање текста“

кандидата

Саше Петалинкара.

1. УВОД

1.1. Хронологија одобравања и израде дисертације

Саша Петалинкар уписао је докторске академске студије на студијском програму
Интелигентни системи при Универзитету у Београду школске 2018/2019. године.

Веће за студије при Универзитету у Београду је на седници одржаној 17. фебруара 2025.
године, Одлуком број 06-3901/IV4.1/2-25, образовало Комисију за оцену научне
заснованости теме докторске дисертације „Машинско учење и велики језички модели у
развоју семантичких мрежа и њиховој примени на аутоматско разумевање текста“ и
испуњености услова кандидата. Комисију су чинили проф. др Владан Девеџић са
Факултета организационих наука, доц. др Милош Утвић са Филолошког факултета и др
Михаило Шкорић, научни сарадник Рударско-геолошког факултета Универзитета у
Београду.

Комисија је сачинила извештај од 1. децембра 2025. године, у којем је предложила
прихватање теме и одобравање израде дисертације. За менторе су предложене др Ранка
Станковић, редовни професор Рударско-геолошког факултета, и др Јелена Граовац,
ванредни професор Математичког факултета Универзитета у Београду. Према
записнику са седнице Већа за студије при Универзитету одржане 22. децембра 2025.
године, извештај Комисије о научној заснованости теме једногласно је усвојен.

Молба за именовање Комисије за преглед и оцену докторске дисертације, са потписима
кандидата и ментора, примљена је на Универзитету у Београду 24. августа 2026. године,
што је потврђено пријемним печатом на објављеном документу. У молби су за чланове

Комисије предложени др Владан Девеџић, академик и редовни професор Факултета организационих наука, др Ненад Митић, редовни професор Математичког факултета у пензији, и др Александра Марковић, виши научни сарадник Института за српски језик САНУ.

Веће за студије при Универзитету у Београду је 21. септембра 2026. године, Одлуком број 06-4574/LX/563/3-25, именovalo Комисију за преглед и оцену у саставу: проф. др Владан Девеџић, проф. др Ненад Митић и др Александра Марковић, виши научни сарадник. Одлука предвиђа достављање реферата у року од 30 дана од именовања.

1.2. Научна област

Према подацима у дисертацији, истраживање припада области интелигентних система, са ужим усмерењем на обраду природног језика и рачунарску лингвистику. Рад повезује методе машинског учења и примену великих језичких модела са развојем, провером и применом лексичко-семантичких ресурса за српски језик.

1.3. Биографски подаци о кандидату

Сапа Петалинкар рођен је 24. децембра 1979. године у Бору, где је завршио основну и средњу школу. Математички факултет Универзитета у Београду уписао је школске 1998/1999. године. На смеру Рачунарство и информатика дипломирао је 25. јуна 2009. године, са просечном оценом 8,14, и стекао стручни назив дипломирани математичар.

Мастер академске студије на студијском програму Математика уписао је школске 2010/2011. године на истом факултету. Студије је завршио 24. јануара 2012. године одбраном мастер рада „Обрада елемената логичког изгледа текста под системом UNITEX“, са оценом 10 на мастер раду и остварених 60 ЕСПБ бодова, чиме је стекао академско звање мастер математичар.

Докторске академске студије на студијском програму Интелигентни системи при Универзитету у Београду уписао је школске 2018/2019. године. Према извештају о оцени научне заснованости теме његове докторске дисертације, на положеним испитима и Завршном раду 1 остварио је 120 ЕСПБ бодова, са просечном оценом 9,70.

Професионално искуство стицао је у Медијском архиву Ебарт, где је од 2012. године пружао подршку системима за архивирање и анализу медијских објава. Од 1. марта 2026. године запослен је на Рударско-геолошком факултету Универзитета у Београду на пословима пројектанта информатичке инфраструктуре. Одлука о његовом избору у звање истраживача-сарадника донета је 25. јуна 2026. године. Члан је Друштва за језичке ресурсе и технологије ЈеРТех, у оквиру којег учествује у развоју ресурса и алата за обраду српског језика. Његова истраживачка интересовања обухватају обраду природног језика, семантичке мреже, анализу сентимента и примену машинског учења и великих језичких модела у обради српског текста.

1.4. Библиографија кандидата

У наставку је наведено 16 публикација чији је кандидат аутор или коаутор: четири рада у научним часописима, једно поглавље у монографији, шест радова објављених у зборницима и пет саопштења објављених у изводу. Радови у часописима под редним бројем 1 и 3 су директно везани за резултате докторске дисертације.

Радови у научним часописима

1. **Saša Petalinkar**; Ranka Stanković; Milica Ikonić Nešić; Cvetana Krstev. 2026. „A semi-automated LLM-based framework for word sense disambiguation in Serbian”. *Intelligent Data Analysis*. DOI: <https://doi.org/10.1177/1088467X261469292>. Извор: <https://journals.sagepub.com/doi/10.1177/1088467X261469292>. (M23)
2. Milica Ikonić Nešić; **Saša Petalinkar**; Olivera Kitanović; Ranka Stanković; Miloš Utvić. 2026. „CNN-based named entity linking: Serbian use case”. *Poznan Studies in Contemporary Linguistics*, 62(2), 249–280. DOI: <https://doi.org/10.1515/psicl-2025-0064>. Извор: <https://api.crossref.org/works/10.1515%2Fpsicl-2025-0064>. (M22)
3. **Saša Petalinkar**; Ranka M. Stanković; Milica Ikonić Nešić. 2025. „Comparative analysis of methods for creating a sentiment lexicon of the Serbian WordNet”. *The Electronic Library*, 43(4), 547–577. DOI: <https://doi.org/10.1108/EL-08-2024-0253>. Извор: <https://doi.org/10.1108/EL-08-2024-0253>. (M22)
4. **Saša Petalinkar**. 2025 (година свеске: 2024). „Creating a Synthetic Evaluation Dataset for Serbian SentiWordNet Using Large Language Models”. *Infotheca — Journal for Digital Humanities*, 24(1), 53–70. DOI: <https://doi.org/10.18485/infotheca.2024.24.1.3>. Извор: https://infoteka.bg.ac.rs/ojs/index.php/infoteka/article/view/2024.24.1.3_en. (M52)

Поглавље у монографији

5. Mihailo Škorić; **Saša Petalinkar**. 2025. „Quality Textual Corpora and New South Slavic Language Models”. *South Slavic Languages in the Digital Environment JuDig*, 337–348. DOI: <https://doi.org/10.18485/judig.2025.1.ch19>. Извор: <https://dr.rgf.bg.ac.rs/s/repo/item/10727>. (M14)

Радови објављени у зборницима

6. Ranka Stanković; Cvetana Krstev; **Saša Petalinkar**; Milica Ikonić Nešić; Jelena Bogdanović; Milica Dinić Marinković; Aleksandra Marković; Marina Bagi; Marijana Đukić. 2026. „SrELEXIS-WSD: Hybrid Semi-Automated WSD for Serbian with Large Language Models—Results and Challenges”. *Proceedings of the UniDive 2026 Workshop*, 30–33. Извор: <https://dr.rgf.bg.ac.rs/s/repo/item/11056>. (M33)
7. Ranka Stanković; **Saša Petalinkar**; Milica Ikonić Nešić. 2026. „From Names to Places: Serbian Entity Linking Through the Lens of GIS”. *Proceedings of the 1st ReLDI Conference on Language Science and Technology*, 25–26. септембра 2025. године, стр. 68–71. DOI целог зборника: <https://doi.org/10.5281/zenodo.18488312>. Извор: <https://dr.rgf.bg.ac.rs/s/repo/item/10885>. (M33)
8. Milica Ikonić Nešić; **Saša Petalinkar**; Ranka Stanković; Ruslan Mitkov. 2025. „From Zero to Hero: Building Serbian NER from Rules to LLMs”. *Proceedings of the First Workshop on Comparative Performance Evaluation: From Rules to Language Models*, 87–96. DOI: <https://doi.org/10.26615/978-954-452-102-8-010>. Извор: <https://aclanthology.org/2025.r2lm-1.10/>. (M33)
9. Milica Ikonić Nešić; **Saša Petalinkar**; Mihailo Škorić; Ranka Stanković; Biljana Rujević. 2024. „Advancing Sentiment Analysis in Serbian Literature: A Zero and Few-Shot

Learning Approach Using the Mistral Model". *Proceedings of the Sixth International Conference on Computational Linguistics in Bulgaria (CLIB 2024)*, 58–70. DOI: <https://doi.org/10.47810/CLIB.24.05>. Извор: <https://aclanthology.org/2024.clib-1.5/>. (M33)

10. Milica Ikonić Nešić; **Saša Petalinkar**; Mihailo Škorić; Ranka Stanković. 2024. „BERT Downstream Task Analysis: Named Entity Recognition in Serbian". *Disruptive Information Technologies for a Smart Society. Lecture Notes in Networks and Systems*, vol. 860, 333–347. DOI: https://doi.org/10.1007/978-3-031-71419-1_29. Извор: https://doi.org/10.1007/978-3-031-71419-1_29. (M33)
11. Milica Ikonić Nešić; **Saša Petalinkar**; Ranka Stanković; Miloš Utvić; Olivera Kitanović. 2024. „SrpCNNeL: Serbian Model for Named Entity Linking". *Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS). Annals of Computer Science and Information Systems*, vol. 39, 465–473. DOI: <https://doi.org/10.15439/2024F8827>. Извор: <https://dr.rgf.bg.ac.rs/s/repo/item/9165>. (M33)

Саопштења објављена у изводу

12. Milica Ikonić Nešić; **Saša Petalinkar**; Ranka Stanković. 2025. „Towards Linguistic Completeness in Knowledge Graphs: Generating Serbian Inflections with Language Models". *Artificial Intelligence Conference Book of Abstracts*, 129–130. Извор: <https://dr.rgf.bg.ac.rs/s/repo/item/10709>. (M34)
13. **Saša Petalinkar**; Milica Ikonić Nešić. 2024. „Automating Synset Example Creation: A Case Study with Serbian Wordnet and ChatGPT". *South Slavic Languages in the Digital Environment JuDig: Book of Abstracts*, 56. Извор: <https://dr.rgf.bg.ac.rs/s/repo/item/9151>.
14. Mihailo Škorić; **Saša Petalinkar**. 2024. „New XLM-R-based language models for Serbian and Serbo-Croatian". *Artificial Intelligence Conference: Book of Abstracts*, 59–60. Извор: https://www.mi.sanu.ac.rs/~ai_conf/previous_editions/2024/AI_Conference_Book_of_Abstracts.pdf#page=63. (M34)
15. **Saša Petalinkar**; Milica Ikonić Nešić; Ranka Stanković; Mihailo Škorić. 2024. „Comparison of Zero and Few-Shot Learning Approach Using the LLMs for Sentiment Analysis in Serbian Literature". *Artificial Intelligence Conference: Book of Abstracts*, 55–56. Извор: https://www.mi.sanu.ac.rs/~ai_conf/previous_editions/2024/AI_Conference_Book_of_Abstracts.pdf#page=59. (M34)
16. Milica Ikonić Nešić; **Saša Petalinkar**; Ranka Stanković. 2024. „Development and Evaluation of Named Entity Linking Models for Serbian Language with Wikidata Integration". *Artificial Intelligence Conference: Book of Abstracts*, 47–48. Извор: https://www.mi.sanu.ac.rs/~ai_conf/previous_editions/2024/AI_Conference_Book_of_Abstracts.pdf#page=51. (M34)

2. ОПИС ДИСЕРТАЦИЈЕ

2.1. Обим, структура и основни садржај

Дисертација је написана на српском језику, претежно ћириличким писмом, уз наслов, резиме и кључне речи на енглеском језику. Достављена верзија обухвата 189 страна у PDF формату: осам ненумерисаних уводних страна и 181 нумерисану страну. Главни текст заузима странице 1–153 и организован је у седам поглавља. Следе литература на страницама 154–165 и три прилога – термилошки глосар, индекс табела и индекс слика – на страницама 166–181. Рад садржи 42 табеле, 25 слика и 139 библиографских јединица.

2.2. Преглед поглавља

Прво поглавље, „Увод“ (стр. 1–4), одређује предмет и циљеве истраживања, описује значај Српског ворднета за обраду природног језика и формулише општу и помоћне хипотезе. У њему су најављени главни доприноси и објашњена организација дисертације.

Друго поглавље, „Преглед досадашњих истраживања“ (стр. 5–33), представља лексичко-семантичке ресурсе, вишезначност, ворднете и SentiWordNet, као и приступе засноване на машинском учењу и великим језичким моделима који су релевантни за истраживање. Посебно су размотрени обликовање упита, генерисање података, дестилација, разрешење вишезначности речи и поравнање ресурса.

Треће поглавље, „Методологија“ (стр. 34–59), описује поступке за одређивање сентимента синсета, формирање синтетичког евалуационог скупа, разрешење вишезначности речи применом генеративног модела и пренос знања на мање моделе. Приказани су и поступци превођења синсета, поравнања са описним речником, екстерне евалуације сентимента и циљане провере у машинском превођењу.

Четврто поглавље, „Коришћени скупови података и велики језички модели“ (стр. 60–69), представља Принстонски и Српски ворднет, ELEXIS-WSD и његов српски део, Речник савременог српског језика, Senti-pol-sr, коришћене моделе и скуп новинских реченица за евалуацију сентимента.

Пето поглавље, „Резултати“ (стр. 70–126), приказује развијени програмски омотач за ворднет, резултате евалуације лексикона сентимента, изградњу и ревизију инвентара значења и квантитативну и квалитативну проверу разрешења вишезначности речи. Обухваћени су дестилација, формирање српских синсета, поравнање са речником и две провере употребе ресурса у даљим задацима.

Шесто поглавље, „Дискусија“ (стр. 127–147), тумачи резултате по истраживачким целинама, разматра покривеност ресурса и разлике у грануларности значења и одређује домет аутоматизације. Посебно се разликују налази о сентименту од ограничених исхода циљане провере машинског превођења.

Седмо поглавље, „Закључак“ (стр. 148–153), повезује налазе са хипотезама и издваја доприносе, ограничења и правце даљег рада. Након њега налазе се литература (стр. 154–165), термилошки глосар (стр. 166–178), индекс табела (стр. 179–180) и индекс слика (стр. 181).

2.3. Предмет, циљеви и полазне хипотезе

Предмет истраживања је примена машинског учења и великих језичких модела у развоју лексичко-семантичких ресурса за српски језик и њиховој употреби у аутоматском разумевању текста. Основни циљ је повезивање поступака за обогаћивање Српског ворднета, проверу предложених допуна и примену ресурса на корпусном материјалу, уз очување стручне лексикографске контроле.

Посебни циљеви односе се на процену сентимента синсета, изградњу и ревизију инвентара значења, додељивање одговарајућих значења лексичким јединицама у тексту, испитивање преноса знања на мање моделе, формирање кандидата за српске синсете и поравнање значења из описног речника са Српским ворднетом. Употребљивост ресурса испитана је у анализи сентимента и циљаној поставци машинског превођења.

У достављеној дисертацији постављена је општа хипотеза *H0* о могућности унапређења развоја ресурса и аутоматског обележавања текста. Она је разрађена кроз *H1* о септменту, *H2* о репрезентацији и поравнању значења, *H3* о контролисаној генеративној допуни и *H4* о употребљивости у одабраним задацима. Хипотеза *H4* разрађена је кроз *H4a*, која се односи на анализу сентимента, и *H4b*, која се односи на циљану проверу машинског превођења.

3. ОСТВАРЕНИ НАУЧНИ ДОПРИНОС

3.1. Приказ научних доприноса

Научни доприноси дисертације приказани су у следећим повезаним целинама:

1. Развој и евалуација слоја сентимента Српског ворднета. Развијени су лексикони и модели који процењују сентимент на основу глоса синсета, уз поступак формирања и ручне провере синтетичког евалуационог скупа. Резултати су изложени у §5.2, а спољна провера на новинским реченицама у §5.11. Senti-pol-sr и полазно пресликавање S0 коришћени су као постојеће основе за поређење и не представљају нове ресурсе настале у дисертацији.
2. Поступак разрешења вишезначности речи и ресурси за његову проверу. Поступак повезује појаве лексичких јединица у корпусу са предложеним значењима и обухвата ревизију инвентара WSD-sr-sens-inv-v1/v2. Припремљен је ручно проверени ELEXIS-WSD-sr-gold, са одвојеним развојно-аналитичким делом од 120 реченица и тест скупом од 300 реченица. На једночланим јединицама издвојеног тест скупа GPT-4.1 у другој итерацији постиже тачност 92,4%, у односу на 79,2% у првој итерацији (табела 5-13, стр. 96–97). Ове вредности се односе на наведени подскуп и поставку, а не на све типове јединица или све корпусе.
3. Испитивање преноса знања посредством сребрних података. У §5.8 поређени су полазни и дестиловани модели на скупу од 1.921 јединице, при чему су у укупан број јединица за израчунавање тачности укључене и непокривене јединице. Тачност модела simple расте са 54,4% на 68,5%, а multilingual-e5-large са 59,2% на 72,6%; код модела Tesla опада са 53,4% на 51,3% (табела 5-16, стр. 104). Резултати показују да поступак може побољшати тачност појединих модела, али да побољшање није остварено код свих испитаних модела.

4. **Поступци формирања кандидата за српске синсете.** Упоређена су три радна тока на 578 полазних синсета, са укупно 1.734 генерисана записа. Ручно је проверено 90 записа за 30 синсета (§5.9, стр. 108–112). Концептуално усмерен поступак постиже највишу просечну оцену глоса у провереном узорку. Међутим, ограничен обим провере коју је спровео један анотатор не омогућава општи закључак о предности тог поступка нити оправдава непосредан унос свих кандидата у ресурс.
5. **Поравнање описног речника са Српским ворднетом.** Развијен је поступак предлагања и класификације односа између речничких значења и синсета. Обрада обухвата 3.414 кандидатских парова; ручно проверени скуп садржи 83 пара, од којих су 42 оцењена као тачна, 20 као блиска и 21 као нетачан (§5.10, стр. 113–119). Поступак припрема материјал за лексикографску редакцију, при чему модел за сваки пар значења који је кандидат за евалуацију предлаже: да ли се значења подударају, да ли је значење у речнику шире или уже од значења синсета, да ли су значења само семантички повезана или веза није утврђена.

Ове целине обухватају седам доприноса побројаних у §7.1 дисертације: поступак за разрешење вишезначности речи и златни скуп приказани су заједно, док су скупови кандидата повезани са поступцима који их производе. Овим груписањем обједињени су међусобно повезани резултати.

3.2. Верификација доприноса кроз публикације

Рад „Comparative analysis of methods for creating a sentiment lexicon of the Serbian WordNet“ тематски је непосредно повезан са развојем лексикона сентимента и њиховом евалуацијом (§3.1 и §5.2). Рад „Creating a Synthetic Evaluation Dataset for Serbian SentiWordNet Using Large Language Models“ односи се на формирање синтетичког скупа за евалуацију (§3.2 и §5.2.3–5.2.4). Рад „A Semi-Automated LLM-Based Framework for Word Sense Disambiguation in Serbian“ повезан је са поступком разрешења вишезначности речи, анотацијом и ревизијом инвентара значења речи (§3.3 и §5.3–5.7).

Радови настали у коауторству са Михаилом Шкорићем односе се на корпусна истраживања и основу обучавања језичких модела. Саопштење „Automating Synset Example Creation: A Case Study with Serbian Wordnet and ChatGPT“ повезано је са генерисањем примера за синсете, а „SrELEXIS-WSD: Hybrid Semi-Automated WSD for Serbian with Large Language Models—Results and Challenges“ са поступком анотације и евалуације значења. Остали радови о препознавању и повезивању именованих ентитета показују ширину истраживачког рада кандидата. Само коришћење модела описаног у коауторском раду не значи да сваки резултат тог рада представља самосталан допринос ове дисертације.

3.3. Програмске реализације и расположивост резултата

Практични резултат истраживања представља скуп програмских компонената за рад са Српским ворднетом и примену његових података у обради текста. Библиотека `srpskiwn` обезбеђује приступ синсетовима, литералима, глосама и семантичким релацијама, као и припрему скупова поларитета сентимента и претварање синсета у текстуалне улазе за моделе. Ова реализација непосредно подржава поступке описане у §5.1 и §5.2 дисертације.

Експериментална програмска подршка обухвата припрему и испитивање лексикона сентимента, генерисање синтетичког евалуационог скупа, полуаутоматску анотацију речи у тексту одговарајућим значењем и проверу инвентара значења речи, дестилацију модела, предлагање српских синсета, поравнање значења речи у различитим речницима и два начина примене података о значењу – анализу сентимента и циљану проверу машинског превођења. Њена вредност је у повезивању описаних поступака са проверљивим улазним подацима, подсшавањима и излазним подацима. Могућност поновног извођења експеримената зависи од доступности података и верзија програма коришћених у експериментима.

Јавно су доступни репозиторијум модела и лексикона за анализу сентимента (<https://github.com/sasa5linkar/Serbian-WordNet-Sentiment-Lexicon-Analysis>), репозиторијум синтетичког евалуационог скупа (<https://github.com/sasa5linkar/SWN-synth-eval-set>), програм за превођење и предлагање синсета (https://github.com/sasa5linkar/wordnet_autotranslate-) и заједнички репозиторијум уз рад о разрешењу вишезначности речи (<https://github.com/te-sla/A-Semi-Automated-LLM-Based-Framework-for-Word-Sense-Disambiguation-in-Serbian>).

На платформи Hugging Face, у оквиру налога Tanor (<https://huggingface.co/Tanor>), доступна су 24 модела за процену сентимента глоса синсета при изградњи лексикона S5–S7. За сваку од три породице — BERTi^ć, GPT2-Orao и Jerteh-355 — објављена су по четири пара класификатора позитивног и негативног поларитета, са ознакама итерација 0, 2, 4 и 6. Картице модела садрже опис намене, податке о лиценцама и везе ка пратећем коду.

Дана 25. септембра 2026. године објављена су и следећа програмска издања, којима су поступци из дисертације учињени доступним за даљу употребу и проверу:

- srpskiwn, верзија 1.0.2 – библиотека за учитавање Српског ворднета, претрагу синсета, семантичке релације и вредности сентимента (§5.1–5.2):
<https://github.com/sasa5linkar/srpskiwn>.
- serbian-wsd-distillation, верзија 0.1.0 – припрема парова контекста и дефиниције, обука и евалуација дестилованих модела за разрешење вишезначности речи, са подсшавањима и збирним резултатима (§3.4 и §5.8):
<https://github.com/sasa5linkar/serbian-wsd-distillation>.
- serbian-sentiment-wsd-evaluation, верзија 0.1.0 – евалуација лексикона сентимента на нивоу реченице, са избором значења или усредњавањем вредности, забележеним праговима и табелама резултата (§5.11):
<https://github.com/sasa5linkar/serbian-sentiment-wsd-evaluation>.
- serbian-wordnet-dictionary-alignment, верзија 0.1.0 – поравнање речничких значења са Српским ворднетом; објављени су код, идентификатори и ознаке за 3.414 кандидатских парова, без изворних дефиниција и глоса, и поступак израчунавања збирних резултата (§5.10):
<https://github.com/sasa5linkar/serbian-wordnet-dictionary-alignment>.

- o serbian-wsd-mt-evaluation, верзија 0.1.0 – припрема примера, преводилачки поступци и евалуација утицаја напомена о значењу на машинско превођење, са метаподацима експеримената и самосталним примером:
<https://github.com/sasa5linkar/serbian-wsd-mt-evaluation>.
- o serbian-wordnet-sentiment-toolkit, верзија 0.1.1 – алат за анализу сентимента који повезује обраду текста, избор значења и вредности сентимента, уз приказ доприноса појединачних јединица и покривености ресурса. Садржи програмски интерфејс, интерфејс командне линије, документацију и примере:
<https://github.com/sasa5linkar/serbian-wordnet-sentiment-toolkit>.

На дан 28. септембра 2026. године постале су јавно доступне и тежине три дестилована модела за разрешење вишезначности речи из §5.8, заједно са конфигурацијама, токенизаторима и картицама модела:

- o E5 Large, експериментална ознака mling:
<https://huggingface.co/Tanor/serbian-wsd-distilled-e5-large>.
- o MiniLM, експериментална ознака simple:
<https://huggingface.co/Tanor/serbian-wsd-distilled-minilm>.
- o TeslaXLM, експериментална ознака tesla:
<https://huggingface.co/Tanor/serbian-wsd-distilled-teslaxlm>.

Пратећа документација повезује моделе са кодом за дестилацију, евалуацију сентимента и практичну примену и садржи примере читавања конкретних ревизија. Програмска издања имају лиценце, податке за цитирање, инсталационе пакете и самосталне примере, чиме је олакшана поновна употреба развијених поступака.

Јавна издања представљају програмске и моделске резултате истраживања, док њихово објављивање не представља нову експерименталну потврду. За поновно извођење појединих експеримената потребни су и одговарајући спољни ресурси, изворне предикције по јединицама и коначне реченичне ознаке, који нису у целини обухваћени овим издањима. Практични алат за сентимент изведен је из истраживања и не замењује засебну проверу резултата из §5.11.

Коауторски рад о разрешењу вишезначности речи, корпуси и инвентари обухватају доприносе коаутора и аноатора. Постојећи Српски ворднет, Princeton WordNet, Senti-pol-sr, ELEXIS и коришћене библиотеке и модели представљају полазне ресурсе са сопственим ауторима.

4. ОЦЕНА ДИСЕРТАЦИЈЕ

4.1. Оригиналноост и однос према постојећим истраживањима

Дисертација повезује више поступака развоја Српског ворднета: процену сентимента синсета, разрешење вишезначности речи, ревизију инвентара значења, пренос знања на мање моделе и припрему кандидатских записа за лексикографску проверу. Предложена решења су усмерена на српски језички материјал и на ограничења непосредног преношења оцена, дефиниција и категорија из синглеских ресурса.

4.2. Методологија и поузданост резултата

Методологија комбинује аутоматску обраду, поређење са полазним методама и ресурсима и ручну језичку проверу. Код разрешења вишезначности речи разграничени су развојно-аналитички део од 120 реченица и издвојени тест скуп од 300 реченица. Сребрни подаци за обуку мањих модела разликују се од ручно проверених података за евалуацију. Резултати се приказују по типовима јединица, моделима и итерацијама, чиме се избегава свођење различитих поставки на једну збирну вредност.

При превођењу синсета и поравнању ресурса ручно проверени узорци мањи су од целокупних аутоматски обрађених скупова. Удео позитивних ознака које је доделио модел зато се не може тумачити као тачност. На пример, 74,0% позитивно означених кандидатских парова у поравнању није 74,0% потврђених исправних поравнања; ни збир тачних и блиских одговора у ручном узорку не представља удео потпуно тачних одговора.

У опису екстерне евалуације у анализи сентимента постоји несагласност у погледу избора прага: на стр. 57 дисертације наводи се избор на развојном скупу, а на стр. 119 избор најбољег прага на евалуационом скупу. Доступни програмски код испитује прагове на истом означеном скупу и из њега издваја најбољи резултат за извештај. Изворни резултати одговарајућег извршавања програма нису били доступни, па није било могуће у потпуности проверити њихову подударност са табелама 5-33 и 5-34. Ова неусаглашеност ограничава тумачење побољшања и статистичких поређења као потврде учинка на независном евалуационом скупу.

4.3. Исход провере хипотеза

Према резултатима и закључку достављене дисертације, хипотезе су поткрепљене у различитој мери:

- *H0*: подржана у ограниченом, полуаутоматском оквиру који укључује ручну проверу и лексикографску редакцију; није показано да се развој ресурса у целини може препустити моделу.
- *H1*: налази поређења лексикона пружају емпиријску подршку прилагођавању сентимента српском материјалу у односу на непосредно пресликавање из SentiWordNet-a. Домет спољне потврде на новинским реченицама ограничен је неусаглашеним описом избора прага.
- *H2*: највећа подршка добијена је у задатку разрешења вишезначности речи; код речничког поравнања односи се на предлагање и класификацију кандидатских веза, уз неопходну ручну проверу.
- *H3*: делимична подршка контролисаном формирању кандидата и концептуално усмереном поступку; обим ручне провере ограничава могућност уопштавања налаза.
- *H4a*: у табелама тезе S7-avg и S7-wsd постижу више вредности макро-F1 од S0. За S7-avg у односу на L0 наведена коригована р-вредност износи 0,0525 и није испод прага 0,05. Неусаглашен опис избора прага ограничава тумачење статистичке потврде побољшања.

- *H4b*: није потврђена у испитаној поставци. Једноставно додавање напомене о значењу није донело стабилно побољшање машинског превода. Налаз се не може проширити на све могуће начине укључивања значења у преводилачке системе.

Емпиријска подршка хипотези *H4* као целини ограничена је и разликује се у зависности од задатка.

4.4. Применљивост и ограничења

Практична применљивост резултата огледа се у припреми и провери лексичко-семантичких ресурса за српски језик, означавању значења у корпусном материјалу, процени сентимента и усмеравању лексикографског рада ка предложеним допунама и везама. Мањи модели представљају могућност економичније примене знања добијеног сложенијим поступком анотације, уз промене тачности које зависе од модела.

Границе примене одређене су покривеношћу инвентара значења речи, квалитетом и грануларношћу самих значења, обимом ручно проверених узорака и начином укључивања ресурса у конкретан задатак. Генерисани синсети и предложене везе остају кандидати до стручне провере. Налази добијени на новинским текстовима и конкретном скупу за разрешење вишезначности речи не доказују исти учинак у свим доменима. Укључивање података о значењу у испитаној поставци није донело стабилно побољшање машинског превода.

Правци даљег рада обухватају проширење ручно проверених података, укључивање више анотатора, унапређење покривености и повезаности ресурса, независну проверу избора прагова и контролисаније начине укључивања података о значењу у машинско превођење.

4.5. Самосталност и провера оригиналности

Кандидат је током израде докторске дисертације показао способност за самосталан научноистраживачки рад. Ова способност огледа се у критичком сагледавању литературе, избору и примени истраживачких метода, програмској реализацији поступака, обради и тумачењу резултата, као и у разматрању ограничења спроведених истраживања.

Оригиналност докторске дисертације проверена је на начин прописан Правилником о поступку провере оригиналности докторских дисертација које се бране на Универзитету у Београду („Гласник Универзитета у Београду”, бр. 204 од 22. јуна 2018. године). Према извештају софтвера iThenticate од 25. септембра 2026. године, укупан степен подударања текста износи 2%, уз искључену библиографију и текст под наводницима. Подударање са сваким појединачним извором мање је од 1%.

Прегледана подударања обухватају делове раније објављеног и цитираног коауторског рада кандидата, административно навођење наслова дисертације, уобичајене академске формулације, стручне називе и формалне податке. Образложење наведених подударања дато је у посебној оцени извештаја о провери оригиналности.

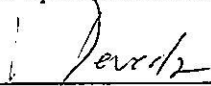
5. ЗАКЉУЧАК И ПРЕДЛОГ КОМИСИЈЕ ЗА ОДБРАНУ ДОКТОРСKE ДИСЕРТАЦИЈЕ

Комисија за оцену докторске дисертације констатује да је дисертација „Машинско учење и велики језички модели у развоју семантичких мрежа и њиховој примени на аутоматско разумевање текста“ кандидата Саше Петалинкара оригинално и самостално научно дело и да су се стекли услови за њену јавну одбрану. Комисија стога предлаже Већу за студије при Универзитету у Београду да прихвати с позитивном оценом реферат о оцени докторске дисертације под насловом „Машинско учење и велики језички модели у развоју семантичких мрежа и њиховој примени на аутоматско разумевање текста“, коју је урадио кандидат Саша Петалинкар под менторством проф. др Ранке Станковић и проф. др Јелене Граовац.

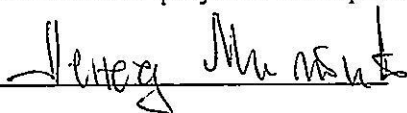
У Београду, _____

ЧЛАНОВИ КОМИСИЈЕ:

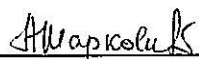
Проф. др Владан Девеџић
Факултет организационих наука Универзитета у Београду



Проф. др Ненад Митић, редовни професор у пензији
Математички факултет Универзитета у Београду



Др Александра Марковић, виши научни сарадник
Институт за српски језик САНУ



ОЦЕНА ИЗВЕШТАЈА О ПРОВЕРИ ОРИГИНАЛНОСТИ ДОКТОРСKE ДИСЕРТАЦИЈЕ

Извршен је преглед извештаја о провери оригиналности докторске дисертације „Машинско учење и велики језички модели у развоју семантичких мрежа и њиховој примени на аутоматско разумевање текста“ аутора Саше Петалинкара која је спроведена помоћу софтвера iThenticate.

Утврђено је следеће:

- Укупан степен подударања текста, према извештају iThenticate од 25.09.2026. године, износи 2%. Из поређења су искључени библиографија и текст под наводницима. Наведено је 65 извора, а подударање са сваким појединачним извором мање је од 1%.
- Подударност са сајтом Филолошког факултета (извор 3) односи се на наслов ове дисертације, наведен у Извештају о раду Факултета за 2025. годину (стр. 407). Доц. др Милош Утвић у том извештају наводи чланство у комисији за оцену научне заснованости теме и испуњености услова кандидата. Реч је о административном навођењу кандидатове дисертације.
- Прегледана подударања обухватају резултате из раније објављеног коауторског рада кандидата (Petalinkar, Stanković и Ikonić Nešić, 2025), који је цитиран у дисертацији, као и кратке уобичајене академске формулације, библиографске наводе, стручне називе, ознаке модела и мера, интернет адресе и формалне податке. Подударности са радом Марка Делића односе се на називе основних целина и уобичајена упућивања на табеле и слике.
- Извештај садржи и техничку ознаку о 53 визуелно слична знака из различитих писама на 23 странице. У тексту су уочени појединачни случајеви мешања латиничних и ћириличних слова унутар речи. PDF извештај не садржи појединачни списак свих означених позиција. Ова ознака сама по себи не представља налаз неоригиналности.

Изјављујемо да извештај о провери оригиналности докторске дисертације доказује потпуну оригиналност докторске дисертације „Машинско учење и велики језички модели у развоју семантичких мрежа и њиховој примени на аутоматско разумевање текста“ аутора Саше Петалинкара.

Минимални утврђени степен преклапања, који је немогуће избећи, ни на који начин не нарушава оригиналност дисертације. Стога предлажемо да се прописани поступак припреме за њену одбрану настави.

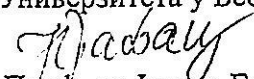
Београд, 25.09.2026.

Ментори:



Проф. др Ранка Станковић
редовни професор

Рударско-геолошки факултет Универзитета у Београду



Проф. др Јелена Граовац
ванредни професор

Математички факултет Универзитета у Београду