

ИЗБОРНОМ ВЕЋУ ФИЛОЛОШКОГ ФАКУЛТЕТА УНИВЕРЗИТЕТА У БЕОГРАДУ

Предмет: Извештај Комисије за избор асистента за ужу научну област *Библиотечка информатика*, предмет *Проналажење информација*

На основу одлуке Изборног већа Филолошког факултета Универзитета у Београду бр. 2624/1 од 19. 9. 2024. године изабрани смо у Комисију за избор кандидата који су се пријавили на конкурс за избор једног асистента за ужу научну област *Библиотечка информатика* (предмет *Проналажење информација*), који је објављен у листу “Послови” бр. 1117-1118 од 6. 11. 2024. године. На основу увида у конкурсну документацију, као и провере и анализе добијеног материјала, а у складу са чланом 84 Закона о високом образовању Републике Србије и чланом 113 Статута Филолошког факултета Универзитета у Београду, Комисија има част да Већу поднесе следећи

ИЗВЕШТАЈ

На расписани конкурс пријавила се једна кандидаткиња — мср Милица Иконић Нешић, докторанткиња Универзитета у Београду, запослена у звању асистента на Катедри за библиотекарство и информатику Филолошког факултета у Београду.

Кандидаткиња мср Милица Иконић Нешић је, уз пријаву на расписани конкурс, приложила сву потребну документацију: оверене копије уверења о држављанству и извода из матичне књиге рођених, оверене фотокопије диплома о завршеним основним и мастер академским студијама, потврду о статусу редовног студента докторских академских студија, биографију са библиографијом и радовима.

А. БИОГРАФИЈА КАНДИДАТКИЊЕ

ОБРАЗОВАЊЕ

Милица Иконић Нешић је рођена 27. 5. 1991. године у Београду. Основну школу је завршила у Барајеву, као носилац Вукове дипломе. Завршила је Математичку гимназију у Београду са просечном оценом 4,95. Дипломирала је на Математичком факултету Универзитета у Београду, на смеру теоријска математика и примена, 29. 9. 2018. године, са просечном оценом 9,27. Степен мастер математике стекла је 29. 9. 2020. године одбраном рада „Мартингални H_p простори”, под менторством професора др Драгољуба Кечкића (просечна оцена на мастер академским студијама 10,00). Тренутно је студент треће године Мултидисциплинарних докторских студија Универзитета у Београду, модул Интелигентни системи (две године мировања због породилског одсуства). На докторским студијама је положила све испите и одбранила приступни рад у року (за три године), са просечном оценом 10,00.

НАСТАВНО-ПЕДАГОШКИ РАД

Кандидаткиња Милица Иконић Нешић је од 2018. до 2020. године радила као сарадник у настави на Математичком факултету Универзитета у Београду за предмете основних академских студија *Анализа 1* на Катедри за анализу и *Геометрија 2* на Катедри за геометрију. Запослена је на Филолошком факултету Универзитета у Београду од фебруара 2021. године као асистент на Катедри за библиотекарство и информатику за предмете *Информатички практикум 2*, *Информатички практикум 3*, *Дигитални текст 1*, *Дигитални текст 2*, *Структура информација 1*, *Структура информација 2*, *Језичке технологије 1*, *Језичке технологије 2*, *Мултимедијални документи*, *Проналажење информација* (основне академске студије), *Напредне методе у проналажењу информација*, *Напредне језичке технологије*, *Дигитална хуманистика и библиотеке - ресурси и алати*, *Технологије семантичког веба* (мастер академске студије).

У периоду од 1.марта 2023. до 1.маја 2024. године користила је породилско одсуство и одсуство са рада ради неге детета.

Укупна просечна оцена педагошког рада за појединачне предмете од претходног избора у звање: *Дигитални текст 1* (4,78), *Информатички практикум 1* (4,70), *Информатички практикум 2* (5,00), *Информатички практикум 3* (4,90).

НАУЧНО-ИСТРАЖИВАЧКИ И СТРУЧНИ РАД

Кандидаткиња Милица Иконић Нешић је, у својству истраживача, ангажована:

- (1) у COST акцији „Distant Reading for European Literary History“ (COST Action CA16204), 2021–2024, <https://www.distant-reading.net/>;
- (2) на међународном истраживачком пројекту „It-Sr-NER: CLARIN Compatible NER and Geoparsing Web Services for Parallel Texts: Case Study Italian and Serbian“, 1. 6. 2022 - 30. 9. 2022. године, <https://github.com/jerteh/It-Sr-NER>;
- (3) на домаћем истраживачком пројекту *TESLA - Text Embeddings – Serbian Language Applications*, подржаног од стране Фонда за науку Републике Србије, #7276, почев од 1. 5. 2024. године, <https://tesla.rgf.bg.ac.rs/>;
- (4) у COST акцији „Global Network on Large-Scale, Cross-domain and Multilingual Open Knowledge Graphs“ CA23147, чији се почетак очекује у марту 2025. године.

Области интересовања кандидаткиње су библиотека информатика, обрада природног језика, рачунарска лингвистика, рачунарска лексикографија, проналажење информација, анализа сентимента, препознавање именованих ентитета, машинско учење, језички модели, семантички веб и математика.

Кандидаткиња је у свим наведеним пројектима примењивала различите технологије и методологије у проналажењу и екстракцији информација из текста.

На пројекту (1) учествовала је у обучавању модела за препознавање именованих ентитета на текстовима старих српских романа, први пут објављених у периоду од 1840-1920. године, а такође и на уносу 120 романа колекције *srpELTeC* (укључујући и метаподатке), као и одговарајућих издања (прва, штампана, дигитална и електронска), у отворену базу знања Википодаци (в. радове 3.1.10, 3.1.9, 3.1.8, 3.1.7 и 4.1.2).

На пројекту (2) кандидаткиња је, уз тим истраживача, развила веб сервис *It-Sr-NER*, компатибилан са инфраструктуром *CLARIN*, за препознавање именованих ентитета у паралелним текстовима, са студијом случаја на италијанском и српском језику (в. радове 3.1.6 и 4.1.1).

Кандидаткиња активно учествује на пројекту (3) где тренутно ради на развоју аотираног корпуса *TESLA-NER-NEL*. Корпус *TESLA-NER-NEL* ће представљати проширење постојећег корпуса *srpELTeC-gold*, садржаће више од 150,000 реченица са обележеним именованим ентитетима (особама, локацијама, професијама, организацијама, демонимима, догађајима и уметничким делима) повезаних са одговарајућим ставкама у википодацима. Са тимом истраживача развила је неколико модела за препознавање именованих ентитета заснованих на архитектури *BERT*, као и модел за повезивање именованих ентитета са википодацима користећи конволуциону неуронску мрежу (*CNN*) (в. радове 3.1.5 и 3.1.2). Такође, на пројекту је учествовала у истраживању и представљању резултата моделовања тема у текстовима корпуса *srpELTeC-gold* на основу анализе сентимента (в. радове 3.1.1 и 3.1.3).

СТРУЧНО-ПРОФЕСИОНАЛНИ ДОПРИНОС

Кандидаткиња Милица Иконић Нешић је члан:

- (1) уредничког одбора часописа *Journal of Natural Language Processing*, <https://www.cambridge.org/core/journals/natural-language-engineering>;
- (2) Групе за језичке технологије Универзитета у Београду и Друштва за језичке ресурсе и технологије - *JePTeX* (<http://jerteh.rs>).
- (3) члан организационог одбора међународне конференције *RANLP 2021*, Бугарска, <https://ranlp.org/ranlp2021/contacts.php>;
- (4) организационог одбора међународне конференције *South Slavic Languages in the Digital Environment - JuDig*, Филолошки факултет, Београд, 2024, <https://judig.jerteh.rs/inner-page.php?p=organizing-committee>.

ДОПРИНОС АКАДЕМСКОЈ И ШИРОЈ ЗАЈЕДНИЦИ

Кандидаткиња је у више наврата била један од представника Катедре за библиотекарство и информатику на Отвореним вратима Филолошког факултета Универзитета у Београду у оквиру промотивних активности Факултета.

ПОЗНАВАЊЕ СТРАНИХ ЈЕЗИКА

Кандидаткиња у биографији наводи да се служи енглеским језиком.

ВЕШТИНЕ ОД ЗНАЧАЈА ЗА УЖУ НАУЧНУ ОБЛАСТ И ПРЕДМЕТЕ ЗА КОЈЕ СЕ КАНДИДАТ БИРА

Кандидаткиња у својој пријави наводи да од програмских језика познаје, C#, Java, Python, док професионално користи технологије INCEPТION, Matlab, Wikidata Query Service, Gephi, OpenRefine, као и алате за обраду природног језика scikit-learn, NumPy, Pandas, TensorFlow, spaCy, gensim, fastText, NLTK и Unitex.

Б. БИБЛИОГРАФИЈА КАНДИДАТКИЊЕ

1. ГРУПА РЕЗУЛТАТА М10

1.1. РАД У ТЕМАТСКОМ ЗБОРНИКУ МЕЂУНАРОДНОГ ЗНАЧАЈА – М14

1.1.1. Крстев Ц., Станковић Р., Шандрих Тодоровић Б., Иконић Нешић, М. (2023). Нове технологије за оживљавање старих текстова. У зборнику радова Међународне научне конференције *Дигитална хуманистика и словенско културно наслеђе II*, Београд, 28-29. јуни 2021. Београд : Савез славистичких друштава Србије.

<https://enauka.gov.rs/handle/123456789/854220>

<http://dr.rgf.bg.ac.rs/s/repo/item/8415>

2. ГРУПА РЕЗУЛТАТА М20

2.1. РАД У МЕЂУНАРОДНОМ ЧАСОПИСУ ИЗУЗЕТНИХ ВРЕДНОСТИ – М21а (1)

2.1.1. Škorić, M., Stanković, R., Ikonić Nešić, M., Byszuk, J., & Eder, M. (2022). Parallel Stylometric Document Embeddings with Deep Learning Based Language Models in Literary Authorship Attribution. *Mathematics*, 10(5), 838.

<https://doi.org/10.3390/math10050838>

<https://repff.fil.bg.ac.rs/handle/123456789/1505>

<https://www.mdpi.com/2227-7390/10/5/838/pdf>

3. ГРУПА РЕЗУЛТАТА М30

3.1. САОПШТЕЊА СА МЕЂУНАРОДНИХ СКУПОВА ШТАМПАНА У ЦЕЛИНИ – М33 (10)

3.1.1. Ikonić Nešić, M., Petalinkar, S., Škorić, M., Stanković, R., & Rujević, B. (2024, September). Advancing Sentiment Analysis in Serbian Literature: A Zero and Few-Shot Learning Approach Using the Mistral Model. In *Proceedings of the Sixth International Conference on Computational Linguistics in Bulgaria (CLIB 2024)* (pp. 58-70).

<https://dr.rgf.bg.ac.rs/s/repo/item/8805>

3.1.2. Ikonić Nešić, M., Petalinkar, S., Stanković, R., Utvić, M., & Kitanović, O. (2024, September). SrpCNNeL: Serbian Model for Named Entity Linking. In *2024 19th Conference on Computer Science and Intelligence Systems (FedCSIS)* (pp. 465-473). IEEE. <https://doi.org/10.15439/2024F8827>
<https://repff.fil.bg.ac.rs/handle/123456789/1570>

3.1.3. Mihajlov, T., Ikonić Nešić, M., Stanković, R., & Kitanović, O. (2024, September). Topic Modeling of the srpELTeC Corpus: A Comparison of NMF, LDA, and BERTopic. In *2024 19th Conference on Computer Science and Intelligence Systems (FedCSIS)* (pp. 649-653). IEEE. <https://doi.org/10.15439/2024F1593>

3.1.4. Stanković, R., Ikonić Nešić M., Perišić, O., Škorić, M., & Kitanović, O. (2024). Towards Semantic Interoperability: Parallel Corpora as Linked Data Incorporating Named Entity Linking. In *The 9th Workshop on Linked Data in Linguistics: Resources, Applications, Best Practices (LDL-2024)@ LREC-COLING-2024* (pp. 115-125). ELRA Language Resources Association and the International Committee on Computational Linguistics. <https://iris.unito.it/handle/2318/1979230>
<https://aclanthology.org/2024.ldl-1.15.pdf>

3.1.5. Ikonić Nešić, M., Petalinkar, S., Škorić, M., & Stanković, R. (2024, March). BERT Downstream Task Analysis: Named Entity Recognition in Serbian. In *Conference on Information Technology and its Applications* (pp. 333-347). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-71419-1_29

3.1.6. Perišić, O., Stanković, R., Ikonić Nešić, M., & Škorić, M. (2023). It-Sr-NER: CLARIN Compatible NER and Geoparsing Web Services for Italian and Serbian Parallel Text. In *Selected Papers from the CLARIN Annual Conference 2022 Prague, Czechia, 2022, 10-12 October* (pp. 99-110). Linköping University Electronic Press. <https://doi.org/10.3384/ecp198010>
<https://dr.rgf.bg.ac.rs/s/repo/item/7464>

3.1.7. Ikonić Nešić, M., Stanković, R., Schöch, C., & Škorić, M. (2022, June). From ELTeC Text Collection Metadata and Named Entities to Linked-Data (and Back). In *Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference* (pp. 7-16). <https://aclanthology.org/2022.ldl-1.2.pdf>
<https://dr.rgf.bg.ac.rs/s/repo/item/6281>

3.1.8. Stanković, R., Košprdić, M., Ikonić Nešić, M., & Radović, T. (2022, June). Sentiment Analysis of Serbian Old Novels. In *Proceedings of the 2nd Workshop on Sentiment Analysis and Linguistic Linked Data* (pp. 31-38). <https://aclanthology.org/2022.salld-1.6.pdf>
<https://dr.rgf.bg.ac.rs/s/repo/item/6277>

3.1.9. Stanković, R., Krstev, C., Todorović, B. Š., Vitas, D., Škorić, M., & Ikonić Nešić, M. (2022, June). Distant Reading in Digital Humanities: Case Study on the Serbian Part of the ELTeC Collection. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC'22) Marseille, Franc.* (pp. 3337-3345). European Language Resource Association (ELRA)

<https://aclanthology.org/2022.lrec-1.356.pdf>

<https://dr.rgf.bg.ac.rs/s/repo/item/6279>

3.1.10. Šandrih Todorović, B., Krstev, C., Stanković, R., & Ikonić Nešić, M. (2021). Serbian NER&Beyond: The Archaic and the Modern Intertwined. In *Deep Learning Natural Language Processing Methods and Applications—Proc. of the Int. Conf. Recent Advances in Natural Language Processing (RANLP 2021)*, edited by Galia Angelova et al (pp. 1252-1260).

<https://dr.rgf.bg.ac.rs/s/repo/item/5139>

3.2. САОПШТЕЊЕ СА МЕЂУНАРОДНОГ СКУПА ШТАМПАНО У ИЗВОДУ – М34 (1)

3.2.1. Ranka Stanković, Christian Chiarcos, Milica Ikonić Nešić. Leveraging Linked Data, NIF, and CONLL-U for Enhanced Annotation in Sentence Aligned Parallel Corpora. *UniDive 2nd general meeting*, University of Naples L’Orientale, Italy, 7-9 February 2024.

<https://doi.org/10.5281/zenodo.11208944>

4. ГРУПА РЕЗУЛТАТА М50

4.1. РАДОВИ У НАЦИОНАЛНИМ ЧАСОПИСИМА – М53 (2)

4.1.1. Perišić, O., Stanković, R., Ikonić Nešić M., & Škorić, M. (2023). It-Sr-NER: Web Services for Recognizing and Linking Named Entities in Text and Displaying Them on a Web Map. *Infotheca — Journal for Digital Humanities*, 23(1), 61-77.

<https://doi.org/10.18485/infotheca.2023.23.1.3>

<https://dr.rgf.bg.ac.rs/s/repo/item/7790>

4.1.2. Ikonić Nešić, M., Stanković, R. & Rujević, B. (2021). Serbian ELTeC Sub-Collection in Wikidata. *Infotheca — Journal for Digital Humanities*, 21(2).

<https://doi.org/10.18485/infotheca.2021.21.2.4>

<https://repff.fil.bg.ac.rs/handle/123456789/1428>

Укупна научноистраживачка продукција. Кандидаткиња мср Милица Иконић Нешић је у коауторству укупно објавила 15 оригиналних радова и то:

- један рад у тематском зборнику међународног значаја – М14,
- један рад у међународном часопису изузетних вредности – М21а,
- десет саопштења са међународних скупова штампана у целини – М33,
- једно саопштење са међународног скупа штампано у изводу – М34 и
- два рада у националном часопису – М53.

Наведени резултати сведоче о посвећености кандидаткиње научно-истраживачком раду, посебно узимајући у обзир дужину временског периода у ком су настали, као и здравствене проблеме детета и саме кандидаткиње током породилског одсуства.

В. КВАЛИТАТИВНА ОЦЕНА НАУЧНО-ИСТРАЖИВАЧКОГ РАДА

Будући да је кандидаткиња мср Милица Иконић Нешић у звању асистента од фебруара 2021. године, сви њени радови су објављени после претходног избора у звање. Комисија у оцени издваја неке од радова.

ПРИКАЗ И ОЦЕНА РАДОВА ОБЈАВЉЕНИХ ПОСЛЕ ИЗБОРА У ЗВАЊЕ

В.1. Радови 3.1.9 (2022) и 1.1.1 (2023) представљају српски део вишејезичне колекције srpELTeC.

COST акција „Удаљено читање за европску историју књижевности“ један је од најважнијих пројеката усмерених на припрему вишејезичног корпуса ELTeC. Колекција ELTeC је првобитно замишљена да обухвати по 100 романа из периода 1840–1920 за 12 европских језика, укључујући српски, а у међувремену се проширила текстовима још 9 европских језика. У раду 1.1.1 је представљен детаљан процес припреме српске потколекције (srpELTeC). Овај процес обухвата избор романа према строгим критеријумима, структурну анотацију, обраду врсте речи, лематизацију и примену метода за анализу великих текстуалних корпуса.

Рад 3.1.9 детаљно разматра, не само поменуте кораке израде srpELTeC колекције: избор и преузимање текстова, њихову припрему и обраду, структурну анотацију, анотацију врсте речи, лематизацију и препознавање именованих ентитета, већ и примену савремених рачунарских метода за анализу и разумевање књижевних текстова, као и примену колекције у дигиталним хуманистичким наукама. Овај рад истиче значај доступности српске потколекције на различитим платформама, чиме је омогућена слободна доступност истраживачима и другим корисницима.

В.2. Рад 2.1.1 (2023) истражује нови приступ решавању задатка препознавања ауторства анализом књижевних текстова на седам језика претходно поменуте колекције ELTeC, међу којима је и srpELTeC. Укупно је анализиран 7051 сегмент од по 10.000 токена из 700 докумената, лематизираних и аотираних врстом речи. Коришћена је стилometriја, као метода статистичке анализе текста, заснована на вектору документа (енгл. document embedding), нумеричкој (векторској) репрезентацији докумената као целине. Аутори су искористили пакет Stylo у програму R како би развили 4 модела заснована на самосталним векторима документа који се међусобно разликују по томе да ли су засновани на речима, лемама, триграмима врста речи или маскама врста речи. Аутори су потом развили композитне моделе користећи постојећи модел mBERT за сваки тестирани језик, и комбинујући самосталне векторе докумената различитим матричним функцијама (просек, производ, минимум, максимум и норма l^2) како би произвели различите изведене векторе и упоредили их. Испоставило се да је композитни модел за већину језика (укључујући и српски) дао боље резултате у поређењу са самосталним моделима, нарочито у случају када су за потребе процене примењене мере тачност (енгл. accuracy и F1). Утврђено је да mBERT није имао значајан позитиван утицај на перформансе предложених метода.

В.3. Радови 3.1.8 (2022) и 3.1.1 (2024) представљају *анализу сентимента* српских старих романа колекције srpELTeC.

Рад 3.1.8 представља прву студију анализе сентимента на поменутој колекцији srpELTeC. Циљ рада је да истражи могућности анализе сентимента српских романа из периода 1840–1920, уз прилагођавање постојећих метода специфичностима српског језика. Посебна пажња посвећена је развоју и проширењу лексикона сентимента који обухвата информације о поларитету и флективне облике. У раду се упоређују методе анализе сентимента засноване на лексиконима NRC, AFFIN и Bing са методама машинског учења, као што су логистичка регресија, наивни Бајес, стабла одлучивања, случајне шуме, SVM и k-NN. Истражује се и неуспешност примене модела обучених на другим корпусима, као што су рецензије филмова, за анализу сентимента у старим књижевним текстовима.

У раду 3.1.1 користи се велики језички модел (LLM) Mistral за технику учења засновану на такозваним приступима „zero-shot“ и „few-shot“. Главни приступ уводи иновације кроз осмишљавање истраживачких упита (енгл. prompt) који укључују текст са инструкцијама за класификацију без примера и на основу неколико примера, омогућавајући језичком моделу да класификује осећања у позитивне, негативне или објективне категорије. Ова методологија има за циљ да поједностави анализу сентимента ограничавањем одговора, чиме се повећава прецизност класификације. Резултати анализе сентимента у оба сценарија, "zero-shot" и "few-shot", за разлику од поменуте методологије у раду 3.1.8, показали су да је приступ „zero-shot“ бољи, постигавши прецизност од 68,2%.

В.4. Радови 3.1.9 (2022) и 3.1.5 (2024) представљају резултате *обучавања модела за препознавање именованих ентитета* у тексту применом конволуционих неуронских мрежа (CNN) и BERT трансформера, тим редом. Рад 3.1.2 (2024) приказује методологију и технику тренирања модела за *повезивање препознатих именованих ентитета* у тексту са отвореном базом знања Википодаци.

У раду 3.1.9 се користи srpELTeC-gold, анотирани корпус од 14 целих романа и 11 делова из различитих романа колекције srpELTeC, за обуку модела за препознавање седам различитих типова именованих ентитета (енгл. Named Entities): особа, локација, организација, демонима, догађаја, занимања и уметничких дела. Модел користи конволуционе неуронске мреже (CNN). Процена модела помоћу метрике F1 даје приближан резултат 91% на тестираном скупу података.

Рад 3.1.5 упоређује различите архитектуре и технике за обучавање модела за препознавање именованих ентитета (енгл. NER) у текстовима на српском језику интеграцијом архитектуре BERT у библиотеку отвореног кода spaCy. Модели су обучени да препознају седам различитих типова именованих ентитета (особе, локације, организације, професије, догађаји, демоними и уметничка дела), а обучавани су на скупу srpELTeC-gold, јавним новинским чланцима и реченицама генерисаним из базе знања Википодаци и лексикографске базе знања Leximika. Представљене су различите конфигурације и неколико процеса обучавања модела који се разликују по сложености и функционалности. Неке конфигурације су посвећене искључиво задацима препознавања именованих ентитета, док друге укључују додатне функције попут аутоматске анотације врстом речи и лематизације. Један од кључних аспеката овог рада је тестирање различитих верзија модела BERT, са различитим архитектурама, величинама и предобученим корпусима српског језика. Овај приступ омогућава да се процени компромис између сложености модела и његових перформанси, те пружа дубље разумевање тога како различите конфигурације утичу на ефикасност и ефективност задатка препознавања именованих ентитета на српском језику.

Рад 3.1.2 приказује развој модела SrpCNNeL за повезивање именованих ентитета (енгл. Named Entity Linking, скр. NEL) са базом знања Википодаци за српски језик. Модел је обучен да препозна и повеже раније споменутих седам различитих типова именованих ентитета (особе, локације, организације, професије, догађаји, демоними и уметничка дела) на скупу података који садржи реченице из романа колекције srpELTeC, правних докумената, као и реченице генерисане из базе знања Википодаци и лексикографске базе знања Leximika. Процена резултованог модела применом метрике F1 даје приближан резултат 80% на тестираном скупу података.

В.5. У радовима 4.1.2 (2021), 3.1.7 (2022), 3.2.1 (2024) и 3.1.4 (2024), представљене су *технике семантичког веба* и представљање података као *повезаних отворених података* (енгл. Linked Open Data).

Радови 4.1.2 и 3.1.7 представљају процес који је довео до повезивања романа колекције srpELTeC и проширене колекције старих српских романа srpELTeC-ext са отвореном базом знања Википодаци, при чему је потоњи проширење истраживања представљеног у првом на примеру седам потколекција колекције ELTeC (енглеска, португалска, француска, словеначка, немачка, мађарска и српска). Док рад 4.1.2 у оквиру процеса израде повезаног скупа података укључује само преузимање метаподатака о романима, препознавање именованих ентитета, трансформацију, мапирање и попуњавање базе знања ставкама за одговарајуће романе и њихова издања (прво, штампано, дигитално и електронско), дотле рад 3.1.7 обухвата и извоз у формат NIF (NLP Interchange Format). OpenRefine и QuickStatements су алати који су изабрани као најбоље опције приликом реализације задатка повезивања са Википодацима и њиховог ажурирања. У оба рада су представљени примери SPARQL-упита за преузимање аутора, наслова романа, места издања и других метаподатака са различитим опцијама визуализације, док рад 3.1.7 укључује и статистичке прегледе.

Саопштење штампано у изводу (3.2.1) описује како се повезани подаци и формат NIF могу искористити за представљање електронског речника вишечланих одредница (енгл. multiword expression, скр. MWE), односно паралелних корпуса, са циљем да се одреднице повежу са својим појављивањима у изабраним корпусима. У питању је само корак у истраживању чији је циљ проширење паралелног корпуса ELEXIS-WSD (тренутно 10 језика са тенденцијом проширења) у коме су аотирана значења.

У раду 3.1.4 се приказују резултати истраживања у вези са припремом паралелних корпуса, при чему је фокус на трансформацији у RDF графове применом формата NIF за лингвистичку анотацију. Описан је паралелни корпус коришћен у овој студији случаја, потом процес његове анотације врстом речи и лематизације, затим процес препознавања именованих ентитета (NER) у корпусу и њихово повезивање (NEL), а на крају конверзија анотираног текста у RDF, односно формат NIF. Произведене датотеке у формату NIF су оцењене применом SPARQL-упита на платформи Apache Jena Fuseki. Такође се разматра повезивање повезаних података (Linked Data) и истраживања дигиталне хуманистике, као и неки недостаци везани за опсежност трансформације. Аутори констатују да концепт семантичке интероперабилности у контексту повезаних података и паралелних корпуса осигурава да подаци размењени између система носе дељена и добро дефинисана значења, што омогућава ефикасну комуникацију и разумевање.

В.6. У раду 3.1.3 (2024) пореде се перформансе модела за *моделовање тема* на романима колекције srpELTeC. као што су латентна Дирихлеова алокација (LDA), ненегативна матрична факторизација (NMF) и BERTopic, мерењем кохерентности тема (енгл. Topic Coherence, скр. TC) и разноврсности тема (енгл. Topic Diversity, скр. TD), као и квалитативном евалуацијом добијених тема. Аутори су, у случају BERTopic модела, упоређивали вишејезичне векторске реченичне репрезентације (енгл. multilingual sentence transformer embeddings) са једнојезичним векторским репрезентацијама за српски језик (Jertex-355). Резултати показују да је NMF постигао најбоље резултате у погледу кохерентности тема, док су BERTopic и Jertex-355 имали најбољу разноврсност тема. Такође, Jertex-355 је надмашио вишејезичне векторске репрезентације у погледу обе мере (TC и TD).

В.7. Радови 3.1.6 (2023) и 4.1.1 (2023) представљају резултате пројекта „It-Sr-NER: Веб сервис за препознавање, повезивање и мапирање именованих ентитета” за српски и италијански језик. Пројекат је реализован у сарадњи између Универзитета у Торину и Друштва за језичке ресурсе и технологије (JeRTeh), са циљем да се развије *веб сервис* It-Sr-NER за потребе анотације именованих ентитета, као што су особе, локације, организације, етничке групе, догађаји и уметничка дела у тексту, као и за приказивање локација на мапи.

ЗАКЉУЧНО МИШЉЕЊЕ И ПРЕДЛОГ КОМИСИЈЕ

Кандидаткиња Милица Иконић Нешић испуњава услове конкурса има просечну оцену на свим нивоима студија већу од 8, просечну оцену 9,27 на ОАС и просечну оцену 10,00 на МАС. На докторским студијама је положила све испите и одбранила приступни рад у року (за три године), са просечном оценом 10,00.

Када је реч о смислу за наставни рад, кандидаткиња мср Иконић Нешић од фебруара 2021. год. учествује у извођењу наставе на високошколском нивоу и има позитивне оцене педагошког рада у студентским анкетама.

Кандидаткиња испољава смисао за научноистраживачки рад, при чему су њена истраживања усмерена првенствено на изучавања библиотечке информатике, проналажења и екстракције информација, семантичког веба, (аутоматске) обраде природног језика и рачунарске лингвистике, па тиме истраживања укључују области од значаја за извођење наставе из предмета за које је конкурс расписан (Проналажење информација).

На основу свега наведеног, Комисија једногласно и са пуном увереношћу предлаже Изборном већу Филолошког факултета Универзитета у Београду да усвоји овај извештај и **изабере мср Милицу Иконић Нешић у звање и на радно место асистента за ужу научну област Библиотечка информатика, предмет Проналажење информација.**

У Београду,

24. 12. 2024. године

Чланови Комисије:

др Милош Утвић, доцент

Универзитет у Београду – Филолошки факултет

др Гордана Стокић Симончић, редовни професор

Универзитет у Београду – Филолошки факултет

др Ранка Станковић, ванредни професор

Универзитет у Београду – Рударско-геолошки факултет