

## ИЗБОРНОМ ВЕЋУ ФИЛОЛОШКОГ ФАКУЛТЕТА УНИВЕРЗИТЕТА У БЕОГРАДУ

Одлуком Изборног већа Филолошког факултета Универзитета у Београду бр. 1801/1 од 29. маја 2024. године именована је Комисија за писање извештаја о пријављеним кандидатима за избор у звање доцента за ужу научну област Библиотека информатика, предмет *Дигитални текст*, по конкурс који је објављен 19. јуна 2024. године у листу Националне службе за запошљавање *Послови* бр. 1097-1098. У Комисију су именовани: др Гордана Стокић Симончић, редовни професор Филолошког факултета Универзитета у Београду (председник Комисије), др Гордана Рудић, редовни професор Педагошког факултета у Сомбору, Универзитета у Новом Саду, и др Ранка Станковић, ванредни професор Рударско-геолошког факултета Универзитета у Београду. На основу прегледа конкурсне документације, Комисија подноси Изборном већу Филолошког факултета Универзитета у Београду следећи

### ИЗВЕШТАЈ

На конкурс за избор доцента за ужу научну област Библиотека информатика, предмет *Дигитални текст*, објављен у листу *Послови* број 1097-1098, од 19. јуна 2024. године, пријавио се др Милош Утвић, као једини кандидат. Он је поднео уредну пријаву на конкурс и сву неопходну документацију која сведочи да др Милош Утвић испуњава услове конкурса.

#### **I Биографија, општи подаци, стручна и наставна активност**

Милош Утвић је рођен 29. јануара 1976. године у Прокупљу. Основну школу и гимназију је завршио у Куршумлији. Дипломирао је на Математичком факултету Универзитета у Београду, смер рачунарство и информатика, 29. маја 2001. године. Степен магистра рачунарства је стекао 18. фебруара 2008. године на истом факултету, одбранивши магистарски рад са темом *Коначни аутомати у регуларној именованој деривацији*. Докторску дисертацију под насловом *Изградња референтног корпуса савременог српског језика* је одбранио 28. априла 2014. године на Филолошком факултету Универзитета у Београду и стекао степен доктора библиотечно-информатичких наука.

Од 2000. до 2001. године је радио на Математичком факултету Универзитета у Београду као стручни сарадник за вежбе из предмета *Преводиоци и интерпретатори* на Катедри за рачунарство и информатику. Од 3. децембра 2001. године ради као асистент-приправник за предмете *Информатика I-IV* на Катедри за библиотекарство и информатику на Филолошком факултету Универзитета у Београду. У звање асистента на Филолошком факултету Универзитета у Београду је биран 15. октобра 2008. године и поново 12. октобра 2011. године, и у том периоду је држао вежбе за предмете *Математика за библиотекаре и лингвисте*, *Информатичка писменост*, *Обрада текста*, *Интернет и веб технологије*, *Програмирање*, *Структура информација* (основне студије, 2006-2009), *Информатика за библиотекаре 1-2*, *Информатички практикум 1-4*, *Дигитални текст 1-2*, *Базе података и библиотечки информациони системи*, *Проналажење информација*, *Мултимедијални документи* (основне студије, 2009-) и *Електронско издаваштво и дигиталне библиотеке* (мастер студије 2010-).

Крајем септембра 2006. одржао је специјалан курс за постдипломце о креирању и одржавању паралелизованих корпуса на Институту за славистику Универзитета Карл Франц у Грацу.

Током школске 2009/2010. и 2009/2010. изводио је вежбе на предмету *Информатика* студентима мастер студија Језици, пословање и међународна трговина које су заједнички организовали Универзитет у Београду (Филолошки факултет) и Универзитет у Орлеану (Француска).

У звање доцента за ужу научну област Библиотечка информатика на Филолошком факултету Универзитета у Београду је изабран 18. новембра 2014. године, за предмет *Дигитални текст 2* и реизабран 17. децембра 2019. године. Почев од школске 2014/2015. године држао је предавања на предметима *Информатички практикум 1-3*, предавања и вежбе на предметима *Дигитални текст 2*, *Статистика у библиотекама*, *Базе података и библиотечки информациони системи*, *Базе података* (основне академске студије), предавања на предметима *Обликовање и одржавање садржаја на вебу* и *Електронско издаваштво и дигиталне библиотеке* (мастер академске студије), као и предавања на предмету *Математичка лингвистика* (докторске академске студије). Почев од школске 2020/2021. године држи предавања и за предмет *Информатика за библиотекаре 1* (основне академске студије), а ангажован је и на предмету *Дигиталне библиотеке и корпусна лингвистика* на мастер студијском програму Рачунарство у друштвеним наукама Универзитета у Београду.

Као члан Катедре за библиотекарство и информатику Филолошког факултета Универзитета у Београду учествовао је са осталим члановима у формирању репозиторијума доиФил — репозиторијума дигиталних идентификатора објеката Филолошког факултета Универзитета у Београду 2016. године. У сарадњи са Михаилом Шкорићем је развио софтвер и веб сајт репозиторијума доиФил (<http://doi.fil.bg.ac.rs>). Од 2016. године обавља дужност администратора метаподатака репозиторијума доиФил и ради као један од редактора DOI-бројева које додељује

Филолошки факултет Универзитета у Београду у својству издавача и овлашћеног члана агенције за регистрацију DOI-бројева Crossref.

Члан је Групе за језичке технологије Универзитета у Београду, Друштва за језичке ресурсе и технологије (<http://jerteh.rs>), као и Комисије за корпус Одбора за стандардизацију српског језика САНУ. Од програмских решења је развио десетак подсистема у оквиру Корпуса савременог српског језика, а у сарадњи са др Ранком Станковић и интегрисано окружење за припрему паралелизованих корпуса и адаптацију постојећих веб сучеља (енгл. web interface) за претрагу корпуса засновану на примени веб сервиса и екстерних лексичких ресурса. У сарадњи са осталим члановима Групе за језичке технологије Универзитета у Београду, учествовао је у изградњи више дигиталних језичких ресурса за српски језик:

- Корпус савременог српског језика СрпКор (<http://www.korpus.matf.bg.ac.rs>) и његов подскуп СрпЛемКор (<http://www.korpus.matf.bg.ac.rs/SrpLemKor>);
- паралелни енглеско-српски корпус СрпЕнгКор (<http://www.korpus.matf.bg.ac.rs/SrpEngKor>);
- паралелни француско-српски корпус СрпФранКор (<http://www.korpus.matf.bg.ac.rs/SrpFranKor>);
- специјализовани корпус из области рударства РудКор (<https://noske.rgf.rs>).

Као истраживач, учествовао је на следећим пројектима:

- 1743 *Интеракција текста и речника*, Министарство за науку Републике Србије (2002-2004);
- 148021 *Теоријско-методолошки оквир за модернизацију описа српског језика*, Министарство за науку Републике Србије и САНУ (2005–2010);
- *Ресурси Средње и Југоисточне Европе* (енг. *Central and South-East European Resources*, скр. *CESAR*), ICT Policy Support Programme, Grant agreement no.: 271022 (2011–2013), <https://cordis.europa.eu/project/id/271022>;
- ОН 178006 *Српски језик и његови ресурси: теорија, опис и примене*, Министарство за образовање и науку Републике Србије (2011–2019);
- ИИИ 47003 *Инфраструктура за електронски подржано учење у Србији*, Министарство за образовање и науку Републике Србије (2011–2019);
- 7276 *TESLA (Text Embeddings - Serbian Language Applications = Векторизација текста – апликације за српски језик)*, Фонд за науку Републике Србије, Програм ПРИЗМА (2023-), <https://tesla.rgf.bg.ac.rs/>.

Од страних језика говори енглески језик, а служи се и руским језиком.

Рецензент је часописа *Библиотекар*, *Читалиште* и *Инфотека*. У својству члана редакције часописа *Инфотека* обавља дужност редактора DOI-бројева. Био је ангажован као рецензент:

- радова за Зборник радова I Међународне научне конференције *Дигитална хуманистика и словенско културно наслеђе*,
- радова за Зборник радова међународне научне конференције *RANLP-2021 (Recent Advances in Natural Language Processing)*,
- радова за Зборник радова Међународне научне конференције *Дигитална хуманистика и словенско културно наслеђе 2*.
- за потребе акредитације семинара *Вики библиотекар* као програма сталног стручног усавршавања у библиотечко-информационој делатности (2018–2024).

Учествовао је у комисијама за одбрану:

- мастер рада Катарине Перић *Отворени приступ и академски дигитални репозиторијуми* (у својству члана комисије, датум одбране 28. IX 2015),
- мастер рада Милице Арсеновић *Тестирање софтвера* (у својству члана комисије, датум одбране 21. IX 2016);
- мастер рада Михаила Шкорића *Сврставање појмова на скали позитивно-негативно, применом истраживања података над корпусом текстова који садрже емотиконе* (у својству члана комисије, датум одбране 23. IX 2016);
- мастер рада Марије Пантић *Придевски атрибути уз именице „човек“, „жена“, „мушкарац“ и „муж“* (у својству члана комисије, датум одбране 29. IX 2017);
- мастер рада Анђеле Стошић *Развој и коришћење дигиталних колекција: на примеру Градске библиотеке „Владислав Петковић Дис“ у Чачку, Библиотеке града Београда и Народне библиотеке „Стеван Сремац“ у Нишу* (у својству члана комисије, датум одбране 10. VII 2018);
- мастер рада Анастасје Мандић *Аутоматска конверзија електронских библиографија из формата BibTeX у формат BibLaTeX* (у својству ментора, датум одбране 27. IX 2018);
- мастер рада Милице Васковић *Информациона писменост у школској библиотеци 21. века* (у својству члана комисије, датум одбране 1. VI 2021);
- мастер рада Немање Антанасијевића *Аутоматизована евалуација модела за препознавање именованих ентитета у обради текста* (у својству члана комисије, датум одбране 16. IX 2022);
- мастер рада Катарине Меглић *Дигитални репозиторијуми и политика отворене науке у Србији* (у својству члана комисије, датум одбране 4. VI 2024);
- докторске дисертације *Кумулативни ефекти експлоатације вишејезичних корпуса у настави страних језика* др Зорана Ж. Ристовића (у својству члана комисије, датум одбране 6.7.2016),  
<https://nardus.mpn.gov.rs/handle/123456789/6196>;
- докторске дисертације *Аутоматско препознавање и нормализација временских израза у неструктурираним новинским и медицинским текстовима на српском језику* др Јелене Јаћимовић (у својству члана комисије, датум одбране 30.8.2016),  
<https://nardus.mpn.gov.rs/handle/123456789/7930>;

- докторске дисертације *Развој модела за управљање рударском пројектном документацијом* др Александре Томашевић (у својству члана комисије, датум одбране 21.9.2018), <https://nardus.mpn.gov.rs/handle/123456789/10128>;
- докторске дисертације *Мрежа отворених података и језички ресурси у процесу изградње српско-немачког литерарног корпуса* др Јелене Андоновски (у својству члана комисије, датум одбране 24. I 2020), <https://nardus.mpn.gov.rs/handle/123456789/17601>;
- докторске дисертације *Citatna analiza kao metoda vrednovanja naučnog rada istraživača iz Srbije u periodu od 2000. do 2016. godine* др Сађе Антонић (у својству члана комисије, датум одбране 10. III 2020), <https://nardus.mpn.gov.rs/handle/123456789/17508>
- докторске дисертације *Речници у дигиталном добу – информатичка подршка за српски језик* др Биљане Рујевић (у својству члана комисије, датум одбране 31. VIII 2022), <https://nardus.mpn.gov.rs/handle/123456789/21293>
- докторске дисертације *Развој библиотечких и језичких ресурса за организовање и проналажење информација о просторном планирању* др Милене Милинковић (у својству члана комисије, датум одбране 27. IX 2022. године), <https://nardus.mpn.gov.rs/handle/123456789/21290>
- докторске дисертације *Композитне псеудограматике засноване на паралелним језичким моделима српског језика* др Михаила Шкорића (у својству члана комисије, датум одбране 6. VI 2023. године), <https://nardus.mpn.gov.rs/handle/123456789/21587>

## II Библиографија. Подаци о објављеним радовима

### (1) Библиографија радова до претходног избора у звање (2003–2019)

#### Научни радови (31):

#### Рад у врхунском међународном часопису (1)

1. Krstev, Cvetana, Obradović, Ivan, Utvić, Miloš and Vitas, Duško. 2014. A System for Named Entity Recognition Based on Local Grammars. *Journal of Logic and Computation* 24 (2): 473–489. <https://doi.org/10.1093/logcom/exs079> (ISSN: 0955-792X).
2. Tomašević, Aleksandra, Ranka Stanković, Miloš Utvić, Ivan Obradović, and Božo Kolonja. 2018. "Managing Mining Project Documentation Using Human Language Technology." *The Electronic Library* 36: 993–1009. <https://doi.org/10.1108/EL-11-2017-0239>. (ISSN: 0264-0473)

#### Рад у часопису међународног значаја верификованог посебном одлуком (1)

3. Драгићевић, Рајна М., and Милош В. Утвић. 2019. "Умножавање мовираних фемининума на -(к)иња у савременом српском језику." *Српски језик : студије српске и словенске* 24: 187–200. <https://doi.org/10.18485/sj.2019.24.1.9>. (ISSN: 0354-9259)

### Рад у водећем часопису националног значаја (3)

4. Утвић, Милош. 2014. "Листе учестаности Корпуса савременог српског језика." Александар Милановић, Живојин Станојчић and Љубомир Поповић (ур.). *Научни састанак слависта у Вукове дане. Српски језик и његови ресурси: теорија, опис и примене. 3/43. научни састанак слависта у Вукове дане, Београд, 12–15. IX 2013.* Београд: Међународни славистички центар, Филолошки факултет. 241–262. [http://informatika.fil.bg.ac.rs/misko/bibliografija/2014/Utvic-MSK\\_43-3.pdf](http://informatika.fil.bg.ac.rs/misko/bibliografija/2014/Utvic-MSK_43-3.pdf). (ISSN: 0351-9066)
5. Драгићевић, Рајна М., and Милош В. Утвић. 2017. "Корпусно истраживање ширења компоненте „мулти“ у српском језику данас." Рајна Драгићевић and Александар Милановић (ур.). *Научни састанак слависта у Вукове дане. Српски језик и његови ресурси: теорија, опис и примене. 3/46. научни састанак слависта у Вукове дане, Београд, 15–18. IX 2016.* Београд: Међународни славистички центар, Филолошки факултет. 73–86. <https://doi.org/10.18485/msc.2017.46.3.ch5>. (ISSN: 0351-9066)
6. Утвић, Милош В., Иван М. Обрадовић, Ранка М. Станковић, Александра Ђ. Томашевић, and Биљана Ђ. Лазић. 2018. "Изградња специјалних корпуса савременог српског језика на примеру корпуса из области рударства." Божо Ћорић and Александар Милановић (ур.). *Научни састанак слависта у Вукове дане. Српски језик и његови ресурси: теорија, опис и примене. 3/47. научни састанак слависта у Вукове дане, Београд, 2017.* Београд: Међународни славистички центар, Филолошки факултет. 103–118. <https://doi.org/10.18485/msc.2018.47.3.ch7>. (ISSN: 0351-9066)

### Рад у националном часопису (1)

7. Утвић, Милош. 2011. Анотација Корпуса савременог српског језика. *Инфотека XII* (2): 39–51. (ISSN: 1450-9687). [http://infoteka.bg.ac.rs/pdf/Srp/2011-2/INFOTHECA\\_XII\\_2\\_Decembar\\_39-51.pdf](http://infoteka.bg.ac.rs/pdf/Srp/2011-2/INFOTHECA_XII_2_Decembar_39-51.pdf)
8. Крстев, Цветана, Милош Утвић, and Јелена Јаћимовић. 2015. "Ако коза лаже, рог не лаже — где су и ко су жене у српској дневној штампи." *Књижевство — часопис за студије књижевности, рода и културе* 5. <https://doi.org/10.18485/knjiz.2015.1.24>. (eISSN: 2217-7809)

## Саопштење са међународног скупа штампано у целини (9)

9. Utvić, Miloš. 2006. Regular Derivation of Name Entities in Serbian. In *Proceedings of the 5th International Conference of Formal Approaches to South Slavic and Balkan Languages, FASSBL-5*, ed. S. Koeva and M. Dimitrova-Vulchanova, 257–264. Sofia: The Department of Computational Linguistics, Institute of Bulgarian Language. (264 p.)
10. Krstev, Cvetana, Stanković, Ranka, Obradović, Ivan, Vitas, Duško and Utvić, Miloš. 2010. Automatic Construction of a Morphological Dictionary of Multi-Word Units. In *Advances in Natural Language Processing, 7th International Conference on NLP, IceTAL 2010, Reykjavik, Iceland, August 16-18, 2010*, ed. H. Loftsson, E. Rúnvaldsson and S. Helgadóttir, 226–237. Vol. 6233. Berlin: Springer. (ISBN: 978-3-642-14769-2, 448 p.) [https://link.springer.com/chapter/10.1007/978-3-642-14770-8\\_26](https://link.springer.com/chapter/10.1007/978-3-642-14770-8_26) (ISBN: 978-3-642-14769-2, 448 p.)
11. Utvić, Miloš. 2010. The Regular Derivation in Serbian: Principles and Classification Using NooJ. In *Applications of Finite-State Language Processing: Selected Papers from the 2008 International NooJ Conference*, ed. T. Váradi, J. Kuti and M. Silberztein, 142–151. Cambridge: Cambridge Scholars Publishing. <https://www.cambridgescholars.com/resources/pdfs/978-1-4438-2573-3-sample.pdf> (ISBN: 978-1-4438-2573-3, 1-4438-2573-5, 225 p.)
12. Utvić, Miloš, Obradović, Ivan, Krstev, Cvetana and Vitas, Duško. 2010. The Effects of Multi-Word Tagging on Text Disambiguation. In *Proceedings of the 29th International Conference on Lexis and Grammar, September 15-18, 2010*, ed. D. Vitas and C. Krstev, 333–342. Belgrade: Faculty of Mathematics, University of Belgrade, Serbia. <http://informatika.fil.bg.ac.rs/misko/bibliografija/2010/LGC2010-MUIOCVDV.pdf> (ISBN: 978-86-7589-080-5, 356 p.)
13. Krstev, Cvetana, Vitas, Duško, Obradović, Ivan and Utvić, Miloš. 2011. E-Dictionaries and Finite-State Automata for the Recognition of Named Entities. In *Proceedings of the 9th International Workshop on Finite State Methods and Natural Language Processing*, ed. A. Maletti and M. Constant, 48–56. Blois, France: Association for Computational Linguistics. <http://informatika.fil.bg.ac.rs/misko/bibliografija/2011/Blois-W11-4407.pdf> (ISBN: 978-1-62276-464-8, 154 p.)
14. Yang, Ying, Utvić, Miloš, Pavlović-Lažetić, Gordana and Vitas, Duško. 2011. Named Entities in Chinese. In *Proceedings of the NooJ 2010 International Conference and Workshop, May 27-29, 2010, Komotini, Greece*, ed. Z. Gavriilidou, E. Chadjipapa, L. Papadopoulou and M. Silberztein, 112–124. Democritus University of Thrace. (281 p.) <http://informatika.fil.bg.ac.rs/misko/bibliografija/2011/NooJ%202010%20-%20Name%20Entities%20in%20Chinese.pdf>

15. Stanković, Ranka, Krstev, Cvetana, Obradović, Ivan, Trtovac, Aleksandra and Utvić, Miloš. 2012. A Tool for Enhanced Search of Multilingual Digital Libraries of E-journals. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, 23–25 May 2012*, ed. N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk and S. Piperidis, 1710–1717. [http://www.lrec-conf.org/proceedings/lrec2012/pdf/375\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/375_Paper.pdf) (ISBN: 978-2-9517408-7-7, 4148 p.)
16. Stanković, Ranka, Utvić, Miloš, Vitas, Duško, Krstev, Cvetana and Obradović, Ivan. 2012. On The Compatibility of Lexical Resources for NooJ. In *Automatic Processing of Various Levels of Linguistic Phenomena: Selected Papers from the 2011 International Nooj Conference*, ed. K. Vučković, B. Bekavac and M. Silberstein, 96–108. Cambridge: Cambridge Scholars Publishing. <https://www.cambridgescholars.com/resources/pdfs/978-1-4438-3711-8-sample.pdf> (ISBN: 978-1-4438-3711-8, 280 p.)
17. Zečević, Anđelka and Utvić, Miloš. 2012. An Authorship Attribution for Serbian. In *Local Proceedings of BCI'12, September 16–20, 2012, Novi Sad, Serbia*, ed. Z. Budimac, M. Ivanović and M. Radovanović, 109–112. Vol. 920. Novi Sad: Faculty of Sciences, University of Novi Sad. <http://ceur-ws.org/Vol-920/p109-zecevic.pdf> (ISBN: 978-86-7031-200-5, 150 p.)
18. Krstev, Cvetana, Duško Vitas, Miloš Utvić, and Branislava Šandrih. 2017. "The New Clothes for an Old Cookbook." Edited by Zygmunt Vetulani and Patrick Paroubek. *Human Language Technologies as a Challenge for Computer Science and Linguistics : 8th Language & Technology Conference*. Poznań, Poland: Fundacja University im. Adama Mickiewicza w Poznaniu. 59–63. <http://poincare.matf.bg.ac.rs/cvetana/biblio/ltc-042-krstev.pdf> (ISBN: 978-83-64864-94-0)
19. Radojičić, Marija, Ivan Obradović, Ranka Stanković, Miloš Utvić, and Sebastijan Kaplar. 2018. "A Mathematical Learning Environment Based on Serbian Language Resources." Edited by Ivan Milićević. *Proceedings / TIE 2018 : 7th International Scientific Conference Technics and Informatics in Education*. Čačak: University of Kragujevac, Faculty of Technical Sciences. 248–254. <http://www.ftn.kg.ac.rs/konferencije/tie2018/Radovi%20TIE%202018/EN/Proceedings%20TIE%202018.pdf> (ISBN: 978-86-7776-226-1)

#### **Рад у тематском зборнику међународног значаја (1)**

20. Obradović, Ivan, Stanković, Ranka and Utvić, Miloš. 2008. Integrisano okruženje za pripremu



paralelizovanog korpusa. In *Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen*, ed. B. Tošović, 563–578. Muenster: LitVerlag.

[http://informatika.fil.bg.ac.rs/misko/bibliografija/2007/IO\\_RS\\_MU\\_GRAZ\\_2007a.pdf](http://informatika.fil.bg.ac.rs/misko/bibliografija/2007/IO_RS_MU_GRAZ_2007a.pdf) (ISBN: 978-3-8258-0144-1, 846 p.)

21. Stanković, Ranka, Ivan Obradović, and Miloš Utvić. 2014. "Developing Termbases for Expert Terminology under the TBX Standard." Edited by Gordana Pavlović-Lažetić, Cvetana Krstev, Ivan Obradović and Duško Vitas. *Natural Language Processing for Serbian: Resources and Applications*. Belgrade: University of Belgrade, Faculty of Mathematics. 12–26. <https://jerteh.rs/wp-content/uploads/2015/05/Stankovic.pdf> (ISBN: 978-86-7589-088-1)
22. Utvić, Miloš. 2014. "Corpus n-Grams and Electronic Dictionaries of Serbian." Edited by Gordana Pavlović-Lažetić, Cvetana Krstev, Ivan Obradović and Duško Vitas. *Natural Language Processing for Serbian: Resources and Applications*. Belgrade: University of Belgrade, Faculty of Mathematics. 113–123. <https://jerteh.rs/wp-content/uploads/2015/05/Utvic.pdf> (ISBN: 978-86-7589-088-1)
23. Драгићевић, Рајна, and Милош Утвић. 2016. "Препозитивне компоненте у српском и руском језику на материјалу интернет-сајтова." Branko Tošović / Arno Wonisch (Hrsg.). *Wortbildung und Internet*. Graz: Institut für Slawistik der Karl-Franzens-Universität Graz – Kommission für Wortbildung beim Internationalen Slawistenkomitee. 91–106. (ISBN: 978-3-9504084-6-1)
24. Утвић, Милош, and Михаило Шкорић. 2017. "Репозиторијум дигиталних идентификатора објеката – доиФил: изградња, стање и перспективе." Александра Вранеш and Љиљана Марковић (ур.). *Научна конференција Библиоинфо — 55 година од покретања наставе библиотекарства на високошколском нивоу, Београд 18. мај 2017. : зборник радова*. Београд: Филолошки факултет Универзитета у Београду. 75–90. <https://doi.org/10.18485/biblioinfo.2017.ch7> (ISBN: 978-86-6153-589-5)

#### **Саопштење са међународног скупа штампано у изводу (1)**

25. Utvić, Miloš. 2006. Derivation from Multi-word Proper Names Using NooJ. In *9th INTEX/NooJ Conference (Abstracts)*, ed. D. Vitas and M. Silberztein, 69–70. Belgrade, Serbia: Faculty of Mathematics, University of Belgrade, LASELDI, Université de Franche-Comté, Maison des Sciences de l'Homme Claude Ledoux. (ISBN: 86-7589-055-9, 84 p.)

#### **Одбрањен магистарски рад**

26. Utvić, Miloš. 2008. Konačni automati u regularnoj imenskoj derivaciji. Magistarski rad,

Математички факултет, Универзитет у Београду. (105 стр.)  
<http://informatika.fil.bg.ac.rs/misko/bibliografija/2008/msMilosUtvic.pdf>

### **Одбрањена докторска дисертација**

27. Utvić, Miloš. 2014. Izgradnja referentnog korpusa savremenog srpskog jezika. Doktorska disertacija, Filološki fakultet Univerziteta u Beogradu. (396 str.)  
<http://informatika.fil.bg.ac.rs/misko/bibliografija/2014/Utvic-PhD.pdf>

### **Стручни радови:**

#### **Рад у тематском зборнику националног значаја (1)**

28. Утвић, Милош. 2008. Настава природно-математичке групе предмета у школској библиотеци. У *Школски библиотекар сарадник у настави*, ур. А. Вранеш, Љ. Марковић и П. Станојевић, 185–190. Београд: Филолошки факултет Универзитета у Београду, Библиотекарско друштво Србије. (ISBN: 978-86-86419-55-2, 208 стр.)

#### **Рад у часопису националног значаја (1)**

29. Утвић, Милош. 2010. BibTeX - креирање персоналних библиографија. *Панчевачко читалиште: научни часопис за теорију и праксу библиотекарства* IX (17): 35–38. (ISSN: 1451-3048) <http://www.citaliste.com/casopis/br17/9.pdf>

### **Остало:**

30. Утвић, Милош. 2003. Регуларни изрази. *Свет речи* VII (15–16): 66–68. (ISSN: 1450-6211)
31. Утвић, Милош. 2007. Корпус савременог српског језика. *Свет речи* XI (23-24): 34-36. (ISSN: 1450-6211)

## **(2) Библиографија радова од претходног избора у звање (2019–2024)**

### **Научни радови (8):**

#### **Рад у међународном часопису изузетних вредности (1)**

1. Škorić, Mihailo, Miloš Utvić, и Ranka Stanković. 2023. „Transformer-Based Composite Language Models for Text Evaluation and Classification“. *Mathematics* 11 (22). <https://doi.org/10.3390/math11224660>  
<http://informatika.fil.bg.ac.rs/misko/bibliografija/2023/mathematics-11-04660.pdf>

Системи за обраду природних језика, засновани на паралелној архитектури и прилагођени за српски, претходно су успешно тестирани на задацима аутоматског обележавања врстом речи (енгл. part-of-speech tagging) и идентификације аутора задатог документа (енгл. authorship attribution) применом минималног језичког моделирања, при чему су постигнути значајно бољи резултати у поређењу са просечним унапређењима која се добијају за седам европских језика на истом експерименту. Циљ овог рада је да представи предности употребе композитних језичких модела у обради и евалуацији (вредновању) писаних текстова на језицима богате флективне морфологије, посебно на српском. Скуп података заснован на перплексији, као главној методолошкој оцени, креиран је помоћу низа генеративних трансформера претходно обучених на различитим репрезентацијама корпуса на српском језику, те њиховом применом над три припремљена скупа реченица (експертски преводи, искварени преводи и аутоматски преводи). Рад описује упоредну анализу израчунатих перплексија како би се измерила класификациона способност различитих језичких модела на задацима бинарне класификације. Током експеримента тестирана су три самостална и два композитна језичка модела, заснована на излазним перплексијама самосталних језичких модела. Добијени резултати издвајају, као најбољу архитектуру композитних језичких модела за класификацију у случају поменутих задатака (обележавање врстом речи, идентификација ауторства), сложени наслагани (енгл. stacked) класификатор са вишеструким бројем својстава екстрахованим из вектора перплексије. Осим тога, резултати намећу корисност употребе синтаксичких и семантичких језичких модела, који успешно препознају различите специфичне карактеристике текста који се обрађује, као и корисност употребе вектора перплексије као адекватне алтернативе перплексији при обучавању композитних језичких модела ове врсте.

#### **Рад у међународном часопису (1)**

2. Драгићевић, Рајна М., и Милош В. Утвић. 2022. „Лексиколошки допринос граматички границе Предрага Пипера“. *Јужнословенски филолог* 78 (2): 131–46. <https://doi.org/10.2298/JFI2202131D>  
<https://dais.sanu.ac.rs/bitstream/handle/123456789/13890/dragic.utvic.piper.2022.pdf>

Разматрајући језичка средства за изражавање границе у српском језику, проф. П. Пипер је навео неколико типичних именица које представљају лексичка средства у том категоријално-семантичком комплексу. У овом раду истражују се и остале именице које имају ту улогу у српском језику. Полази се од претпоставке да улогу граничника у српском језику вероватно врше оне именице које се чешће од свих осталих употребљавају са предлозима, тј. којима је реч иза које се користе најчешће предлог. На основу електронског корпуса СрпКор2013 направљен је списак свих таквих именица, а затим су направљене и листе именица које се најчешће користе директно иза предлога који сугеришу неки тип граничности, а то су: *до, од, из, на, око, пред, у*. Извршена је, затим, селекција грађе на основу семантичких и прагматичких критеријума и тако је добијена листа именица-граничника које су описане у овом раду.

### **Рад у врхунском часопису националног значаја (3)**

3. Утвић, Милош В., Ранка М. Станковић, Александра Ђ. Томашевић, Михаило Ђ. Шкорић, и Биљана Ђ. Лазић. 2019. „Претрага корпуса заснована на употреби екстерних лексичких ресурса путем веб-сервиса“. У *Научни састанак слависта у Вукове дане. Српски језик и његови ресурси: теорија, опис и примене. 3/48. научни састанак слависта у Вукове дане, Београд, 2018.*, ур. Божо Ђорић и Александар Милановић, 48 (3): 279–98. Београд: Међународни славистички центар, Филолошки факултет. <https://doi.org/10.18485/msc.2019.48.3.ch12>  
[http://doi.fil.bg.ac.rs/pdf/eb\\_ser/msc/2019-3/msc-2019-48-3-ch12.pdf](http://doi.fil.bg.ac.rs/pdf/eb_ser/msc/2019-3/msc-2019-48-3-ch12.pdf)

У раду се разматра хибридни приступ претрази корпуса, илустрован на примеру алатки OCWB и NoSketch Engine, примењених на специјални корпус из области рударства (РудКор) и Корпус савременог српског језика (СрпКор). Разматрани приступ комбинује постојеће могућности алатки OCWB и NoSketch Engine, које своју претрагу заснивају на лингвистичкој анотацији корпуса, са новим могућностима претраге у виду консултовања екстерних језичких ресурса (морфолошки електронски речници српског језика и лексичка база података Српски ворднет). Хибридни приступ је реализован надоградњом веб-сучеља која поменуте алатке користе за претрагу корпуса, као и употребом веб-сервиса заснованих на технологији RESTful. Веб-сервиси омогућавају да се задати упит, у коме се појављује лексема X, по потреби прошири лексемама које се са лексемом X налазе у некој лексичкој релацији (синонимија, антонимија, хиперонимија итд.).

4. Драгићевић, Рајна М., и Милош В. Утвић. 2020. „Песнички речник Ирине Грицкат“. У *Научни састанак слависта у Вукове дане. Српски језик и његови ресурси: теорија, опис и примене. 3/49. научни састанак слависта у Вукове дане, Београд, 2019.*, ур. Божо Ђорић и Александар

Милановић, 49(3): 55–84. Београд: Међународни славистички центар, Филолошки факултет.  
<https://doi.org/10.18485/msc.2020.49.3.ch4>  
[http://doi.fil.bg.ac.rs/pdf/eb\\_ser/msc/2020-3/msc-2020-49-3-ch4.pdf](http://doi.fil.bg.ac.rs/pdf/eb_ser/msc/2020-3/msc-2020-49-3-ch4.pdf)

Рад се бави избором лексике Ирене Грицкат у збирци песама *Циклоида* и указује на везу између њене поезије, посебно лексике у њој, са закључцима у њеним лингвистичким истраживањима и размишљањима у есејима које је посветила науци. У раду се објављује и фреквенцијски речник аутосемантичних лексема употребљених у овој збирци, чиме се та лексика даје на увид читаоцима и омогућава лакши приступ за даљу анализу.

5. Ђорђевић, Ана, и Милош Утвић. 2022. „Квалитативни библиометријски индикатори у научним радовима депонованим у институционални репозиторијум Хемијског факултета“. *Читалиште : научни часопис за теорију и праксу библиотекарства* 2022 (41): 30–44.  
<https://doi.org/10.19090/cit.2022.41.30-44>  
<https://citaliste.ff.uns.ac.rs/index.php/cit/article/view/167/115>  
<http://informatika.fil.bg.ac.rs/misko/bibliografija/2022/Citaliste-2022-DjordjevicUtvic.pdf>

Репозиторијум Хемијског факултета Универзитета у Београду и Иновационог центра *Cherry* представља добро организовану дигиталну библиотеку научних резултата који се у оквиру тих институција реализују. За потребе овог рада истражена је промена броја цитата научних радова садржаних у репозиторијуму на основу екстерне апликације *Аутори, пројекти, публикације* у којој су интегрисани параметри за преглед броја цитата из индексних база *Scopus* и *Web of Science*. У раду се такође посматра и промена броја цитата код радова из репозиторијума који се појављују на платформи за алтернативну метрику *Altmetric*. Представљена су два метода истраживања, у првом се разматра промена броја цитата за све унете и цитиране записе радова у репозиторијуму, док се у оквиру другог промена броја цитата проучава на нивоу свих записа радова одређеног аутора. Онај чији су радови изабрани за други метод овог истраживања је у извештају Хемијског факултета за 2021. годину проглашен најцитиранијим аутором. Код оба метода се посматрају три врсте записа са различитим степеном доступности: трајно закључани, доступни након ембарго периода (зелени отворени приступ) и они који су потпуно у отвореном приступу (златни отворени приступ). Метаподаци свих врста записа радова у репозиторијуму *Cherry* су јавно доступни уз строгу контролу библиотекарца – администратора репозиторијума. Овим истраживањем се запажа разлика која настаје у броју цитата појединачних радова у септембру 2021. године и годину дана касније, у септембру 2022. Циљ рада је да се представи начин на који би могла да се успостави корелација између нивоа доступности радова и њихове цитираности, са посебним освртом на радове у зеленом отвореном приступу, што је полазна тачка за посматрање сврсисходности репозиторијума како за Хемијски факултет и Иновациони центар, тако и за друге институције које су део развијене мреже репозиторијума у Србији.

**Рад у националном часопису (1)**

6. Stanković, Ranka, and Miloš Utvić. 2019. „Vebran Web Services for Corpus Query Expansion“. *Infotheca* 19(2): 99–118. <https://doi.org/10.18485/infotheca.2019.19.2.5>  
<http://informatika.fil.bg.ac.rs/misko/bibliografija/2019/infotheca-2019-19-2-5.pdf>

Овај рад разматра даљи развој веб сервиса Вебран и њихову примену на унапређење претраге корпуса. Веб сервис Вебран су у претходном истраживању већ оспособљени да, користећи спољашње лексичке ресурсе за српски језик (углавном електронске морфолошке речнике за српски и Српски ворднет), семантички проширују корисничке упите за претрагу корпуса. Међутим, неизбежне су грешке произашле из аутоматске анотације корпуса врстом речи и лемом применом програма претходно обученог на далеко мањем, ручно аотираном корпусу. Као последица, кориснички упит који тражи флективну парадигму задате леме не мора пронаћи све релевантне резултате, а због претходне хомографски узроковане погрешне аутоматске лематизације може произвести и нерелевантне резултате. Рад описује веб сервисе за морфолошко проширење упита приликом претраге корпуса, тј. генерисање флективне парадигме задате леме како би се из корпуса српског језика добили релевантнији резултати. У случајевима када корисник задаје лему присутну у лексичким ресурсима које користе веб сервисе, одзив је 100%, тј. добијају се сви релевантни резултати, а евентуално и неки нерелевантни јер на непотпуну прецизност поново утиче хомографија (различите леме могу имати хомографски исте облике). Посебну пажњу рад посвећује генерисању флективних парадигми полилексичких јединица (енгл. multi word units) за шта је такође развијен посебан веб сервис.

#### **Саопштење са међународног скупа штампано у целини (2)**

7. Stanković, Ranka, Branislava Šandrih, Cvetana Krstev, Miloš Utvić, and Mihailo Škorić. 2020. „Machine Learning and Deep Neural Network-Based Lemmatization and Morphosyntactic Tagging for Serbian“. In *Proceedings of the Twelfth International Conference on Language Resources and Evaluation, LREC 2020, Marseille, France, May 11–16, 2020*, eds. Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi et al, 3947–55. <http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.487.pdf>

Рад описује потребу за обучавањем нових језичких модела за обележавање српских текстова врстом речи (енгл. tagger models), превасходно мотивисану проширивањем постојећег скупа етикета или обележја (енгл. tagset) граматичком категоријом рода. У припремној фази су усклађивани ручно аотирани ресурси, добијени као резултат различитих пројеката током дужег временског периода, што је омогућено развојем алатки које подржавају делимичну аутоматизацију. Иако су разматрани различити програми за аутоматско обележавање (енгл. taggers), као и скупови етикета, рад се фокусира на програме *TreeTagger* и *spaCy* и на усклађивање шеме анотације примењене у електронским морфолошким речницима српског са шемама *MULTEXT-East* и *Universal Part-of-Speech tagset*. Обучени језички модели користиће се за

креирање нове верзије Корпуса савременог српског језика и Корпуса српске књижевности. Упоредна евалуација програма за аутоматско обележавање, уз истраживање како на перформансе утиче избор одговарајућих скупова етикета, утврдила је за оба нова модела прецизност обележавања по токену приближно једнаку 98%. Рад такође најављује и нови анотирани скуп података SR BASIC (око 200 хиљада анотираних токена) који ће моћи да се користи за обуку нових језичких модела.

8. Stanković, Ranka, Christian Chiarcos, Miloš Utvić, and Olivera Kitanović. 2023. „Towards ELTeC-LLOD: European Literary Text Collection Linguistic Linked Open Data“. У *Proceedings of the 4th Conference on Language, Data and Knowledge (LDK 2023) - Vienna, Austria, 12–15 September 2023.*, eds. Sara Carvalho, Anas Fahad Khan, Ana Ostroški Anić, Blerina Spahiu, Jorge Gracia, John P. McCrae, Dagmar Gromann, Barbara Heinisch, and Ana Salgado, 180–91. Lisboa: NOVA FCSH - CLUNL. <https://doi.org/10.34619/srmk-inji>  
<http://informatika.fil.bg.ac.rs/misko/bibliografija/2023/LDK2003-Stankovicetal.pdf>

Рад описује студију случаја којом се анализира генерисање текстуалних корпуса повезаних података (енгл. Linked Data) применом формата NLP Interchange Format (NIF). NIF је формат заснован на језицима за репрезентацију знања (RDF/OWL) који претендује да постигне међусобну размену информација и њихово коришћење између различитих алатки за аутоматску обраду природних језика, језичких ресурса и анотација. Иако рад укратко разматра сличне формате који користе RDF за потребе лингвистичке анотације, посебна пажња је посвећена формату NIF зато што, за разлику од неких поменутих формата, поседује експлицитне структуре података за одвојену (енгл. stand-off) анотацију (на пример, именованих ентитета, морфосинтаксичких описа, резултате синтаксичке анализе граматиком зависности, анализу сентимента текста итд.), има довољно обиман RDF речник и у широкој је употреби. Подскуп корпуса ELTEC, састављеног од 900 романа на 9 европских језика, насталих током периода 1840-1920, послужио је као основа за истраживање. Рад описује трансформацију анотираних верзија поменутих романа из полазног формата TEI level-2 у формат NIF, отворени програмски код који имплементира одговарајуће кораке трансформације (енгл. pipeline), као и резултате за сличне случајеве примене.

#### **IV Оцена о наставном ангажовању и педагошком раду**

Др Милош Утвић има 24 године педагошког искуства, врло кратко на Математичком факултету, а потом у континуитету на Катедри за библиотекарство и информатику Филолошког факултета УБ. У звањима сарадника, асистента и доцента до сада успешно је изводио све облике наставе на курсевима који су му поверени. У протеклом изборном периоду (2019–2024) био је члан три комисије за одбрану мастер радова и учествовао у раду пет комисија за преглед, оцену и одбрану докторске дисертације на Универзитету у Београду.

Увидом у анкете о вредновању педагошког рада наставника, које се спроводе на Филолошком факултету, истичу се највише оцене за редовно одржавање предавања и вежби, разумљивост и начин излагања материје, помоћ студентима да савладају градиво, спремност да се

одговори на коментаре студената, пружање корисних информација за њихов будући рад, објективност и непристрасност у оцени знања студената, професионалност и етичност наставника у укупној комуникацији са студентима.

## V ЗАКЉУЧАК И ПРЕДЛОГ КОМИСИЈЕ

На основу пажљивог увида у биографију, библиографију и радове кандидата, Комисија је закључила да др Милош Утвић испуњава све услове за реизбор у звање доцента. Има одличне референце, објављује радове у релевантним публикацијама, бави се актуелним темама, пре свега у областима библиотечке информатике и рачунарске лингвистике, показује истрајност и преданост у научном и педагошком раду. Комисија је мишљења да су испуњени формални, законски, статутарни и суштински услови за поновни избор др Милоша Утвића у звање доцента.

Имајући у виду успешну реализацију свих облика наставне активности и вредност научних радова др Милоша Утвића, Комисија са задовољством предлаже Изборном већу Филолошког факултета Универзитета у Београду да прихвати овај извештај те да др Милоша Утвића поново изабере у звање и на радно место доцента за ужу научну област Библиотечка информатика, предмет *Дигитални текст*.

Београд, 22. септембар 2024.

### КОМИСИЈА

др Гордана Стокић Симончић, редовни професор

др Гордана Рудић, редовни професор

др Ранка Станковић, ванредни професор