

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Извештај комисије за оцену испуњености услова за избор др **Вука Батановића** у научно звање **научни сарадник**

Одлуком Наставно-научног већа Електротехничког факултета Универзитета у Београду бр. 1686/2, од 03.02.2021. године, образована је Комисија за оцену испуњености услова за избор др Вука Батановића у научно звање научни сарадник, у саставу:

1. др Бошко Николић, редовни професор, Универзитет у Београду – Електротехнички факултет
2. др Милош Цветановић, ванредни професор, Универзитет у Београду – Електротехнички факултет
3. др Зоран Шеварац, ванредни професор, Универзитет у Београду – Факултет организационих наука

По пријему релевантне документације, Комисија је обавила анализу научне и стручне активности кандидата, на основу чега подноси следећи

ИЗВЕШТАЈ

1. Биографски подаци о кандидату

Вук Батановић је рођен 02.06.1987. у Београду. Основне академске студије на Електротехничком факултету Универзитета у Београду је уписао 2006. године. Дипломирао је на модулу Рачунарска техника и информатика 2010. године, са просечном оценом 9,56. Мастер академске студије на истом модулу Електротехничког факултета је уписао 2010. и завршио 2011. године, са просечном оценом 10. У пролећном семестру школске 2011/2012. године уписао је докторске академске студије на модулу Софтверско инжењерство истог факултета. Испите на докторским студијама је положио са просечном оценом 10, а докторску дисертацију под насловом „Методологија решавања семантичких проблема у обради кратких текстова написаних на природним језицима са ограниченим ресурсима“ одбранио је 2020. године. У току студија усавршавао се присуствовањем већем броју летњих школа из обраде природних језика и машинског учења.

Од 2018. године ради као истраживач-сарадник у Иновационом центру Електротехничког факултета у Београду. Од школске 2017/2018. године ангажован је у настави на мастер академским студијама Електротехничког факултета, на предметима Обрада природних језика и Проналажење скривеног знања. Сарађивао је на међународном истраживачком пројекту *Regional Linguistic Data Initiative* (ReLDI) и словеначком пројекту CLARIN.SI (*Common Language Resources and Technology Infrastructure*), као и на иновационом пројекту Програма Уједињених нација за развој (UNDP) о аутоматизацији семантичке претраге правних

документа на српском језику. Од септембра 2020. године ангажован је на пројекту AVANTES (*Advancing Novel Textual Similarity-based Solutions in Software Development*), одобреног у оквиру програма Фонда за науку Републике Србије за развој пројеката из вештачке интелигенције. Добитник је награде „Блажо Мирчевски“ за најбољи рад младог аутора презентован на конференцији ТЕЛФОР 2016.

Научна интересовања су му усмерена на решавање семантичких проблема из обраде природних језика, пре свега применом метода машинског учења. Нарочито га интересују специфичности обраде кратких текстова и решења примењива и у језицима са ограниченим ресурсима.

2. Библиографија

Кандидат Вук Батановић бира се у звање научни сарадник по први пут, тако да се вреднују сви његови досадашњи научни резултати.

Кандидат је приложио електронске верзије свих својих радова. Комисија је утврдила аутентичност научних радова објављених у научним часописима категорије М20 провером на сајту издавача и преко приложених DOI референци. Поред тога, Комисија је утврдила аутентичност научних резултата категорије М30, М50 и М60 провером на интернету. За техничка решења Комисија је утврдила аутентичност користећи податке из РИС система Министарства просвете, науке и технолошког развоја.

У наставку су дате табеле са објављеним научним резултатима кандидата, разврстаним по категоријама. Број поена по категоријама дефинисан је према Правилнику о стицању истраживачких и научних звања („Службени гласник РС“, бр. 159/2020). За сваки рад укључен у SCOPUS индексну базу наведен је и број хетероцитата.

Сходно Прилогу 2 наведеног Правилника, за одређивање коефицијента M и импакт фактора међународних часописа разматран је период од две године пре публиковања и година публиковања, и усвајана она година у којој је часопис најбоље рангиран, односно у којој је имао највећи импакт фактор. На основу одељка 1.4 Прилога 1 наведеног Правилника, за научне резултате са више од три коаутора коришћена је формула за ефективне поене $K/(1+0,2(n-3))$, $n>3$, где је n број аутора, а K номинални број поена у датој категорији.

P. бр.	Подаци о раду	Цит.	Катег. рез.	Поени	Ефект. поени
Радови објављени у научним часописима међународног значаја М20					
1	B. Furlan, V. Batanović, B. Nikolić, “Semantic similarity of short texts in languages with a deficient natural language processing support”, <i>Decision Support Systems</i> , vol. 55, no. 3, pp. 710–719, 2013. ISSN: 0167-9236, DOI: 10.1016/j.dss.2013.02.002 https://doi.org/10.1016/j.dss.2013.02.002	25	M21a IF (2012) 2,201 rang: 6/79	10	10

2	<p>V.Batanović, M. Cvetanović, B. Nikolić, "A versatile framework for resource-limited sentiment articulation, annotation, and analysis of short texts", <i>PLoS ONE</i>, vol. 15, no. 11, Article ID: e0242050, 2020. ISSN: 1932-6203, DOI: 10.1371/journal.pone.0242050</p> <p>https://doi.org/10.1371/journal.pone.0242050</p>		M22 IF (2018) 2,776 rang: 24/69	5	5
3	<p>V. Batanović, D. Bojić, "Using Part-of-Speech Tags as Deep-Syntax Indicators in Determining Short-Text Semantic Similarity", <i>Computer Science and Information Systems</i>, vol. 12, no. 1, pp. 1–31, 2015. ISSN: 1820-0214 (Print), 2406-1018 (Online), DOI: 10.2298/CSIS131127082B</p> <p>https://doi.org/10.2298/CSIS131127082B</p>	8	M23 IF (2015) 0,623 rang: 83/106	3	3

Зборници међународних научних скупова М30

4	<p>V. Batanović, B. Nikolić, "Using Language Technologies to Automate the UNDP Rapid Integrated Assessment Mechanism in Serbian", in <i>Proceedings of the International Conference on Language Technologies for All: Enabling Linguistic Diversity and Multilingualism Worldwide (LT4All)</i>, Paris, France, 2019.</p> <p>https://lt4all.elra.info/media/papers/O5/137.pdf</p>		M33	1	1
5	<p>V. Batanović, N. Ljubešić, T. Samardžić, "SETimes.SR – A Reference Training Corpus of Serbian", in <i>Proceedings of the Conference on Language Technologies & Digital Humanities 2018 (JT-DH 2018)</i>, pp. 11–17, Ljubljana, Slovenia, 2018.</p> <p>http://www.sdjt.si/wp/wp-content/uploads/2018/09/JTDH-2018_Batanovic-et-al_SETimesSR-A-Reference-Training-Corpus-of-Serbian.pdf</p>		M33	1	1
6	<p>N. Ljubešić, Ž. Agić, F. Klubička, V. Batanović, T. Erjavec, "hr500k – A Reference Training Corpus of Croatian", in <i>Proceedings of the Conference on Language Technologies & Digital Humanities 2018 (JT-DH 2018)</i>, pp. 154–161, Ljubljana, Slovenia, 2018.</p> <p>http://www.sdjt.si/wp/wp-content/uploads/2018/09/JTDH-2018_Ljubesic-et-al_hr500k-A-Reference-Training-Corpus-of-Croatian.pdf</p>		M33	1	0,71

7	<p>V. Batanović, M. Cvetanović, B. Nikolić, "Fine-grained Semantic Textual Similarity for Serbian", in <i>Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)</i>, pp. 1370–1378, Miyazaki, Japan, 2018.</p> <p>http://www.lrec-conf.org/proceedings/lrec2018/pdf/442.pdf</p>	1	M33	1	1
8	<p>V. Batanović, B. Nikolić, "Sentiment Classification of Documents in Serbian: The Effects of Morphological Normalization", in <i>Proceedings of the 24th Telecommunications Forum (TELFOR 2016)</i>, Belgrade, Serbia, 2016. ISBN: 978-1-5090-4087-2 (Print), 978-1-5090-4086-5 (Online), DOI: 10.1109/TELFOR.2016.7818923</p> <p>https://doi.org/10.1109/TELFOR.2016.7818923</p>	2	M33	1	1
9	<p>V. Batanović, B. Nikolić, M. Milosavljević, "Reliable Baselines for Sentiment Analysis in Resource-Limited Languages: The Serbian Movie Review Dataset", in <i>Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)</i>, pp. 2688–2696, Portorož, Slovenia, 2016.</p> <p>http://www.lrec-conf.org/proceedings/lrec2016/pdf/284_Paper.pdf</p>	5	M33	1	1
10	<p>N. Ljubešić, T. Samardžić, T. Erjavec, D. Fišer, M. Miličević Petrović, S. Krek, V. Batanović, "The "ReLDI effect": Collaborative development of manually annotated datasets for Slovene, Croatian and Serbian", in <i>Proceedings of the International conference E-dictionaries and E-lexicography (eReL 2019)</i>, pp. 95–96, 2019.</p> <p>http://ihjj.hr/mreznik/uploads/7b5602698002896a0fd4f2dce210796.pdf</p>		M34	0,5	0,28
Радови у часописима националног значаја М50					
11	<p>V. Batanović, B. Nikolić, "Sentiment Classification of Documents in Serbian: The Effects of Morphological Normalization and Word Embeddings", <i>Telfor Journal</i>, vol. 9, no. 2, pp. 104–109, 2017. ISSN: 1821-3251 (Print), 2334-9905 (Online), DOI: 10.5937/telfor1702104B</p> <p>https://doi.org/10.5937/telfor1702104B</p>	1	M52	1,5	1,5

12	V. Batanović , D. Bojić, "Evaluation and Classification of Syntax Usage in Determining Short-Text Semantic Similarity", <i>Telfor Journal</i> , vol. 6, no. 1, pp. 64–68, 2014. ISSN: 1821-3251 (Print), 2334-9905 (Online), DOI: 10.5937/telfor1401064B https://doi.org/10.5937/telfor1401064B	M52	1,5	1,5
----	--	-----	-----	-----

Зборници научних скупова националног значаја М60

13	V. Batanović , N. Ljubešić, T. Samardžić, M. Miličević Petrović, „Otvoreni resursi i tehnologije za obradu srpskog jezika“, <i>Zbornik radova sa konferencije Primena slobodnog softvera i otvorenog hardvera 2020 (PSSOH 2020)</i> , Beograd, Srbija, 2020. DOI: 10.5281/zenodo.4113230 https://doi.org/10.5281/zenodo.4113230	M61	1,5	1,25
14	V. Batanović , D. Bojić, „Evaluacija i klasifikacija korišćenja sintaksnih informacija u određivanju semantičke sličnosti kratkih tekstova“, <i>Zbornik radova sa 21. Telekomunikacionog foruma (TELFOR 2013)</i> , pp. 821–824, Beograd, Srbija, 2013. ISBN: 978-1-4799-1419-7 (Print), 978-1-4799-1420-3 (Online), DOI: 10.1109/TELFOR.2013.6716356. https://doi.org/10.1109/TELFOR.2013.6716356	M63	0,5	0,5
15	V. Batanović , B. Furlan, B. Nikolić, „Softverski sistem za određivanje semantičke sličnosti kratkih tekstova na srpskom jeziku“, <i>Zbornik radova sa 19. Telekomunikacionog foruma (TELFOR 2011)</i> , pp. 1249–1252, Beograd, Srbija, 2011. ISBN: 978-1-4577-1499-3 (Print), 978-1-4577-1500-6 (Online) DOI: 10.1109/TELFOR.2011.6143778 https://doi.org/10.1109/TELFOR.2011.6143778	M63	0,5	0,5
16	D. Drašković, V. Batanović , B. Nikolić, „Softverski sistem za učenje ekspertskega sistema“, <i>Zbornik radova sa 18. telekomunikacionog foruma (TELFOR 2010)</i> , pp. 1129–1132, Beograd, Srbija, 2010. http://2010.telfor.rs/files/radovi/TELFOR2010_10_17.pdf	M63	0,5	0,5

Одбрањена докторска дисертација М70

17	V Batanović, „Metodologija rešavanja semantičkih problema u obradi kratkih tekstova napisanih na prirodnim jezicima sa ograničenim resursima“, doktorska disertacija, Univerzitet u Beogradu – Elektrotehnički fakultet, 2020. https://www.etf.bg.ac.rs/uploads/files/javni_uvid/izvestaji/doktorske/2020/10/Vuk_Batanovic_Doktorska_dizertacija.pdf	M70	6	6
Техничка решења M80				
18	V. Batanović, B. Furlan, B. Nikolić, „Softverski sistem za određivanje semantičke sličnosti kratkog teksta napisanog na srpskom jeziku“, 2013.	M81	8	8
19	V. Batanović, B. Furlan, B. Nikolić, „Korpus parafraza srpskog jezika“, 2013.	M86	1	1

3. Анализа научних резултата

Кандидат Вук Батановић се у свом научном раду бави облашћу обраде природних језика, са фокусом на решавање семантичких проблема у аутоматској анализи текстуалних података написаних на језицима са ограниченим ресурсима. У оквиру тога, кандидат је највећу пажњу посветио проблему анализе сентимента текстова и проблему одређивања семантичке сличности кратких текстова. Кандидат је у свом научном раду такође дао и значајан допринос развоју слободно доступних језичких алата и анотираних корпуса текстова на српском језику, развијених како за поменуте проблеме, тако и за друге, морфосинтактичке, синтактичке и семантичке проблеме у обради српског језика.

Проблем анализе сентимента текстова кандидат је разматрао у референцама под редним бројем 2, 8, 9, 11, 13 и 17. У раду бр. 9 кандидат је представио први јавно доступан скуп докумената на српском језику који су аутоматски анотирани у погледу поларности сентимента који изражавају, при чему су коришћене три класе – позитивна, негативна и неутрална. Поред тога, представљена је и оригинална метода за оптимизацију процеса балансирања скупова података из овог домена погодна за језике са ограниченим ресурсима. Евалуацију различитих статистичких модела заснованих на машинском учењу на креираном скупу података кандидат је започео у раду бр. 9 и наставио у радовима бр. 8 и 11. Поред тога, кандидат је израдио и први јавно доступан скуп кратких текстова на српском језику који су ручно анотирани у погледу сентимента, за шта је коришћен нов система означавања који је кандидат представио. Овај систем се разликује од претходних по знатној флексибилности тумачења ознака сентимента, тако да се једном додељене ознаке могу касније користити за већи број ужих проблема у анализи сентимента, укључујући одређивање поларности, одређивање субјективности, детекцију сарказма, вишекласну класификацију сентимента, итд. Наведени систем означавања и скуп података су приказани у раду бр. 2, у коме је такође предложена и нова метрика за утврђивање економичности анотације података, као и евалуација већег броја статистичких модела заснованих на машинском учењу. Сви наведени резултати су систематизовани у прегледном раду под редним бројем 13, као и у докторској дисертацији кандидата (р. бр. 17).

Проблем аутоматског одређивања семантичке сличности кратких текстова кандидат је разматрао како на текстовима на енглеском језику, тако и на онима на српском. Евалуација постојећих приступа за решавање једноставнијег проблема детекције парофраза спроведена је у радовима бр. 12 и 14, док је нова метода за одређивање семантичке сличности представљена и евалуирана на енглеском језику у раду бр. 3. У референцама бр. 1, 15 и 19 приказан је први скуп података за проблем детекције парофраза на српском језику, који је касније проширен грануларнијим оценама семантичке сличности и описан у раду бр. 7. У референцама бр. 1, 7 и 18 кандидат је представио и нове моделе за одређивање семантичке сличности кратких текстова, који су погодни за примену у језицима са ограниченим ресурсима, попут српског. Наведени резултати су систематизовани у прегледном раду под редним бројем 13, као и у докторској дисертацији кандидата (р. бр. 17). Поред тога, у раду бр. 4 кандидат је размотрio и примену техника за одређивање семантичке сличности текстова на практичном проблему семантичке претраге правних докумената на српском језику.

Кандидат је такође детаљно разматрао утицај различитих техника морфолошке нормализације текстова у контексту оба наведена семантичка проблема, и развио је обједињени пакет алгоритама за стемовање текстова на српском језику, представљен у раду бр. 9. Поред тога, учествовао је и у развоју референтних анотираних корпуса на српском и хрватском језику за проблеме токенизације и раздвајања текстова на реченице, морфосинтактичког означавања речи, лематизације, синтактичког парсирања реченица и означавања именованих ентитета. Креирани корпуси представљају прве јавно доступне корпусе тог типа и они су описани у радовима број 5, 6, 10 и 13.

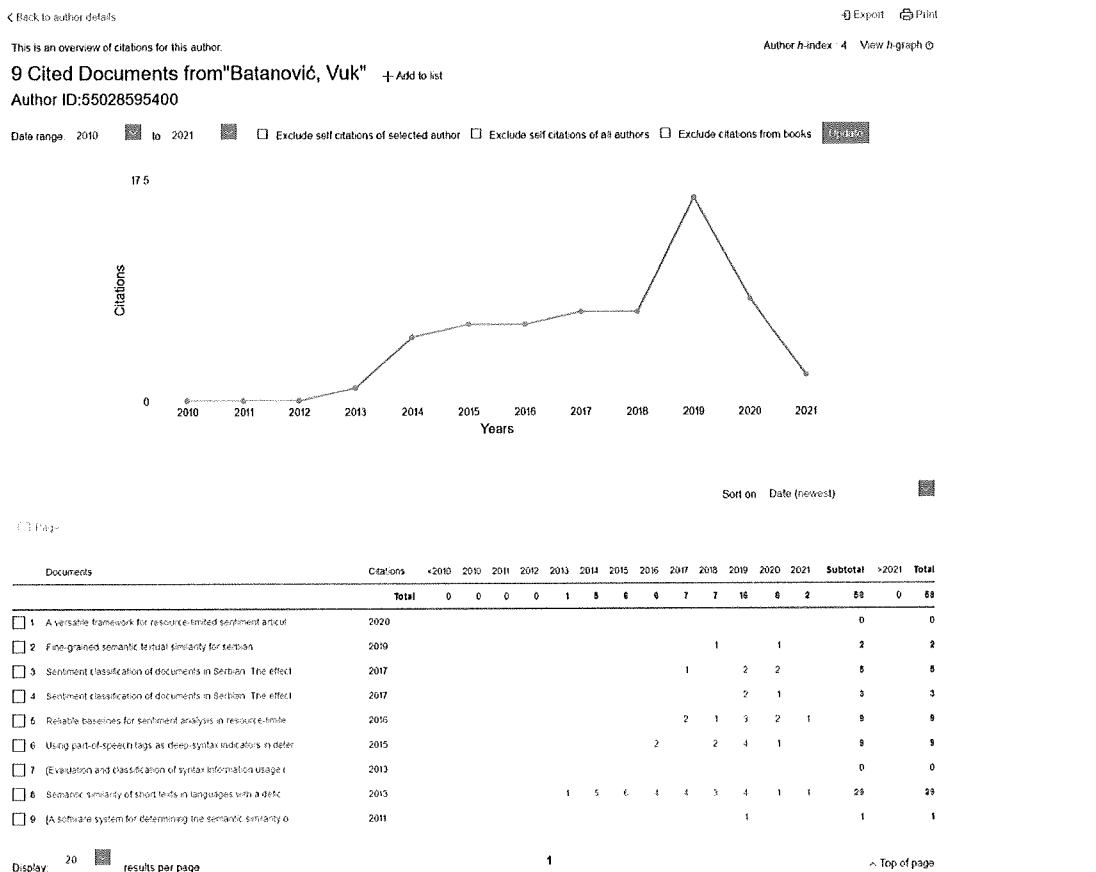
4. Цитираност објављених радова кандидата

У SCOPUS бази индексирано је девет радова кандидата Вука Батановића, који имају укупно 58 цитата од чега 42 хетероцитата. Кандидат у овој бази има Хиршов индекс са вредношћу 4, односно вредношћу 3 ако се посматрају само хетероцитати. Приложен је дијаграм и списак цитираности радова кандидата у SCOPUS индексној бази.

Најцитијанији рад кандидата је референца под редним бројем 1, који има 29 цитата, односно 25 хетероцитата. Значајнији број цитата имају и референце под редним бројем 3 (9 цитата, од чега 8 хетероцитата) и редним бројем 9 (9 цитата, од чега 5 хетероцитата). Радови кандидата су позитивно цитирани у престижним међународним часописима категорија M21a и M21, попут рада бр 1, који је цитиран у часописима *IEEE Access*, *Information Processing and Management*, *Decision Support Systems* и *Scientometrics*, или рада бр. 8, који је цитиран у часопису *Organizational Research Methods*.

У бази Google Scholar радови кандидата имају укупно 131 цитат, док кандидат има Хиршов индекс са вредношћу 6.

Citation overview



Цитираност радова кандидата индексираних у SCOPUS бази

5. Квалитативни показатељи научног ангажмана

Кандидат Вук Батановић се континуирано бави научним радом и објавио је 19 научних публикација, од чега 3 рада у међународним часописима (M20), 2 рада у часописима националног значаја (M50), 7 радова на међународним (M30) и 4 на домаћим конференцијама (M60), докторску дисертацију (M70) и 2 техничка решења (M80). Главни фокус његових истраживања је област обраде природних језика, у коју спада 18 од 19 његових објављених научних резултата. При томе, кандидат је показао висок степен самосталности у научном раду, те је први аутор на 15 од 19 научних резултата.

Кандидат је добитник награде „Блажко Мирчевски“ за најбољи рад младог аутора презентован на конференцији ТЕЛФОР 2016 (референца са редним бројем 8). У приложеној документацији кандидат је навео да је рецензент часописа *Decision Support Systems* (категорија M21a), часописа *IEEE Access* (категорија M21), часописа *Journal of Experimental & Theoretical Artificial Intelligence* (категорија M22), часописа *Computer Science and Information Systems* (категорија M23), као и да је рецензирао радове за конференције *LREC 2020 – 12th International Conference on Language Resources and Evaluation*, *COLING 2018 – 27th International Conference on Computational Linguistics*, *BSNLP 2017 – 6th Workshop on Balto-Slavic Natural Language Processing*, као и неколико *TELFOR* конференција.

Своје научноистраживачке активности кандидат је остварио кроз већи број домаћих и међународних научних и иновационих пројеката. У току 2018. године кандидат је био

укључен у иновациони пројекат Програма Уједињених нација за развој (UNDP) о аутоматизацији семантичке претраге правних докумената на српском језику, који је финансиран кроз позив за иновационе пројекте канцеларије Уједињених нација за координацију развојних активности (UNDOCO). Од септембра 2020. године кандидат је ангажован као руководилац неколико радних пакета на пројекту *Advancing Novel Textual Similarity-based Solutions in Software Development* (AVANTES), одобреног од стране Фонда за науку Републике Србије у оквиру Програма за развој пројеката из вештачке интелигенције (грант бр. 6526093).

Кандидат је био сарадник у реализацији међународног истраживачког пројеката *Regional Linguistic Data Initiative* (ReLDI, грант бр. 160501), који је представљао партнерство између научних организација у Швајцарској, Србији и Хрватској у оквиру програма SCOPES Швајцарске националне фондације за науку (2015 – 2018). Током овог пројекта, као и кроз сарадњу са Институтом Јожеф Стефан из Љубљане у оквиру иницијативе CLARIN.SI (*Common Language Resources and Technology Infrastructure*), кандидат је допринео креирању или унапређењу већег броја језичких алата и анотираних корпуса за аутоматску обраду јужнословенских језика. Поред тога, кандидат учествује у изградњи јавно доступних корпуса новинских текстова и коментара на тему језика за потребе пројекта *(Re-)imaging language, nation and collective identity in the 21st century: Language ideologies as new connections in post-Yugoslav digital mediascapes* који се изводи у оквиру Научно-истраживачког центра Словеначке академије наука и уметности (ZRC SAZU). Овај двогодишњи пројекат (2019–2021) је финансиран од стране Словеначке агенције за истраживачку делатност (ARSS) у оквиру програма *MSCA Horizon 2020 Seal of Excellence Complementary Scheme N6-0123*.

6. Квантитативна оцена научних резултата кандидата

Према Правилнику о стицању истраживачких и научних звања („Службени гласник РС“, бр. 159/2020), кандидат је остварио укупно 44,74 поена (неопходно 16), за категорије M10+M20+M31+M32+M33+M41+M42+M51+M80+M90+M100 остварио је 32,71 поен (неопходно 9), а у категорији радова M21+M22+M23 остварио је 18 поена (неопходно је 5).

На основу наведених података, закључено је да у свакој од три категорије резултата кандидат има значајно више од неопходног броја поена за избор у звање научни сарадник.

Табела за оцену испуњености услова за избор у звање научни сарадник за техничко-технолошке и биотехничке науке

Потребно је да кандидат има најмање XX поена, који треба да припадају следећим категоријама:	Неопходно	Остварено
Укупно \geq	16	44,74
M10 + M20 + M31 + M32 + M33 + M41 + M42 + M51 + M80 + M90 + M100 \geq	9	32,71
M21 + M22 + M23 \geq	5	18

7. Закључак и предлог Комисије

Научноистраживачки рад кандидата Вука Батановића припада области обраде природних језика. Кандидат се највише бавио истраживањима у домену семантичке обраде текстова на природним језицима, са посебним фокусом на кратке текстове и на природне језике са ограниченим ресурсима, што је била и тема његове докторске дисертације. У склопу свог

рада, кандидат је развио већи број језичких алата и ресурса за српски језик у виду анотиралих корпуса текстова неопходних за обучавање и евалуацију различитих статистичких модела заснованих на алгоритмима машинског учења. Научни доприноси кандидата у том погледу нису ограничени само на семантичке проблеме, јер је учествовао у креирању или унапређењу алата и ресурса и за друге типове задатака у обради природних језика.

Значај кандидатовог научног рада је верификован кроз 19 научних публикација, са 42 хетероцитата у индексној бази SCOPUS, као и кроз учешће у већем броју научноистраживачких и иновационих пројекта. При томе, кандидат је остварио значајно већи број поена у свакој од категорија резултата од неопходног броја поена за избор у звање научни сарадник.

Комисија закључује да кандидат задовољава и квантитативне и квалитативне услове за стицање научног звања научни сарадник за техничко-технолошке науке, те стога предлаже Наставно-научном већу Електротехничког факултета Универзитета у Београду избор Вука Батановића у звање научни сарадник.

у Београду,
04.02.2021. године

ЧЛАНОВИ КОМИСИЈЕ

dr Boško Nikolić, redovni profesor
Универзитет у Београду – Електротехнички факултет

dr Miloš Ćvetanović, vanredni profesor
Универзитет у Београду – Електротехнички факултет

dr Zoran Ševarač, vanredni profesor
Универзитет у Београду – Факултет организационих наука