

**ВЕЋУ ЗА СТУДИЈЕ ПРИ УНИВЕРЗИТЕТУ**  
**УНИВЕРЗИТЕТА У БЕОГРАДУ**

**ПРИЈАВА ТЕМЕ ЗА ИЗРАДУ МАСТЕР РАДА**

Обраћам се Већу за мултидисциплинарне студије Универзитета у Београду са молбом да ми се, као студенту мастер академских студија:

**РАЧУНАРСТВО У ДРУШТВЕНИМ НАУКАМА**

одобри израда мастер рада.

**Наслов тезе:**

**Системи за одговарање на питања у јавним позивима за финансирање научноистраживачких пројеката: текстометријска и NLP анализа институционалног дискурса**

**Научни циљеви истраживања:**

У савременим системима научног финансирања, дигитални канали комуникације између институција и истраживача имају важну улогу у јавним позивима и конкурсима за научноистраживачке пројекте. Посебан облик ове комуникације јесу системи за одговарање на питања (у литератури наведени као Q&A системи), у којима се одговори институције на питања апликаната јавно објављују ради појашњења услова конкурса, правила програма и процедура пријављивања.

Иако превасходно служе функционалној комуникацији, ови системи истовремено представљају значајан корпус који би омогућио контрастивну анализу два језичка варијетета: формализоване одговоре институције и питања апликаната која су често мање формална и ближа природној језичкој употреби. Овај контраст је од значаја за разумевање начина на који се информације језички обликују у зависности од улоге учесника у комуникацији, као и за утврђивање разлика које утичу на њихову приступачност и разумљивост.

На основу добијених резултата, истраживање тежи ка формулисању језичких ресурса који могу послужити као основа за развој алата за аутоматизовано одговарање и унапређење комуникације у институционалном контексту.

### **Сврха истраживања:**

Сврха истраживања је да се кроз корпусну и текстометријску анализу испитају језичке разлике између институционалних одговора и питања апликаната у оквиру система за одговарање на питања у јавним позивима за финансирање научноистраживачких пројеката. Истраживање има за циљ да допринесе бољем разумевању начина на који се институционална комуникација језички обликује, као и на који начин различите комуникативне улоге утичу на степен формалности, лексичку и синтаксичку сложеност и разумљивост текста.

Формирање и анализа корпуса институционалне комуникације омогућавају извођење емпиријски заснованих закључака који могу унапредити разумевање институционалних текстова и процедура, и допринети развоју система за аутоматско одговарање на питања у овој области.

### **Предмет рада:**

Истраживање се налази на пресеку корпусне лингвистике, обраде природног језика и проучавања институционалне комуникације, уз примену рачунарских метода у анализи друштвених феномена. Предмет рада је корпус структурираних докумената у ком су обележена питања апликаната и институционалних одговора, прикупљених у оквиру јавних позива за финансирање научноистраживачких пројеката. Посебан фокус ставља се на језичке и дискурсне карактеристике ова два поткорпуса, као и на њихово поређење у оквиру истог институционалног комуникационог система.

Биће формиран специјализовани корпус структурираних докумената која садрже питања и одговоре (Q&A парове) објављене на званичном сајту институције током 9 јавних позива у оквиру различитих програма финансирања научноистраживачких пројеката Фонда за науку Републике Србије. Поред докумената који садрже одговоре на питања, корпус ће укључити и релевантну конкурсну документацију, упутства и друге

јавне документе који дефинишу правила програма и процедуре пријављивања пројеката. Ови документи чине додатни извор доменског знања и представљају основу за формирање базе знања у наредним фазама истраживања.

Употреба рачунарских метода за анализу дискурса омогућава анализу великих количина језичких података и доношење поузданијих закључака о лексичким, синтаксичким и стилским обележјима посматраног дискурса.

Формулисана су следећа истраживачка питања:

1. Које су основне језичке карактеристике и разлике између питања корисника и институционалних одговора у корпусу јавних позива за финансирање научноистраживачких пројеката?
2. У којој мери се институционални одговори одликују већом лексичком и синтаксичком сложености у односу на питања корисника?
3. Да ли се и на који начин разлике између два поткорпуса могу повезати са различитим комуникативним улогама учесника у институционалној комуникацији?
4. У којој мери уочене језичке разлике утичу на разумљивост и приступачност јавних информација?

Полазећи од литературе из области социолингвистике, стилистике и анализе институционалног дискурса, претпоставка је да се институционални одговори одликују вишим степеном формалности, већом лексичком и синтаксичком комплексношћу, чешћом употребом пасивних конструкција, клишеа и фраза, односно да су одговорима више заступљени обрасци административног и бирократског језика у односу на језик корисничких питања.

#### **Методе рада:**

Предложени материјал представљао би доменски специфичан корпус институционалне комуникације у области финансирања научноистраживачких пројеката и пружио би основу за контрастивну анализу језика апликаната и институционалних одговора. У истраживању ће се примењивати методе корпусне лингвистике, текстометрије и обраде природног језика (NLP). Корпус ће се анализирати помоћу квантитативних метода које обухватају фреквенцијску анализу, анализу лексичке разноврсности,

колокација, испитивање дужине и структуре реченица, као и поређење морфосинтаксичких категорија. За обраду и анализу података користиће се напредне функционалности програма Microsoft Excel, софтверски алат за текстометријску анализу ТХМ и програмски језик Python. За аутоматско одређивање лема и врста речи користиће се алат TreeTagger, интегрисан у софтвер ТХМ. За означавање текстова на српском језику примењује се TreeTagger модел SrpKor4Tagging, који омогућава морфосинтаксичко означавање корпуса и представља основу за даљу текстометријску и лингвистичку анализу.

За прву фазу истраживања неопходно је прикупити и хармонизовати податке, што укључује њихову интеграцију, трансформацију и хармонизацију, ради даље рачунарске обраде. Формираће се специјализовани корпус подељен у два поткорпуса који садрже питања апликаната и одговоре институције током јавних позива за финансирање научноистраживачких пројеката, и треба га припремити за даљу обраду коришћењем ТХМ алата. Помоћни, референтни корпус чиниће релевантна конкурсна нормативна документација, упутства за припрему и подношење предлога пројеката, текстови јавних позива и други документи којима се дефинишу правила програма и процедуре пријављивања пројеката.

Биће коришћена фреквенцијска анализа ради утврђивања најчешћих лексичких јединица и образаца у корпусу. Применом метода текстометрије анализираће се кључне речи, конкорданце и колокације, са циљем да идентификујемо типичне језичке обрасце, лексику и фразе које се јављају у институционалној комуникацији. Даља анализа обухватиће испитивање лексичке разноврсности, комплексности у погледу дужине речи и реченица, пасивних конструкција, и других одлика административног и бирократског стила. Биће примењене методе обраде природног језика ради аутоматског издвајања, мерења и поређења језичких карактеристика различитих делова корпуса. На основу добијених резултата биће спроведена упоредна анализа питања корисника и институционалних одговора, са циљем утврђивања њихових лингвистичких сличности и разлика.

Добијени резултати ће такође послужити за свеобухватно сагледавање језичких одлика институционалне комуникације, као и за разматрање могућности примене формираног корпуса у будућим истраживањима и системима заснованим на обради природног језика.

### **Структура (по поглављима):**

#### **1. Увод**

- 1.1. Предмет истраживања
- 1.2. Циљеви и истраживачка питања
- 1.3. Теоријски оквир и контекст институционалне комуникације

#### **2. Методологија**

- 2.1. Формирање и обрада корпуса
- 2.2. Опис и структура корпуса
- 2.3. Методе обраде природног језика (NLP)
- 2.4. Текстометријски приступ

#### **3. Анализа корпуса**

- 3.1. Квантитативна анализа
- 3.2. Језичке одлике институционалних одговора
- 3.3. Контрастивна анализа поткорпуса

#### **4. Дискусија**

- 4.1. Језичке разлике у корпусу
- 4.2. Лексичко поређење са општим и нормативним корпусом
- 4.3. Примена у обради природног језика (NLP)

#### **5. Закључак**

- 5.1. Научни допринос и даљи правци истраживања

#### **6. Литература**

### Оквирна литература:

1. Biber, D. (1988). *Variation across Speech and Writing*. Cambridge University Press.
2. Bugarski, R. (1996). *Jezik u društvu*. Beograd: Čigoja štampa.
3. Drew, P., & Heritage, J. (1992). *Talk at Work: Interaction in Institutional Settings*. Cambridge University Press.
4. Fairclough, N. (1989). *Language and Power*. London: Longman.
5. Halliday, M. A. K., & Matthiessen, C. (2014). *Halliday's Introduction to Functional Grammar*. Routledge.
6. Heiden, S. (2010). The TXM platform: Building open-source textual analysis software compatible with the TEI encoding scheme. In *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation* (pp. 389–398).
7. Jaćimović, J. (2019). Textometric methods and the TXM platform for corpus analysis and visual presentation. *Infotheca – Journal for Digital Humanities*, 19(1), 30–54. <https://doi.org/10.18485/infotheca.2019.19.1.2>
8. Klikovac, D. (2008). *Jezik i moć: ogledi iz sociolingvistike i stilistike*.
9. McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
10. Radovanović, M. (2003). *Sociolingvistika*. Sremski Karlovci – Novi Sad: Izdavačka knjižarnica Zorana Stojanovića.
11. Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*. Manchester, UK.
12. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
13. Stanković, R., Krstev, C., & Vitas, D. (2024). SrpELTeC: A Serbian literary corpus for distant reading. *Primerjalna književnost*, 47(2). <https://doi.org/10.3986/pkn.v47.i2.03>
14. Stanković, R., Šandrih, B., Krstev, C., Utvić, M., & Škorić, M. (2020). Machine learning and deep neural network-based lemmatization and morphosyntactic tagging for Serbian. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 3954–3962). <https://aclanthology.org/2020.lrec-1.487>

15. Stanković, R., Škorić, M., & Šandrih Todorović, B. (2022). Parallel bidirectionally pretrained taggers as feature generators. *Applied Sciences*, 12(10), 5028. <https://doi.org/10.3390/app12105028>
16. Tošović, B. (2002). *Funkcionalni stilovi*. Beograd: Beogradska knjiga.

Интернет извор:

Фонд за науку Републике Србије. <https://fondzanauku.gov.rs/pozivi/>

Напомена: Списак литературе ће бити додатно прилагођен и допуњен приликом израде рада.

### **Радна биографија:**

Ђурђа Перовић је рођена у Ваљеву, 1994. године. Школовала се у Специјализовано-математичком одељењу Ваљевске гимназије, где се први пут сусрела са рачунарством и програмским језицима. Основне академске студије Српског језика и књижевности завршила је на Филолошком факултету Универзитета у Београду. Наставак школовања на мултидисциплинарним мастер академским студијама Рачунарство у друштвеним наукама при Универзитету у Београду представљао је природан наставак њених интересовања ка повезивању лингвистике и рачунарства. Све испите је успешно положила са просечном оценом 10,00.

Од 2021. године запослена је у Фонду за науку Републике Србије, где учествује у припреми, спровођењу и праћењу програма финансирања научноистраживачких пројеката. Током свог рада пружала је подршку истраживачима и научноистраживачким организацијама и учествовала у унапређењу институционалне комуникације, процедура и дигиталних алата за управљање програмима и пројектима.

Ментор

Кандидат

Ранка Станковић

Ђ. Перовић

проф. др Ранка Станковић, редовни професор  
Универзитет у Београду, Рударско-геолошки факултет

Ђурђа Перовић,  
бр. индекса: 33/2025

Будући да је кандидаткиња полагањем испита, обављањем стручне праске и одбраном приступног рада остварила 45 ЕСПБ, комисија предлаже да јој се одобри израда предложеног мастер рада.

**Чланови комисије:**

проф. др Владан Деведић, редовни професор, Универзитет у Београду, Факултет организационих наука,

др Јован Чудомировић, доцент, Универзитет у Београду, Филолошки факултет

## Изјава о академској честитости

Ђурђа Перовић 33/2025

Својим потписом гарантујем да ћу се у току студија, а нарочито у току израде мастер рада, у потпуности уздржати од плагирања, односно копирања туђих идеја и њиховог презентовања као сопствених, без одговарајућег признавања ауторства.

Разумем да копирање туђих идеја може бити извршено из књига и часописа и података који могу бити представљени у табелама, дијаграмима, дизајну, плановима, фотографијама, филму, музици, формулама, веб-сајтовима, компјутерским програмима и на други начин. Јасно ми је да плагијат такође укључује и представљање, употребу и дистрибуирање рада предавача или других студената као сопствених.

Свесна сам да намерним присвајањем идеја других и приказивањем као сопствених чиним повреде Закона о ауторском и сродним правима, као и других закона и одговарајућих аката Универзитета у Београду.

У Београду, 10. јун 2026.

Потпис студента Ђ. Перовић