

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

I ПОДАЦИ О КОМИСИЈИ
<ol style="list-style-type: none">1. Датум и орган који је именовано комисију 18. X 2021. Научно-наставно веће Филолошког факултета2. Састав комисије са знаком имена и презимена сваког члана, звања, назива уже научне области за коју је изабран у звање, датума избора у звање и назив факултета, установе у којој је члан комисије запослен:<ol style="list-style-type: none">1. др Цветана Крстев, редовни професор, библиотечка информатика, 20. V 2014, Филолошки факултет Универзитета у Београду2. др Бранислава Шандрих Тодоровић, доцент, библиотечка информатика, 26. I 2021, Филолошки факултет Универзитета у Београду3. др Балша Стипчевић, доцент, српски језик, 22. IX 2020, Филолошки факултет Универзитета у Београду4. др Ранка Станковић, ванредни професор, математика и информатика, 19.V 2015, Рударско-геолошки факултет Универзитета у Београду5. др Јелена Јаћимовић, научни сарадник, библиотечка информатика, 15. X 2020, Стоматолошки факултет Универзитета у Београду
II ПОДАЦИ О КАНДИДАТУ
<ol style="list-style-type: none">1. Име, име једног родитеља, презиме: Тања (Драгољуб) Ивановић2. Датум рођења, општина, република: 3. IX 1981. Београд, Србија3. Датум одбране, место и назив мастер рада: 27. II 2008. Београд. „Односи с јавношћу у афирмисању потенцијала обновљиве енергије у Србији“4. Научна област из које је стечено академско звање мастера: Организационе науке, менаџмент
III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ: Развој терминологије из области електроенергетике коришћењем метода обраде природних језика
IV ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ: Навести кратак садржај са знаком броја страна поглавља, слика, шема, графика и сл.
Докторска дисертација мастер Тање Ивановић бави се развојем паралелне терминологије из области електроенергетике применом савремених рачунарских метода заснованих на корпусима и језичким ресурсима. Истраживање посебно обухвата: <ul style="list-style-type: none">• преглед термилошких прилика у српском језику, посебно у области електроенергетике, те преглед лексикона и стандарда који донекле покривају термине из ове области;• основне појмове о рачунарским корпусима, једнојезичним и вишејезичним, начинима њихове изградње и могућностима њиховог коришћења, с посебним освртом на коришћење корпуса у термилошком раду;

- изградњу паралелизованог енглеско-српског корпуса из домена електроенергетике који у великој мери покрива правну регулативу ЕУ у овој области;
- екстракцију термина на српском и енглеском језику, као и њихово поравнавање применом информатичких метода, темељну евалуацију добијених резултата и њихово јавно објављивање за потребе истраживања и преводилачког рада.

У истраживању се користе експерименталне методе, као што су корпусне методе претраживања и методе вештачке интелигенције. За процену квалитета остварених резултата коришћена је евалуација од стране два стручњака.

Дисертација обухвата 170 стране, а у оквиру тога 8 поглавља (149 стране), списак коришћене литературе (16 страна, 139 библиографских јединица на латиници, 31 на ћирилици, 19 речника, 94 стандарда), додатак (коришћене скраћенице, 5 страна), уводни и завршни материјал (насловне стране на српском, енглеском и руском, апстракт на српском и енглеском, садржај, списак табела, списак слика, типографске конвенције, биографија кандидата и три прилога, 17 страна). У дисертацији укупно има 32 слике и 24 табеле. Поглавља дисертације су:

1. Уводна разматрања (2 стране).
2. Терминологија (24 стране).
3. Језичке и термилошке прилике у српском језику (22 стране).
4. Формирање и развој терминологије из области електроенергетике у српском језику (32 стране).
5. Примена корпусне лингвистике и улога корпуса у термилошком истраживању (28 страна);
6. Корпусни алати и језички ресурси у обради природних језика (8 страна);
7. Обрада паралелизованог корпуса електроенергетског система (30 страна),
8. Закључна разматрања (3 стране).

Прилози дисертације су:

9. Допатак (5 страна).

V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

У уводном поглављу дисертације кандидаткиња Тања Ивановић представља предмет својих истраживања, даје њихов кратак преглед по поглављима и укратко приказује остварене резултате.

У другом поглављу – „Терминологија“ – кандидаткиња се пре свега бави самим термином *терминологија*, који у зависности од контекста може имати различито значење. Кандидаткиња указује како су овом питању приступали разни истраживачи, чија би се гледишта могла сажети у закључак да појам *терминологија* може означавати: (а) научну дисциплину; (б) скуп термина који представљају систем појмова одређеног предметног поља; (в) публикацију у којој су ти термини сакупљени. Кандидаткиња према овоме одређује и свој правац рада, који представља скуп активности на прикупљању, описивању и презентацији термина изабраног предметног поља (електроенергетика). У наставку овог поглавља кандидаткиња представља рад и достигнућа традиционалних термилошких школа: *бечке*, у којој преовлађује приступ предметног поља, *руске*, у којој преовлађује филозофски приступ, и *прашке*, у којој преовлађује лингвистички приступ. Кандидаткиња се осврће и на критике ових термилошких школа до којих је дошло услед убрзаног развоја науке и технике. Но и поред критика, многе поставке традиционалних термилошких школа се не одбацују већ се модификују или допуњују. Оно у чему се бављење терминологијом данас битно разликује од времена настанка традиционалних термилошких школа јесте укључивање савремених информатичких средстава у развој терминологије, а ту се пре свега мисли на језичке, доменске корпусе и на алате за екстракцију информација (термина) из њих. На крају овог поглавља, кандидаткиња се, имајући у виду значај термина и термилошких система за трансфер знања, осврће на питања стандардизације терминологије, пре свега на рад Међународне организације за стандардизацију (ISO) и њеног Техничког комитета за језик и терминологију ISO/TC 37, као и других међународних и европских организација које се баве стандардизацијом

терминологије, те коначно и радом Института за стандардизацију Републике Србије у овој области.

У трећем поглављу – „Језичке и термилошке прилике у српском језику“ – кандидаткиња Тања Ивановић даје кратак приказ историјских прилика које су условљавале настанак терминологије у српском језику, указујући на допринос појединаца и институција које су препознале значај развоја терминологије за потребе школства, напредак науке и технолошког прогреса у Србији (и Југославији) у XIX и XX веку. Осврћући се на термилошке прилике у Србији данас, кандидаткиња указује на утицаје који на српски језик уопште, а терминологију посебно, долазе из страних језика, а данас пре свега из енглеског. Имајући у виду изузетно динамичан развој науке и технологије у протеклим деценијама, изузетно је важно да се неконтролисано преузимање енглеских термина ограничи, да би се не само избегле двосмислености и погрешна тумачења већ и да би се очувао српски језик у новом, дигиталном окружењу. Стога на развоју терминологије у српском језику треба радити поштујући неке унапред постављене критеријуме настале као резултат вишедеценијских научних истраживања. За постизање ових циљева задужено је више институција у Србији, чији рад кандидаткиња укратко представља. То су: Институт за српски језик САНУ, Одбор за стандардизацију српског језика, Министарство за европске интеграције Републике Србије, Институт за стандардизацију РС и Друштво за језичке ресурсе и технологије, које се посебно бави питањима обраде српског језика у дигиталном окружењу.

Четврто поглавље – „Формирање и развој терминологије из области електроенергетике у српском језику“ – кандидаткиња отпочиње представљањем електроенергетике као научне дисциплине и њене позиције у класификацији наука према Дјуијевој децималној класификацији и Универзалној децималној класификацији. Кандидаткиња потом даје кратак преглед веома динамичног развоја електроенергетике као науке и праксе у Србији. У већем делу овог поглавља кандидаткиња исцрпно представља једнојезичне, двојезичне и вишејезичне речнике из области електротехнике објављене у Србији (и Југославији) који су од значаја, или су били од значаја, за развој терминологије из области електроенергетике. Закључак овог истраживања је да су речници углавном застарели те да су готово сви посвећени ширим областима. У посебни део издвојени су речници из области електроенергетике које издаје Институт за стандардизацију Србије (ИСС), а који су махом настали превођењем одговарајућих постојећих међународних стандарда (речника). Препознајући значај рада ИСС у овој области, кандидаткиња указује и на извесне недоследности које су у овим речницима појављују. Ово поглавље кандидаткиња завршава представљањем значајних организација из области електроенергетике у Србији данас и улогом коју би оне могле имати за развоје терминологије из ове области.

У петом поглављу – „Примена корпусне лингвистике и улога корпуса у термилошком истраживању“ – кандидаткиња указује на значајну улогу коју језички корпуси данас имају за термилошка истраживања. После кратког прегледа историјата настанка и коришћења корпуса и представљања њихових најважнијих карактеристика, кандидаткиња говори о корпусима релевантним за изучавање српског језика, Корпусу савременог српског језика и вишејезичним корпусима у којима је један од језика српски. Посебну пажњу кандидаткиња посвећује двама језичким ресурсима: први од њих је Корпус правних текстова Европске уније (Евротека), који представља енглеско-српски корпус паралелних текстова који садржи делове правне регулативе ЕУ (такозвани *acquis communautaire*), док је други вишејезична термилошка база Евроним настала током превођења правне регулативе ЕУ на српски језик а која представља и важан алат у току самог превођења. Ова два ресурса су од значаја за кандидаткињино истраживање јер правна регулатива ЕУ делом покрива и електроенергетику. На крају овог поглавља је представљена дигитална библиотека паралелних текстова и алат за њихово коришћење *Библиша*, који је кандидаткиња и користила за представљање резултата свог истраживања

У шестом поглављу – „Корпусни алати и језички ресурси у обради природних језика“ – које је логични наставак претходног поглавља, кандидаткиња укратко представља неке рачунарске ресурсе и алате за обраду корпуса које је неопходно користити да би се они могли успешно применити у термилошком раду. Овде се, поред општих алата за претраживање

корпуса, мисли на алате који врше сегментацију полазног текста на токене и реченице, обележавање врстом речи и придруживање леме, а потом и моделирање неких уобичајених структура именичних израза. Да би се ови задаци могли успешно обавити, потребни су електронски речници језика који се обрађују, те кандидаткиња укратко представља такве речнике за српски језик који су јој били на располагању. За кандидаткињин даљи рад, који је описан у наредном поглављу, посебно су значајни алати који од два текста на различитим језицима, од којих је један превод другог, формирају текст паралелизован на нивоу сегмента (најчешће реченице). Кандидаткиња представља неке од ових алата које је она користила у свом раду.

У седмом поглављу – „Обрада паралелизованог корпуса електроенергетског система“ – кандидаткиња Тања Ивановић описује свој рад на припреми доменског двојезичног корпуса, екстракцији двојезичне терминологије, евалуацији и објављивању резултата. После кратког прегледа литературе везане за екстракцију терминологије из паралелизованих корпуса и различитих метода које су у њој помињу, кандидаткиња представља метод који је она користила, а који се заснива на постојању паралелизованог корпуса, листа термина за сваки појединачни језик које су добијене било из постојећих речника или екстракцијом из дела истог корпуса на одговарајућем језику, и система за аутоматско поравнавање сегмената паралелног текста који би могли бити преводни еквиваленти. Кандидаткиња потом детаљно описује садржај корпуса из области електроенергетике који је формирала а који се састоји од 76 различитих докумената, који имају скоро 30.000 сегмената, више од 610.000 речи у енглеском делу и скоро 550.000 у српском делу корпуса. После претварања првобитног материјала у паралелизовани текст приступило се његовој припремној обради, што је подразумевало обележавање врстом речи и придруживање лема. У наредној фази, екстраховани су потенцијални термини на енглеском и српском језику из одговарајућих делова корпуса, за шта су коришћени различити алати. За српски језик коришћене су локалне граматике засноване на електронским речницима српског језика које из текста издвајају именичке фразе које по структури представљају потенцијалне термине – на пример, именица којој претходи придев. Након поравнавања сегмената који потенцијално представљају преводне еквиваленте и њиховог сравњивања са екстрахованим листама термина, добијена је листа од 3.833 потенцијална енглеско-српска термилошка пара.

Евалуацију добијених резултата обавила су два независна евалуатора, кандидаткиња као стручни преводилац у области електроенергетике и један стручњак из ове области. При оцењивању резултата екстракције, евалуатори су се придржавали критеријума за одабир добрих термина које је кандидаткиња поставила, а који су у складу са теоријом и праксом у овој области. Евалуатори су имали три задатка евалуације: екстракција енглеских термина, екстракција српских термина, и упаривање енглеских и српских термина. Поступак евалуације је добро постављен, а евалуатори су добро обавили свој посао, јер су постигли висок степен слагања, што је утврђено високим Коеновим степеном сагласности: 0.8614 за први задатак, 0.8197 за други задатак и 0.9032 за трећи задатак. После усаглашавања евалуатора, коначан резултат овог рада је листа од 2.684 енглеско-српска термилошка пара из области електроенергетике или неке шире техничке области. Добијене резултате кандидаткиња је поредила са две јавно доступне термилошке базе: у бази *Евроним* потврђена су 52 термилошка пара из добијене листе, док је у бази *Електронепедија* потврђено 25 парова из разних предметних области. Премда ови резултати потврђују правилност поступка изградње двојезичне терминологије, они још више показују да на оваквом приступу треба и даље радити да би се ови важни ресурси допунили терминима на српском језику. Добијени резултати су објављени и доступни за коришћење широј стручној јавности. Термини из области електроенергетике могу се претраживати у бази *Терми*, која подржава развој терминологије и већ садржи термине из разних научних области. Такође, паралелни корпус који је кандидаткиња припремила могу консултовати одобрени корисници дигиталне библиотеке *Библиша* и при томе могу своје упите формулисати коришћењем термина из базе *Терми*.

У завршном поглављу – „Закључна разматрања“ – кандидаткиња сумаризује резултате постигнуте овом дисертацијом и даје смернице за будући рад. Постигнути резултати и представљена метода која је до њих довела треба да буду окосница рада на будућем речнику

електроенергетике, који би требало да, осим у штампаном облику, буде доступан и онлајн, било као веб или мобилна апликација.

На крају дисертације мастер Тања Ивановић је приложила 2 додатка:

1. У додатку 1 дат је списак коришћених скраћеница на ћирилици.
2. У додатку 2 дат је списак коришћених скраћеница на латиници.

VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ, уз напомену: Навести називе радова, где и када су објављени.

Ivanović, Tanja, Ranka Stanković, Branislava Šandrih i Cvetana Krstev i. „Corpus-based bilingual terminology extraction in the power engineering domain“. *Terminology*, примљено за штампу 9. IX 2021.

Impact Factor: 0.826 (5-year: 1.208)

Arts & Humanities Citation Index

Ivanović, T. (2016). Lexical analysis of two-word terminological phrases within distribution system. *Infotheca - Journal For Digital Humanities*, 16(1-2). doi:10.18485/infotheca.2016.16.1_2.3

VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА

Резултати изложени у овој дисертацији говоре да је кандидаткиња мастер Тања Ивановић остварила циљеве зацртане у пријави дисертације. Кандидаткиња је прецизно описала стање терминологије из области електроенергетике у српском језику, указала на могућности које за убрзан развоје терминологије пружају савремене језичке технологије, посебно оне које се заснивају на корпусима и лексичким ресурсима, развила је један доменски корпус и применила савремене методе екстракције термина, чиме је добијен паралелан терминолошки ресурс из домена електроенергетике, који је објављен и корисницима доступан онлајн.

Сам текст дисертације, као и списак литературе наведен на крају рада, говоре да је мастер Тања Ивановић користила релевантну и савремену литературу, те да је постављене проблеме обрадила детаљно и сагледавајући их из разних углова. Овим радом мастер Тања Ивановић је изградила један значајан нови терминолошки ресурс за српски (и енглески) језик из области електроенергетике, који је постао доступан широком кругу истраживача, поставила је основе за његово унапређивање и развила методологију за изградњу сличних ресурса у будућности.

VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА **НАПОМЕНА:** Навести позитивну или негативну оцену начина приказа и тумачења резултата истраживања.

Комисија сматра да је кандидаткиња Тања Ивановић у својој дисертацији *Развој терминологије из области електроенергетике коришћењем метода обраде природних језика* успешно обрадила значајну тему коришћењем савремених метода, да је текст дисертације урађен према одобреној пријави дисертације и да је реч о раду који представља оригинално и самостално научно дело.

IX ПРЕДЛОГ:

На основу укупне оцене дисертације, комисија предлаже Научно-наставном већу Филолошког факултета Универзитета у Београду да прихвати извештај о дисертацији *Развој терминологије из области електроенергетике коришћењем метода обраде природних језика* кандидаткиње мастер Тање Ивановић и упути га Већу научних области друштвено-хуманистичких наука Универзитета у Београду, како би кандидаткиња била позвана на усмену одбрану рада.

ПОТПИСИ ЧЛАНОВА КОМИСИЈЕ

1. др Цветана Крстев, редовни професор
Филолошки факултет, Универзитет у Београду
2. др Бранислава Шандрих Тодоровић, доцент
Филолошки факултет, Универзитет у Београду
3. др Балша Стипчевић, доцент
Филолошки факултет, Универзитет у Београду
4. др Ранка Станковић, ванредни професор
Рударско-геолошки факултет, Универзитет у Београду
5. др Јелена Јаћимовић, научни сарадник
Стоматолошки факултет, Универзитет у Београду