

Импресум

ЗА ИЗДАВАЧА:

проф. др Александар Јерков

office@unilib.bg.ac.rs

Универзитетска библиотека „Светозар Марковић“

Филолошки факултет, Универзитет у Београду

ИЗДАВАЧ:

Филолошки факултет, Универзитет у Београду

Универзитетска библиотека „Светозар Марковић“

Заједница библиотека универзитета у Србији

ГЛАВНИ И ОДГОВОРНИ УРЕДНИК:

проф. др Цветана Крстев

cvetana@matf.bg.ac.rs

Филолошки факултет, Катедра за библиотекарство и информатику

ОПЕРАТИВНИ УРЕДНИК:

др Александра Тртовац

aleksandra@unilib.bg.ac.rs

Универзитетска библиотека „Светозар Марковић“

УРЕДНИК ЕЛЕКТРОНСКОГ ИЗДАЊА:

Јелена Андоновски

andonovski@unilib.bg.ac.rs

Универзитетска библиотека „Светозар Марковић“

УРЕЂИВАЧКИ ОДБОР:

проф. др Александра Вранеш, проф. др Александар Јерков, проф. др Биљана Дојчиновић, Филолошки факултет, Универзитет у Београду; проф. др Елизабет Бур, Институт за романистику, Универзитет у Лајпцигу; проф. др Владан Девеџић, Факултет организационих наука, Универзитет у Београду; проф. др Милена Добрева, Факултет за медијске и когнитивне науке, Универзитет на Малти; др Томаж Ерјавец, Одсек технологије знања, Институт „Јожеф Стефан“, Љубљана; проф. др Светла Коева, Институт за бугарски језик, Бугарска академија наука; проф. др Денис Морел, проф. др Агата Савари, Универзитет Франсоа Рабле из Тура; проф. др Иван Обрадовић, Рударско-геолошки факултет, Универзитет у Београду; проф. др Гордана Павловић Лажетић, проф. др Душко Витас, Математички факултет, Универзитет у Београду; проф. др Катерина Здравкова, Факултет за информатичке науке и компјутерско инжињерство, Универзитет „Св.Кирил и Методије“ у Скопљу

ISSN 1450-9687 (штампано издање) Београд, год. 18, бр. 2, децембар 2018.
ISSN 2217-9461 (е-издање)

ИЗРАДА ВЕБ ПОРТАЛА ЧАСОПИСА:

Вукосав Средојевић

Универзитетска библиотека „Светозар Марковић“

ЛЕКТОРИ ЗА ЕНГЛЕСКИ ЈЕЗИК:

др Јелена Јаћимовић, *Стоматолошки факултет, Универзитет у Београду*

др Јелена Митровић, *Катедра за информатику и математику, Универзитет Пасау*

ЛЕКТОР ЗА СРПСКИ ЈЕЗИК:

Свјетлана Ћелић

Универзитетска библиотека „Светозар Марковић“

ДИЗАЈН И ПРИПРЕМА ЗА ШТАМПУ:

Александар Милошевић, Бранислава Шандрих

и **Инфошека** тим

РЕДАКТОР БИБЛИОГРАФИЈЕ И УДК КЛАСИФИКАЦИЈА:

Наташа Дакић

Универзитетска библиотека „Светозар Марковић“

РЕДАКТОР ДОИ БРОЈЕВА:

др Милош Утвић

Филолошки факултет, Универзитет у Београду

РЕДАКЦИЈА ЧАСОПИСА:

Часопис Инфотека

Булевар краља Александра 71, 11000 Београд

+381 11 3370–211

infotheca@unilib.rs

ШТАМПА:

Мамиго плус

Београд

Часопис излази два пута годишње

Садржај

Научни радови

Јелена Анђелковић, Даница Сеничић, Даница Сеничић	
Паралелизовани корпус за домен менаџмента: поступак припреме и могућности примене . . .	7
Цветана Крстев, Денис Морел, Душко Витас	
Интеграција српског језика у вишејезични речник Prolexbase	32

Стручни радови

Василије Милновић	
Презентација српске књижевности, издаваштва и културе на сајмовима књига у свету	57
Весна Абадић, Марија Гордић	
Платформа за отворену науку – обавезност публикавања у отвореном приступу у Републици Србији . .	66

Прикази

Ивана Гавриловић	
Фокус на отвореној науци у Универзитетској библиотеци „Светозар Марковић“	76
Бранислава Шандрих	
Приказ летње школе „ESSLI 2018“	80
Емина Чано Томић	
Конференција COBISS 2018 – 30 година система	87

Паралелизовани корпус за домен менаџмента: поступак припреме и могућности примене

УДК 81'322.2

УДК 81'373:005

САЖЕТАК: У раду је представљен паралелизовани енглеско-српски корпус за домен менаџмента као једна од већих колекција доступна у оквиру алата за претраживање паралелизованих текстова Библиша. Поред описа садржаја корпуса и критеријума за његов одабир, као и самог процеса компилације и припреме текстова за корпус, у раду су кроз примере анализе корпуса приказане и могућности његове примене у различитим врстама лингвистичких студија, првенствено у терминолошким, компаративним и контрастивним истраживањима језика струке, истраживањима специјализованог и академског дискурса, у процесу превођења, али и у настави енглеског језика струке.

КЉУЧНЕ РЕЧИ: паралелизовани корпус, доменски корпус, терминологија, терминолошка јединица, менаџмент

РАД ПРИМЉЕН: 29. децембар 2017.

РАД ПРИХВАЋЕН: 15. новембар 2018.

Јелена Анђелковић

plecasj@fon.bg.ac.rs

Факултет организационих наука

Даница Сеничић

danica.senicic@gmail.com

Ранка Станковић

ranka.stankovic@rgf.bg.ac.rs

Рударско-геолошки факултет

Универзитет у Београду

Србија

1. Увод

У ранијој лингвистичкој литератури под корпусом се подразумевала било која колекција текстова за коју се претпоставља да представља дати језик, дијалекат или други подсистем неког језика, а који се користи ради лингвистичке анализе (Pearson, 1998, 42). У новије време, са развојем рачунарске лингвистике и алата и система за обраду природних језика, мења се и дефиниција корпуса. Данас се

под корпусом подразумевају машински читљиве колекције текстова (McEnergy and Wilson, 2001, 177), односно колекције делова језичког текста у електронском облику, које су одабране на основу екстерних критеријума да представе, колико је то могуће, језик или језички варијетет као извор података за језичка истраживања (Sinclair, 2005) Макенери, Ксао и Тоно (McEnergy et al., 2006, 13) под корпусом подразумевају колекције машински читљивих, аутентичних текстова (укључујући и транскрипте писаног говора) који су узорковани тако да буду репрезентативни за одређени језик или језички варијетет. Постоје различите поделе и врсте језичких корпуса, а сам одабир врсте корпуса зависи од његове намене, то јест, врсте лингвистичке анализе која се спроводи. Док су корпуси општег језика (енгл. *general language corpora*) колекције писаног и / или говорног језика који представљају (или би требало да представе) језик као целину, специјализовани корпуси, корпуси језика за посебне намене или доменски корпуси (енгл. *specialized corpora* / *corpora for specific purposes* / *domain corpora*) јесу електронски доступне колекције текстова које представљају одређену специјализовану област комуникације, односно које су репрезентативне за посебне намене одређеног језика или посебан домен језичке употребе. Често су специјализовани корпуси састављени од текстова који припадају једном жанру (енгл. *genre-specific*), односно репрезентативни су само за одређену научну и стручну област, дисциплину или домен (енгл. *domain-specific* или *discipline-specific*).

Од посебног значаја за терминолошка, контрастивна и компаративна истраживања језика су паралелни и паралелизовани доменски корпуси (корпуси за посебне намене). Паралелни (енгл. *parallel*) су двојезични или вишејезични корпуси који садрже преводно еквивалентне текстове на два или више језика (Tognini-Bonelli, 2001, 6)(Pearson, 1998, 47). У неким случајевима, паралелни корпуси могу садржати текстове на само једном језику, тј. парове еквивалентних различитих превода на исти језик. Код паралелизованих корпуса (енгл. *aligned corpora*), са друге стране, паралелни (преводно еквивалентни) текстови су и паралелизовани, односно поравнати на нивоу пасуса, реченице или појединачних речи.

Основна карактеристика паралелизованих корпуса текстова за посебне намене јесте да су они примарни језички ресурси за вишејезичну обраду у оквиру рачунарске лингвистике. Ови корпуси омогућавају систематичну обраду велике количине терминолошких информација, аутоматску или полуаутоматску екстракцију термина и

њихових еквивалената на страном језику или језицима. Лексичко знање стечено коришћењем паралелизованих корпуса је стога неопходно за развој софтверских система и алата за аутоматску обраду природних језика (енгл. *natural language processing* – NLP), као што су системи за машинско превођење и генерисање билингвалних електронских термилошких глосара, речника, лексикона и термилошких база података (Véronis, 2013, 238).

Иако број и доступност паралелизованих колекција текстова који обухватају и српски језик нисује на завидном нивоу, највише због чињенице да је често тешко пронаћи адекватне преводно еквивалентне текстове, у наставку поглавља представимо неколико постојећих.

Први радови везани за паралелизоване текстове где је један од језика српски везују се за почетак деведесетих година, где се о конкорданцијама паралелизованих текстова говори у (Krstev C., 1994), а касније се паралелизује Платонова *Република* на 17 језика (Vitas, 1998b) и превод Орвеловог романа „1984” поравнат на седам језика (Vitas, 1998a).

Евротекa¹ је двојезични, енглеско-српски, корпус правних текстова или делова правних текстова настао током процеса превођења правних тековина Европске уније на српски језик, за који је задужено Министарство за европске интеграције. Евротеку од 2009. године одржава и надограђује Сектор за комуникације овог министарства. За превођење правних текстова, те као и за креирање овог корпуса, користи се програм SDL Trados.², тј. његов алат Translator’s Workbench

По броју језика који обухвата, издваја се и вишејезични паралелизовани корпус „MULTEXT-East "1984" annotated corpus 4.0” који је расположив на CLARIN.SI репозиторијуму, а развијен у оквиру пројекта MULTEXT-East – *Multilingual Text Tools and Corpora for Eastern and Central European Languages* (Erjavec and Ide, 1998). Корпус MULTEXT-East "1984" састоји се од енглеског оригинала романа „1984” аутора Џорџа Орвела и његовог превода на дванаест источно и централноевропских језика (бугарски, чешки, енглески, естонски, мађарски, македонски, персијски, пољски, румунски, српски, словачки и словеначки). Текстови су поравнати на нивоу реченице, а леме и морфосинтаксички описи су ручно валидирани (Erjavec et al., 2010). Паралелизовани корпус српског и енглеског језика srenWaC³ (Ljubešić

¹ Евротекa (на вебу)

² Trados (на вебу)

³ srenWaC (на вебу)

et al., 2016) састоји се од електронских текстова различите тематике преузетих са домена .rs, а генерисан је аутоматски уз коришћење алата Spidextor.⁴

(Pazienza et al., 2005), као и други, описује различите приступе за екстракцију термина: 1) коришћење статистичких мера при одабиру добрих термина са листе кандидата, 2) идентификовање и препознавање израза користећи само лингвистичка и филтрирање специфичних синтактичких терминолошких образаца и 3) хибридне приступе који покушавају да користе ова два приступа заједно, узимајући у обзир синтаксичка својства и статистичке мере за препознавање термина. (Siddiqi and Sharan, 2015) наводи још два приступа: 4) коришћење метода машинског учења и 5) доменски специфичних ресурса знања (на пример онтологија) при екстракцији термина. У овом раду је коришћен хибридни приступ који комбинује препознавање синтаксичких образаца и статистичке мере, а који је описан у (Stanković et al., 2016a).

Паралелизоване и електронски доступне корпусе на енглеском и српском језику развија Група за језичке технологије на Математичком факултету⁵ и Друштво за језичке ресурсе и технологије (JePTex)⁶ чији су чланови углавном са Универзитета у Београду. Један од паралелизованих корпуса јесте енглеско-српски корпус *СрпЕнгКор*, који се састоји од текстова различитих жанрова (књижевност, новинарство, научни и стручни текстови из области права, медицине, образовања итд.) који су сегментирани и поравнати (у највећем броју случајева) на нивоу реченице.⁷

У оквиру Друштва Јертех је развијен је алат за претраживање паралелизованих колекција стручних текстова – Библиша.⁸ Детаљан опис имплементације и начина употребе је представљен је у (Stanković et al., 2016b). Тренутно је путем Библише доступно неколико колекција стручних текстова, и то за научне и стручне области библиотекарства и информатике, рударства и геологије, стоматологије, архитектуре и урбанизма и других.

У следећим поглављима која следе описан је поступак компилације и обраде једне од већих колекција паралелизованих текстова доступних путем алата Библиша, колекције текстова за домен менаџмента.

⁴ Spidextor (на вебу)

⁵ Група за језичке технологије на Математичком факултету (на вебу)

⁶ Друштво за језичке ресурсе и технологије (на вебу)

⁷ СрпЕнгКор (на вебу)

⁸ Библиша (на вебу)

Поред процеса израде, у раду су истакнуте и могућности примене ове колекције у различитим врстама лингвистичким истраживањима. Акценат није стављен на примарну намену паралелизованих доменских корпуса (у рачунарској лингвистици и менаџменту терминологије), већ на мање истражене могућности примене, и то пре свега у примењеној лингвистици и компаративном и контрастивном проучавању терминологије, језика струке и специјализованог дискурса.

2. Паралелизовани корпус за домен менаџмента

2.1 Садржај корпуса

Паралелизовани енглеско-српски специјализовани корпус за домен менаџмента састоји се од текстова који припадају жанру научног рада, а публиковани су у међународном научном и стручном часопису *Менаџмент: Часопис за теорију и праксу менаџмента* (енгл. *Management: Journal for theory and practice of management*).⁹ Реч је, дакле, о корпусу који је специфичан како за домен, односно научну и стручну област менаџмента (енгл. *domain-specific corpus*), тако и о корпусу који је специфичан за жанр научног рада (енгл. *genre-specific corpus*).

Корпус садржи 17 бројева часописа *Менаџмент* објављених између 2008. и 2012. године¹⁰, са укупно 181 научним радом, око 30.000 реченица, и преко 600.000 речи по језику (прецизније, 611.651 реч на српском језику). Детаљнији приказ садржаја корпуса представљен је у табели 1.

Међународни научни и стручни часопис *Менаџмент: Часопис за теорију и праксу менаџмента* квартално издаје Факултет организационих наука Универзитета у Београду као водећа академска институција за ову област у Србији. Часопис има циљ да “омогући

⁹ У мају 2017. године, часопис је променио назив из *Management: Journal for Theory and Practice of Management* у *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies*.

¹⁰ Часопис *Менаџмент* од 1996. (када је основан) до 2008. (закључно са бројем 46) излази само на српском језику. Од 2008. до 2013. године сваки број са идентичним садржајем се штампа и на енглеском и на српском језику, док се од 2013. године, почевши од броја 66, штампа искључиво на енглеском језику. Чињеница да су паралелни текстови доступни само за период између 2008. и 2012. године ограничила је и величину нашег корпуса на 17 бројева: од броја 47–48 до броја 65.

размену релевантних информација и комуникацију између научника, истраживача, менаџера и појединаца из различитих пословних области, који долазе са универзитета, института, компанија и јавних услужних предузећа”¹¹. У време када су издати бројеви који су обухваћени нашим корпусом, часопис *Менаџмент* се налазио у категорији М51 по категоризацији Министарства просвете, науке и технолошког развоја Републике Србије. Сви радови на српском и енглеском језику, обухваћени корпусом, доступни су на сајту часописа.¹²

издање (број/година)	број радова (по језику)	број реченица (српски језик)
47-48/2008	12	2.187
49-50/2008	14	2.097
51/2009	9	1.503
52/2009	9	1.575
53/2009	10	1.194
54/2010	10	1.817
55/2010	10	1.750
56/2010	10	1.648
57/2010	10	1.502
58/2011	10	1.475
59/2011	10	1.501
60/2011	11	1.426
61/2011	14	2.301
62/2012	12	2.297
63/2012	10	1.815
64/2012	10	1.655
65/2012	10	1.583
Σ	181	29.326

Табела 1. Садржај корпуса

¹¹ Менаџмент: Часопис за теорију и праксу менаџмента (на вебу)

¹² Менаџмент, архива (на вебу). И поред јавне доступности, за њихово коришћење при изради паралелизованог корпуса добијена је дозвола тадашњег уредника часописа Менаџмент, проф. др Александра Марковића.

2.2 Предности и ограничења корпуса

Ауторство. Радови обухваћени корпусом дела су појединачних аутора, али и композитни текстови од више аутора. Аутори текстова су истраживачи у области менаџмента и представници академске заједнице, али и представници привреде из наше земље, региона и света. Увидом у метаподатке, уочили смо да су у 21 раду (11,6%) од укупно 181 рада аутори (или аутор) искључиво страни држављани (ван простора бивше Југославије), а преосталих 160 радова (88,4%) написали су или домаћи аутори или аутори из региона, самостално или у коауторству.

На основу доступних метаподатака и информација са веб-сајта часописа није могуће тачно утврдити који језик су аутори користили за писање изворног текста, као ни да ли је текст преводио сам аутор или стручни преводилац; претпоставка је да су радови страних аутора изворно написани на енглеском језику, а затим преведени на српски, док су радови домаћих и аутора из региона највероватније изворно написани на српском језику. Иако овакав састав текстова у корпусу може значајно утицати на квалитет, односно прецизност и једнозначност терминологије која се у њима налази, он потенцијално може дати бољу слику о терминолошкој и другој језичкој варијацији условљеној прагматичким или социолингвистичким факторима. Из овог разлога није вршена селекција текстова на основу ауторства, односно на основу изворног језика текста.

Прагматички фактори за одабир текстова. Приликом одабира текстова за паралелизовани корпус менаџмента, водили смо се првенствено прагматичким факторима: доступношћу адекватних преводно еквивалентних текстова за домен менаџмента у електронском облику, као и циљевима које желимо да постигнемо изразом и коришћењем оваквог корпуса, а то су лингвистичка и терминолошка истраживања у овој области и сродниом научно-стручним областима. Оба ова фактора са собом доносе одређене предности и ограничења.

Величина корпуса. Величина представљеног корпуса (око 600.000 речи по језику) последица је доступности преведених научних радова из домена менаџмента на српском и енглеском језику. Иако се аутори који се баве корпусном лингвистиком не слажу у потпуности око оптималне величине корпуса (Roe, 1977; Fang, 1993; Gledhill, 2000), односно око идеалне величине специјализованог корпуса (Flowerdew, 2004, 18), сматрамо да је овде представљен корпус менаџмента адекватан

за многе, али не и за све врсте лингвистичких и термилошких анализа у овом домену и сродним стручним доменима, због чега би за такве анализе дати корпус требало додатно проширити.

Жанр. Са функцинално-стилистичког гледишта, наш паралелизовани доменски корпус у потпуности припада једном текстуалном жанру језика струке, тачније жанру истраживачког научног рада. Припадност текстова једном жанру доприноси хомогености корпуса, будући да су приступ тематици и ниво специјалности уједначени, односно да регистарски, стилски и жанровски не постоје велике варијације између текстова. Иако је код општих корпуса препоручљива разноврсност жанрова, у термилошким истраживањима и истраживањима језика струке сасвим су прихватљиви и веома често коришћени корпуси засновани на једном жанру. Имајући у виду да жанр истраживачког научног рада карактеришу једнообразност (одсуство разговорне и дијалекатске лексике), лексичка и термилошка густина, као и информативност, логичност, целовитост и прецизност текста, сматрамо да је одабир овог жанра адекватан за термилошку анализу, док је за одређене врсте лингвистичких истраживања потребно укључити и досупне текстове других жанрова.

Компилација и припрема корпуса. Процес припреме корпуса за анализу путем одговарајућих софтверских алата састојао се од неколико фаза: припрема и екстракција текста, поравнања текстова на нивоу пасуса и реченица, креирање докумената у форматима TEI/XML и TMX, снабдевање метаподацима и укључивање у базу.

2.3 Припрема и екстракција текста

Након одабира текстова увршћених у наш паралелизовани корпус за домен менаџмента и жанр научног рада, сви текстови су појединачно преузети у формату PDF са веб-странице часописа *Менаџмент*. Будући да је формат обичног текста (.txt) стандардан за софтвер намењен за обраду и анализу корпуса, све електронске текстове смо конвертовали у овај формат користећи програм Abby PDF Transformer. Ради лакше идентификације текстова, обележили смо их одговарајућим ознакама (на пример, називом Mng52_01-sr што се односи на датотеку која садржи први чланак часописа из броја 52 на српском језику). Проблеми на који смо повремено наилазили приликом припреме корпуса јесу да конверзија докумената из расположивог формата PDF у формат

обичног текста није препознала неке од карактера који се појављују у корпусу, пре свега дијакритичке ознаке у српским текстовима (сви текстови на српском језику писани су латиничним писмом), као и да се две колоне са текстом у формату PDF приликом конверзије у формат обичног текста често спајају у једну. Како би се умањиле грешке, поједине датотеке су конвертоване у формат програма Microsoft Word (.doc), затим кориговане и снимљене у формату обичног текста.

Након конвертовања текстова одстрањени су сви елементи нерелевантни за лингвистичку анализу, као што су табеле, графикони, формуле, литература, референце, садржај, заглавља, подножја и сл.

Овако припремљени текстови погодни су за лингвистичку анализу неанотираних (или „сирових“) корпуса (енгл. *raw corpora*) на једном језику коришћењем јавно доступних и често бесплатних програма за анализу корпуса какви су, на пример, WordSmith¹³ или AntConc.¹⁴ За анализирање паралелних корпуса, међутим, потребно је извршити поравнање текста на неком од структурних нивоа (одељци, пасуси, реченице или речи), а за адекватно анализирање високо флективног језика какав је српски, неопходна је и морфосинтаксичка анотација овог дела корпуса.

Поравнање текстова на нивоу пасуса. После припреме појединачних текстова, парови одговарајућих изворних текстова (на српском језику) и циљних текстова (на енглеском језику) поравнати су на нивоу пасуса (нпр. Mng52_01-sr и Mng52_01-en). Овај процес обављен је у програму Notepad++, поређењем њихових садржаја и поравнањем пасуса, тако да се пасуси српског оригинала и пасуси енглеског превода, сваки у својој датотеци, морају налазити у истом реду. Приликом овог процеса, наишли смо на многобројне проблеме, као што су непреведени, неадекватно преведени, изостављени или измештени пасуси. Ове проблеме решавали смо или проналаском делова пасуса који недостају у оригиналним документима у формату PDF, или простим одстрањивањем пасуса или делова пасуса за које не постоји еквивалент у другом језику. Основни разлог за овако захтеван и дуготрајајан поступак јесте смањење тзв. буке/шума (енгл. *noise*) током корпусне анализе.

¹³ WordSmith (на вебу)

¹⁴ AntConc (на вебу)

Поравнање текстова на нивоу реченица и креирање XML документа. Трећи корак у припреми паралелизованог корпуса је креирање документа у формату XML (*eXtensible Markup Language*) поравнатих на нивоу реченица. Текстови у формату XML, поред основног текста могу да садрже и додатне интерпретативне лингвистичке податке, као што су информације о структури текста, информације о ауторима и верзијама текста, као и лингвистичку анотацију текста засновану на процесима токенизације, препознавање граница реченица, морфолошке анализе (укључујући лематизацију и анотацију врсте речи, енгл. *Part-of-Speech tagging* или *PoS tagging*), као и плитка синтаксичка анализа (енгл. *shallow parsing*).

Пре поравнања текстова на нивоу реченице, неопходно је сегментирати их на том нивоу. Овај корак је аутоматски обављен помоћу система Unitex (Paumier, 2002) који се користи за креирање и преграживање корпуса. Реченице су сегментиране помоћу локалних граматика, форми за опис и препознавање лингвистичких феномена у тексту. Локалне граматике су имплементирание као коначни аутомати и трансдуктори којима корисник манипулише помоћу њиховог графичког приказа, тј. графова. Локалне граматике за препознавање краја реченице прилагођене су правопису српског језика и саставни су део лингвистичких ресурса за српски језик који се дистрибуирају са самим програмом Unitex. Резултат сегментације на реченице, излазни текст, садржи симбол {S} као граничник реченице, који се даље конвертује у одговарајуће етикете формата TEI/XML за означавање реченица (сегмената) у складу са Смерницама TEI P5.¹⁵, незваничног стандарда за кодирање текста који се највише користи

Обележавање текста (енгл. *markup*) на структурном нивоу пасуса и реченица олакшава процес упаривања изворних и циљних текстова.

У овом истраживању структурни нивои текста су обележени (Слика 1) етикетама <div> (цео документ), <body> (садржина целог документа), <head> (наслови), <p> (пасуси) и <seg> (реченице).

Претраживање корпуса и генерисање конкорданција је подржано је морфолошком и семантичком експанзијом упита (Stanković et al., 2016b), тако да лематизација није била потребна.

Креирање TMX документа. Наредни корак у припреми корпуса јесте креирање документа у формату TMX (Savourel, 2004). Ради се XML

¹⁵ TEI P5 (на вебу)

40	<p><seg>Postoje gledišta da je integrisanje IT sa sistemima ljudskih aktivnosti osnovni problem u oblasti IS i da je zanemarivanje "ljudskog okruženja" tj. punog društvenog konteksta osnovni razlog velikog procenta neuspešnih IS ([13]).</seg>	40	<p><seg>There are also arguments, that integrating IT with the systems of human activities is the basic problem in the IT area and that the real cause of a high percentage of failed IS is the neglect of "human environment", that is, of the entire social context ([13]).</seg>
41	<seg> Direktni korisnici IS često ne učestvuju u donošenju odluka o njegovom razvoju.</seg>	41	<seg> The direct IS users are usually left out of a decision making process concerning IS development.</seg>
42	<seg> Informacioni sistem, osim na svoja direktna korisnike, često utiče i na druge zainteresovane grupe, čiji se interesi, međutim, često ne razmatraju u procesu razvoja IS.</seg></p>	42	<seg> The information system, apart from having an impact upon its direct users, frequently affects other interested parties (stakeholders), whose interests, however, often fail to be taken into consideration in the IS development process.</seg></p>

Слика 1. Пример припремљеног паралелног српско-енглеског текста у формату TEI/XML

спецификацији за размену података из преводилачких меморија (енгл. *Translation Memory eXchange*) и користи се често у алатима намењеним за превођење уз коришћење рачунара (енгл. *computer-aided translation* – CAT). Документи у формату TMX креирани су у програму ACIDE, интегрисаном окружењу за припрему паралелизованих корпуса које је развило Друштво за језичке ресурсе и технологије у Београду (Obrodović et al., 2008, 563). ACIDE нуди графичко сучеље (енгл. *interface*) за поравнање и визуализацију поравнатих текстова, док се за само поравнање користе програмски пакети XAlign и Concordancier развијени у лабораторији LORIA.¹⁶ у Француској (Bonhomme et al., 2001) Пример упарене реченице у TMX формату је приказан је на слици 2.

```
<tu>
  <prop type="Domain">Mitić M., 2010, vol. XV:54, ID: 9.2010.54.9</prop>
  <tuv xml:lang="en" creationid="n21 " creationdate="20161206T160737Z">
    <seg>There are also arguments, that integrating IT with the systems of human activities is
    the basic problem in the IT area and that the real cause of a high percentage of failed IS
    is the neglect of "human environment", that is, of the entire social context ([13]). </seg>
  </tuv>
  <tuv xml:lang="sr" creationid="n21 " creationdate="20161206T160737Z">
    <seg>Postoje gledišta da je integrisanje IT sa sistemima ljudskih aktivnosti osnovni
    problem u oblasti IS i da je zanemarivanje "ljudskog okruženja" tj. punog društvenog
    konteksta osnovni razlog velikog procenta neuspešnih IS ([13]). </seg>
  </tuv>
</tu>
```

Слика 2. Пример преводилачке јединице са енглеском и одговарајућом српском реченицом у формату TMX

¹⁶ LORIA (на вебy)

Напоследку, текстови у формату ТМХ укључени су у базу података креирану у оквиру платформе MongoDB.¹⁷ Укључивањем у базу коришћењем алата Библиша¹⁸ паралелни текстови су доступни за претраживање и даљу анализу.

Снабдевање корпуса метаподацима и укључивање у базу. Претходно припремљени ТМХ документи су укључени у Библишу (Stan-ković et al., 2016a) као седма колекција у оквиру базе паралелизованих текстова. Сама колекција подељена је на 17 потколекција које садрже текстове одговарајућих 17 бројева часописа *Менаџмент*. У оквиру сваке потколекције има између 9 и 12 докумената, односно радова. Свака потколекција (свеска часописа) и чланак имају свој јединствен идентификациони број. На пример, идентификациони број потколекције је 7.2011.59, што значи да је у питању потколекција (свеска часописа) седме колекције (часопис *Менаџмент*), да је свеска издата 2011. године и да је број свеске 59, док 7.2011.59.1 представља први чланак те свеске. Сваки од докумената у оквиру потколекција снабдевен је библиографским метаподацима (на енглеском и српском језику) који се односе на наслове, ауторе радова, њихове афилијације и контакте (имејл адресе), хипервезе ка чланцима у формату PDF, апстрактне и кључне речи на оба језика, као и метаподатке од којих се састоји идентификациони број (редни број чланка, број часописа, година издања).

3. Анализа корпуса и могућности примене

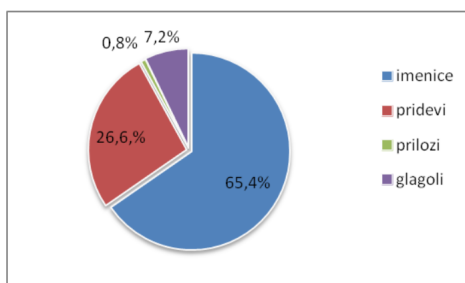
Једна од важних предности које двојезични (у нашем случају паралелизовани) доменски корпус нуди у односу на једнојезичне јесте могућност анализе делова корпуса на сваком од два језика појединачно, али и могућност интерлингвалне компаративне и контрастивне анализе. У наредним редовима представимо неке од основних могућности примене нашег доменског корпуса, прво путем екстракције једночланих и вишечланих терминолошких јединица у српском делу корпуса, а затим и екстракције парова преводних еквивалената на два језика.

¹⁷ MongoDB (на вебу)

¹⁸ Библиша (на вебу)

3.1 Екстракција кључних речи на српском језику

Након обављеног процеса компилације и обраде текста паралелизованог корпуса менаџмента, извршена је екстракција лексичких јединица на српском језику које се издвајају по критеријуму кључности (енгл. *keyness*), односно оних које се значајно фреквентније појављују у доменском корпусу него у референтном корпусу, тј. *Корпусу савременог српског језика СрпКор 2003*¹⁹ (122 милиона речи), насталом на Математичком факултету Универзитета у Београду (Utvić, 2011, 36a-47a). Овај процес обављен је помоћу алата за развој и управљање лексичким ресурсима LeXimig који је развио JePTex, Друштво за језичке ресурсе и технологије (Stanković et al., 2011, 77-84). Анализом првих 500 кључних лексичких јединица издвојених из доменског корпуса установљено је да међу њима има 327 именица, 133 придева, 36 глагола и 4 прилога (Слика 3).



Слика 3. Дистрибуција врста речи у скупу од 500 изабраних кључних лексичких јединица

Детаљнијом анализом и филтрирањем добијене листе, издвојили смо кључне именице (Табела 2), придеве (Табела 3) и глаголе (Табела 4). Параметар кључност се рачуна као однос релативне фреквенције (изражене у милионитим деловима целине, тј. у ppm као јединицама мере) у доменском корпусу (менаџмента у овом случају) и одговарајуће релативне фреквенције у референтном корпусу опште намене, при чему су оба броја пре дељења увећана за један. Ознаке RFr и RFd

¹⁹ СрпКор 2003 (на веб)

редом представљају релативне фреквенције (у ppm) у референтном и доменском корпусу, док су AFr и Afd ознаке одговарајућих апсолутних фреквенција. Дакле, кључност рангира леме према односу фреквенција у доменском и референтном корпусу, а не само према фреквенцијама у доменском корпусу. Леме које се чешће појављују у доменском корпусу него у референтном, узимајући у обзир величине тих корпуса, биће при врху табеле и вероватно су оне термини домена.

лема	кључност	RFr	RFd	AFr	Afd
индикатор	121,398	3,035	488,841	67	299
менаџмент	115,894	16,713	2051,824	369	1255
рачуноводство	115,656	3,306	497,015	73	304
бренд	107,933	1,857	307,365	41	188
портфолио	84,509	1,314	194,555	29	119
интернет	82,145	4,167	423,444	92	259
профитабилност	81,684	1,314	188,016	29	115
перформанса	81,194	4,892	477,396	108	292
подсистем	79,44	0,906	150,413	20	92
сертификација	69,345	1,042	140,603	23	86
конкурентност	67,896	4,529	374,397	100	229
евалуација	65,029	0,725	111,175	16	68
управљање	63,609	38,726	2525,95	855	1545
преференције	62,063	0,544	94,825	12	58
методологија	59,253	6,522	444,698	144	272

Табела 2. Кључне именице у доменском корпусу

Табеле 2, 3 и 4 показују да кључне термилошке јединице не морају уједно бити и најфреквентније. Термилошке јединице које су истовремено и изразито фреквентне и кључне за домен менаџмента истовремено су и најважније термилошке јединице за ову област: *менаџмент* (кључност = 115,894, Afd = 1255) и *управљање* (кључност = 63,609, Afd = 1545), али и термилошке јединице чија се синонимна употреба често доводи у питање (о овоме ће бити више речи у одељку 3.3). Са друге стране, на пример, глаголи попут *операционализовати* (кључност = 15,316, Afd = 13), *инкорпорирати* (кључност = 15,039, Afd

лема	кључност	RFr	RFd	AFr	Afd
пројектни	215,922	3,578	987,491	79	604
корпоративан	146,506	2,31	483,936	51	296
одржив	123,302	3,397	541,158	75	331
мотивациони	95,619	0,498	142,238	11	87
ефективан	85,99	2,491	299,19	55	183
рачуноводствени	84,553	2,763	317,174	61	194
екстерни	83,349	2,582	297,555	57	182
иновативан	83,031	1,178	179,841	26	110
управљачки	76,955	4,303	407,095	95	249
стратегијски	68,548	5,979	477,396	132	292
организациони	68,428	20,518	1471,427	453	900
проблемски	64,179	2,582	228,889	57	140
конкурентски	62,603	5,571	410,365	123	251
менаџерски	61,658	2,808	233,793	62	143

Табела 3. Кључни придеви у доменском корпусу

лема	кључност	RFr	RFd	AFr	Afd
фокусирати	47,821	3,397	209,27	75	128
имплементирати	40,691	2,038	122,619	45	75
генерисати	33,435	2,355	111,175	52	68
израчунавати	31,611	1,359	73,571	30	45
класификовати	20,779	2,038	62,127	45	38
базирати	19,884	10,644	230,524	235	141
рангирати	19,208	2,627	68,667	58	42
позиционирати	17,702	0,996	34,333	22	21
дефинисати	17,287	53,628	943,348	1184	577
формализовати	17,149	0,679	27,794	15	17
операционализовати	15,316	0,453	21,254	10	13
обухватати	15,087	36,778	568,952	812	348
креирати	15,048	14,494	232,159	320	142
инкорпорирати	15,039	1,132	31,063	25	19

Табела 4. Кључни глаголи у доменском корпусу

= 19) и *позиционирати* (кључност = 17,702, Afd = 21) имају релативно висок параметар кључности иако нису фреквентни у доменском корпусу.

3.2 Екстракција вишечланих термина на српском језику

Имајући у виду да су термилошке јединице најчешће вишечлане (Krstev et al., 2015), а не једночлане, издвојена листа једночланих термилошких јединица није довољна за термилошку анализу. Стога је потребно укључити и вишечлане термилошке јединице до којих се долази аутоматском екстракцијом помоћу синтаксичких графова развијених у оквиру програма Unitex. Коришћењем алата Лексмир екстрахују се кандидати према дефинисаним синтаксичким обрасцима (Stanković et al., 2016b)(Krstev et al., 2015), препознате лексичке јединице се лематизују, како би се објединила појављивања исте вишечлане леме. Узимајући у обзир вишезначност овакве лематизације, имплементиране су различите стратегије разрешавања вишезначности. Потом се за сваког термилошког кандидата из генерисане листе, рачунају различите статистичке мере на основу којих се рангирају кандидати, након чега следи евалуација и одређивање термина који ће ући у термилошки речник (Stanković et al., 2016b).

Увидом у двадесет најфреквентнијих вишечланих термилошких јединица на српском језику у доменском корпусу за менаџмент (Табела 5) можемо закључити да су међу екстрахованим вишечланим терминима најфреквентније именичке синтагме са придевом као зависним чланом у препозицији (grf01, образац придев именица, у ознаци AXN, код кога се придев слаже у роду, броју и падежу са именицом), као у примерима *људски ресурси*, *информациони систем*, *електронско пословање*, *управљачко рачуноводство* итд.

Екстракција термилошких јединица из српског дела нашег доменског корпуса и њихова детаљнија анализа може бити од великог значаја за проучавање семантичких, прагматичких и социолингвистичких аспеката термилошких јединица менаџмента, затим контрастивних и компаративних проучавања стручне терминологије у домену менаџмента (посебно у односу на енглески језик), али може представљати и допринос термилошкој језичкој политици, планирању, систематизацији и стандардизацији терминологије овог домена. Иако примери наведени у овом и претходном одељку указују првенствено на примену корпуса у оквиру термилошких студија, доменски корпус за менаџмент је могуће

користити и у проучавању српског језика струке и специјализованог дискурса менаџмента, затим у анализи академског писања и анализи жанра научног рада на српском језику, али и другим врстама лингвистичких истраживања, што је у плану за будући истраживачи рад.

Graph	Pattern	лема	фрек.	бр. речи	рел. фрек.
grf01	AXN	људски ресурси	258	2	421.81
grf01	AXN	организациона наука	207	2	338.43
grf01	AXN	информациони систем	190	2	310.63
grf01	AXN	управни одбор	181	2	295.92
grf01	AXN	електронско пословање	162	2	264.86
grf01	AXN	пројектно финансирање	136	2	222.35
grf01	AXN	пројектни менаџмент	128	2	209.27
grf01	AXN	конкурентска предност	127	2	207.63
grf01	AXN	управљачко рачуноводство	122	2	199.46
grf01	AXN	пројектни менаџер	100	2	163.49
grf01	AXN	финансијски извештај	98	2	160.22
grf03	N2X	реализација пројекта	97	2	158.59
grf03	N2X	управљање ризиком	97	2	158.59
grf01	AXN	информациона технологија	95	2	155.32
grf01	AXN	каматна стопа	95	2	155.32
grf01	AXN	енергетска ефикасност	93	2	152.05
grf03	N2X	процес управљања	92	2	150.41
grf10	2XAXN	јавно-приватно партнерство	90	3	147.14
grf01	AXN	економска криза	86	2	140.6
grf01	AXN	финансијско средство	77	2	125.89

Табела 5. Најфреквентније вишечлане термилошке јединице у доменском корпусу

3.3 Екстракција преводних еквивалената на енглеском језику

У претходна два одељка представљени су основни резултати анализе српског дела нашег доменског корпуса. Прави потенцијал овог паралелизованог корпуса, међутим, може се искористити

софтверским алатом Библиша. Комплетним текстовима који припадају представљеном корпусу могу приступити регистровани корисници путем веб-сајта <http://jerteh.rs/biblisha/>. Ограничено коришћење могуће је и без регистрације и овлашћења – тада је доступно првих девет реченица сваког текста приликом прелиставања докумената или 30 конкорданција при претраживању.

Уношењем упита – једночланих или вишечланих терминолошких јединица у претраживач Библише на једном од два језика – можемо доћи до парова конкорданција на енглеском и српском језику. Парови конкорданција не само да дају преводни еквивалент за задати упит, већ пружају и контекст употребе задате речи или израза на два језика. Уношењем енглеског термина *management* у претраживач Библише, на пример, добијамо све конкорданције ове терминолошке јединице на енглеском језику, као и паралелне реченице (преводе) на српски језик, из којих се могу издвојити преводи овог термина на српски (Табела 6), при чему су пронађени термини аутоматски истакнути плавом бојом. Треба поменути да Библиша, осим морфолошке експанзије упита, пружа могућност да се упит прошири и семантички коришћењем семантичке мреже WordNet и више терминолошких база. Уз то систем проналази еквиваленте у другом језику, чиме се омогућава издвајање паралелизованих реченица које: 1) проналазе еквиваленте у оба језика, 2) само у српском или 3) само у енглеском, што омогућава кориснику да користи систем за различите намене.

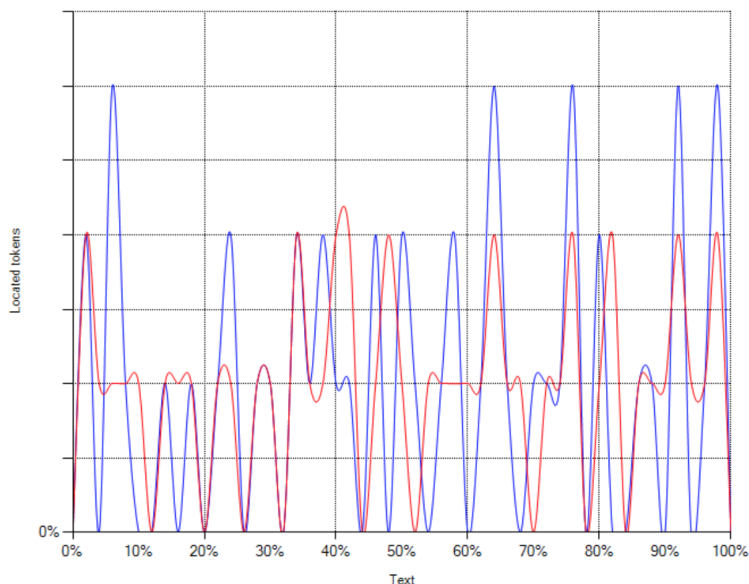
Табела 6 првенствено указује на различите преводне еквиваленте енглеског термина *management* на српски језик, и то као *менаџмент*, *управљање*, *руковођење*, *управа*, *управљачки*, *руководи*, *менаџерски* итд., као и на примере у којима енглески термин није преведен него задржан у оригиналном облику. Уочени српски еквиваленти енглеског термина *management* указују како на његову полисемичност (*management* као процес управљања или као група људи која управља), тако и на синонимну употребу термина *менаџмент*, *управљање* и *руковођење* (као процеса), односно термина *менаџмент*, *управа* и *руководство* са (као тима људи). Осим тога, детаљнијом анализом могуће је утврдити и околности, односно контекст у коме се енгл. *management* преводи придевом, глаголом, или на неки други начин, али се тиме нећемо бавити у овом раду.

Детаљнијим упитом, односно претраживањем парова састављених од енглеског термина и сваког од његових преводних еквивалената на српском језику појединачно (нпр. енгл. *management* и срп. *управљање*,

енгл. *management* и срп. *менаџмент*, или енгл. *management* и срп. *руководђење*) можемо доћи до конкретних примера конкорданција у којима је овај енглески термин преведен на одређени начин. Слика 4 приказује дистрибуцију енглеског термина *management* и српског термина *менаџмент* кроз наш доменски корпус који је уређен дијахроно, од првих радова публикованих 2008 (лево) до последњих у корпусу публикованих 2012 (десно).

Метаподаци	Concordances (en)	Конкорданције (ср): 1260
Милићевић et al., 2009, vol. XIV:53, ID: 7.2009.53.1	The modern business prefers an integrative approach in the implementation of management tools in tracking ...	У савременом бизнису је пожељан интегративни приступ коришћењу менаџерских алата у праћењу ...
Милосављевић et al., 2012, vol. XVII:62, ID: 7.2012.62.5	As an integral function of global management , human resource management has a task to explore and define, ... what it is that makes the organization unique on the market.	Менаџмент људских ресурса, као интегрална функција глобалног менаџмента , има задатак да истражује и дефинише, ... шта је то што организацију чини јединственом на тржишту.
Ђурић Д., 2010, vol. XIV:55, ID: 7.2010.55.2	n40 Essentially, financial accounting is part of management accounting and reporting.	n40 Суштински, финансијско рачуноводство представља део руководљачког рачуноводства и извештавања.
Mason R., 2012, No. 63, ID: 7.2012.63.1	The executives in our studies, even those that recognize the opportunities of this still emerging environment, still have widely divergent views on the most effective management models for realizing these opportunities.	Директори обухваћени нашом студијом, чак и они који су свесни могућности које даје ово ново окружење које је још у настајању, још увек се разликују у ставовима о томе који су најефективнији модели за реализацију ових могућности.
Пинтерић У., 2008, vol. XIII:49/50, ID: 7.2008.49-50.7	n78 Slovenian scientists wrote about introducing new public management elements into the work at all levels of Slovenian public administration ...	n78 Словеначки аутори писали су о увођењу елемената нове јавне управе у пословање на свим нивоима словеначке јавне управе ...
Станић С., 2008, vol. XIII:49/50, ID: 7.2008.49-50.6	n75 The internal factors include: the amount of media budget, the competence of management and administrative structure within the media department of the company or the hired marketing agency.	n75 Интерни фактори обухватају: величину медијског буџета, способности руководеће и административне структуре у оквиру медијског одељења компаније или ангажоване маркетинг агенције.
Домазет et al., 2009, vol. XIV:51, ID: 7.2009.51.4	n160 The Customer Relationship Management combines the business strategy and technology aiming to identify, attract and retain long-term relations with customers ...	n160 Customer Relationship Management комбинује пословну стратегију и технологију са циљем да идентификује, привуче и одржи дугорочне односе сакупцима ...
Вулић et al., 2012, No. 63, ID: 7.2012.63.7	The management is making organizational improvements in the country.	Руководство се организационо усавршава у земљи.
Hitka et al., 2009, vol. XIV:51, ID: 7.2009.51.8	n11 Managers from the area of manpower management have to deal with ... the problem ...	n11 Менаџери који управљају људском радном снагом морају да нађу прави одговор на питање ...
Панић С., 2012, No. 63, ID: 7.2012.63.9	To facilitate a two-way communication between the management and the employees, the company implemented three modes of communication...	Да би олакшала двосмерну комуникацију између управе и запослених, компанија је увела 3 начина комуникације...
Барјактаровић et al., 2011, vol. XVI:61, ID: 7.2011.61.1	A very important constituent of the overall bank management process is the implementation of the corporate governance principles.	Врло битан елемент целокупног процеса управљања банком јесте и примењивање принципа корпоративног управљања.

Табела 6. Конкорданције енглеског термина *management* и паралелне конкорданције на српском језику



Слика 4. Дистрибуција српског термина менаџмент и енглеског *management* кроз корпус

Дат је и преглед преводних еквивалената термина *management* на српски језик, заједно са бројем конкорданција у којима се дати одређени преводни еквиваленти појављују (Табела 7), стога можемо закључити да одговор на упит садржи различите флективне облике термина на српском језику, а не само облик номинатива (лему) у ком је задат упит.

Могућности примене паралелизованог доменског корпуса за менаџмент које су илустроване изнетим примерима могу бити од користи стручним преводиоцима, истраживачима у области компаративне и контрастивне лингвистике и терминологије, наставницима и студентима енглеског језика струке, као и другим профилима корисника.

Прво, претраживање доменског паралелизованог корпуса за менаџмент путем алата Библиша може помоћи стручним преводиоцима да у процесу превођења разреше термилошке и друге језичке недоумице и пронађу одговарајући преводни еквивалент у контексту језичке употребе, посебно имајући у виду недостатак адекватних

расположивих терминографских и лексикографских ресурса на српском језику.

енгл. <i>management</i>		пример		
Преводни еквивалент	Број конкорданција	извор	енглески	српски
управљање	1431	Барјактаровић et al., 2011, vol. XVI:61, ID: 7.2011.61.1	A very important constituent of the overall bank management process is the implementation of the corporate governance principles.	Врло битан елемент целокупног процеса управљања банком јесте и примена принципа корпоративног управљања .
менаџмент	1089	Митрић et al., 2012, No. 65, ID: 7.2012.65.5	The fields of her scientific and professional interests are related to Accounting and Finance.	Њени главни истраживачки и наставни интереси везани су за област рачуноводства и финансијског менаџмента .
руковођење	14	Michalski G., 2008, vol. XIII:49/50, ID: 7.2008.49-50.12	n95 Operating cycle management should also contribute to realization of this fundamental aim.	n95 Постизању овог основног циља треба да допринесе и руковођење пословним циклусом.
управа	26	Savoiu et al., 2008, vol. XIII:49/50, ID: 7.2008.49-50.1	n16 55 Development of Slovenian selfgovernment in the new public management perspective	n16 55 Развој локалне самоуправе у Словенији у светлу Нове јавне управе
управљачки	83	Петровић С., 2009, vol. XIV:51, ID: 7.2009.51.6	n120 • Organizations are too complicated to be understood by means of one management model...	n120 • Организације су превише компликоване да би могле бити схваћене коришћењем једног управљачког модела...
руководећи	2	Петковић М., 2009, vol. XIV:51, ID: 7.2009.51.1	n11 ... that are presented by the number and the density of communications among organizational parts, management positions or members of a team.	n11 ... које се представљају бројем и густином комуникација између делова организације, руководећих позиција или чланова једног тима.
менаџерски	33	Петковић et al., 2012, No. 64, ID: 7.2012.64.7	... organizational design is a management lever (tool) used to achieve a balance between effectiveness and efficiency...	...организациони дизајн менаџерска полуга (алат), којом се балансира између ефективности и ефикасности...

Табела 7. Преводни еквиваленти енглеског термина *management* са примерима из корпуса

Друго, овај корпус је такође користан ресурс у компаративним и контрастивним проучавањима српског и енглеског језика струке, нпр. у контрастирању карактеристика академског писања на два језика, али и проучавању терминолошке (синонимијске и друге) варијације, недоследности и празнина које се неминовно јављају у српском језику као пасивном примаоцу научно-технолошког трансфера и трансфера

знања из развијених (махом англофоних) земаља.

Треће, за разлику од једнојезичних корпуса који се већ дуже време наилазе на примењујућу у учењу страних језика и креирању наставних материјала, педагошка употреба паралелизованих корпуса представља релативну новину о којој говоре, на пример, [Danielsson and Mahlberg \(2003\)](#), [Granger \(1998\)](#), а за српски језик и [Ristović \(2012\)](#).

Наш паралелизовани доменски корпус пружа могућности директне и индиректне примене у настави. Индиректно, енглески део овог корпуса може се користити као полазна основа за креирање наставних материјала и тестова, али и наставних планова и програма за стручни и академски енглески језик у области менаџмента и организације (нпр. за студенте менаџмента или за специјализоване курсеве енглеског језика), и то ослањањем на листе фреквентних и кључних речи и израза ради креирања тзв. „лексичких силабуса“ ([McEnergy and Xiao, 2011](#)). Директна експлоатација корпуса односи се на учење вођено подацима (енгл. *data-driven learning*), процес у оквиру кога ученици и студенти самостално користе корпус ([Römer, 2011](#)). Другим речима, студенти енглеског језика струке за област менаџмента, уз надзор наставника и адекватну обуку за коришћење корпуса и Библише, могли би самостално да користе корпус ради откривања одређених граматичких, лексичких, дискурзивних и других правила и карактеристика, како би извршили анализу грешака у академском писању (најчешће насталих услед интерференције матерњег, тј. српског језика), али и у процесу превођења стручних текстова са једног језика на други.

4. Закључак

Посебан значај корпуса представљен у претходним поглављима лежи у чињеници да пре његове израде није постојао ниједан други електронски доступан и обележен корпус српског језика из области менаџмента нити сродних дисциплина (економија, маркетинг, организација и друге). Осим тога, електронски корпус може се непрестано надограђивати новим материјалом, чиме он остаје релевантан и савремен. Паралелизовани доменски корпуси радова представљају примарни ресурс за екстракцију термилошких јединица и израду секундарних термилошких ресурса – двојезичких термилошких речника и термилошких база и њихове сталне надоградње новим терминима. Осим тога, овакви корпуси су корисни и у области статистичког и неуронског машинског превођења. Као

ресурс за превођење, паралелизовани корпус који је поравнат на нивоу реченице може се користити за креирање преводилачких меморија и као подршка процесу превођења. Иако је у раду приказана екстракција само на српском језику, овакав ресурс је могуће користити и за двојезичну екстракцију терминологије. Осим наведених примена паралелизованог корпуса за домен менаџмента, треба нагласити многобројне могућности његове примене у другим врстама лингвистичких истраживања, пре свега у оквиру термилолошких, компаративних и контрастивних истраживања језика, студија превођења, учења и подучавања енглеског као страног језика струке, али и у анализи дискурса менаџмента, семантичким, прагматичким и социолингвистичким студијама.

Литература

- Bonhomme, P, TMH Nguyen and S O'ROURKE. "XAlign: l'aligneur de Langue & Dialogue, 2001", ,
- Danielsson, P and M Mahlberg, "There is more to knowing a language than knowing its words: Using parallel texts in the bilingual classroom". *English for Specific Purposes World. Online Journal for Teachers* Vol. 3, no. 6 (2003)
- Erjavec, Tomaž and Nancy Ide. "The MULTTEXT-East Corpus". У *Proceedings of the First International Conference on Language Resources and Evaluation*, 971–74. Citeseer, 1998,
- Erjavec, Tomaž, Ana-Maria Barbu, Ivan Derzhanski, Ludmila Dimitrova, Radovan Garabik et al.. "MULTTEXT-East" 1984" annotated corpus 4.0", (2010)
- Fang, Cheng-yu, "Building a corpus of the English of computer science". *English Language Corpora: Design, Analysis and Exploitation. Amsterdam and Atlanta, GA: Rodopi* (1993): 73–8
- Flowerdew, Lynne. "The argument for using English specialized corpora to understand academic and professional language". *Discourse in the professions: Perspectives from corpus linguistics* (2004): 11–33
- Gledhill, Chris. "The discourse function of collocation in research article introductions". *English for Specific Purposes* Vol. 19, no. 2 (2000): 115–135
- Granger, Sylviane. "The computer learner corpus: A testbed for electronic EFL tools". *Linguistic databases* (1998): 175–88

- Krstev, Cvetana, Ranka Stankovic, Ivan Obradovic and Biljana Lazic. “Terminology Acquisition and Description Using Lexical Resources and Local Grammars.”. У *TIA*, 81–89. 2015
- Krstev C., Vitas D. “Konkordancije paralelizovanih tekstova”. *Zbornik radova XXXVIII konferencije ETRAN, Niš, juni 1994*, 229–230. 1994
- Ljubešić, Nikola, Miquel Esplà-Gomis, Sergio Ortiz Rojas, Filip Klubička and Antonio Toral. “Serbian-English parallel corpus srenWaC 1.0”, 2016
- McEnery, Anthony M and Anita Wilson. *Corpus linguistics: an introduction..* Edinburgh University Press, 2001
- McEnery, Tony and Richard Xiao. “What corpora can offer in language teaching and learning”. *Handbook of research in second language teaching and learning* Vol. 2 (2011): 364–380
- McEnery, Tony, Richard Xiao and Yukio Tono. *Corpus-based language studies: An advanced resource book*. Taylor & Francis, 2006
- Obradović, I, R Stanković and M Utvić. “An integrated environment for development of parallel corpora”. *Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen* (2008): 563–578
- Paumier, Sébastien. “Manuel d’utilisation du logiciel Unitex”. *Université de Marne-la-Vallée*, 2002
- Pazienza, Maria Teresa, Marco Pennacchiotti and Fabio Massimo Zanzotto. “Terminology extraction: an analysis of linguistic and statistical approaches”. У *Knowledge mining*, 255–279. Springer, 2005,
- Pearson, Jennifer. *Terms in context*, Vol. 1. John Benjamins Publishing, 1998
- Ristović, Zoran. “From corpus to classroom: The use of aligned corpora in English language teaching”. *Infoteka* Vol. 13, no. 2 (2012): 52–66
- Roe, Peter Joseph. “The Notion of Difficulty in Scientific Text.”. Докторска дисертација, University of Birmingham, 1977
- Römer, Ute. “Corpus research applications in second language teaching”. *Annual review of applied linguistics* Vol. 31 (2011): 205–225
- Savourel, Y. “TMX 1.4 b Specification, The Localisation Industry Standards Association (LISA)”, 2004
- Siddiqi, Sifatullah and Aditi Sharan. “Keyword and keyphrase extraction techniques: a literature review”. *International Journal of Computer Applications* Vol. 109, no. 2 (2015)
- Sinclair, John. “Corpus and Text: Basic Principles. Wynne, M.(Ed.) Developing Linguistic Corpora: A Guide to Good Practice: 1-16”, , 2005
- Stanković, Ranka, Ivan Obradović, Cvetana Krstev and Duško Vitas. “Production of morphological dictionaries of multi-word units using a multipurpose tool”. У *Proceedings of the Computational Linguistics-Applications Conference*, 77–84. 2011

- Stanković, Ranka, Cvetana Krstev, Ivan Obradović, Biljana Lazić and Aleksandra Trtovac. "Rule-based automatic multi-word term extraction and lemmatization". У *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC*, 507–514. 2016a
- Stanković, Ranka, Cvetana Krstev, Duško Vitas, Nikola Vulović and Olivera Kitanović. "Keyword-based search on bilingual digital libraries". У *Semanitic Keyword-based Search on Structured Data Sources*, 112–123. Springer, 2016b
- Tognini-Bonelli, Elena. *Corpus linguistics at work*, Vol. 6. John Benjamins Publishing, 2001
- Utvić, Miloš. "Annotating the corpus of contemporary Serbian". У *Proceedings of the INFOtheca '12 Conference*, 2011
- Véronis, Jean. *Parallel Text Processing: Alignment and use of translation corpora* Vol. 13. Springer Science & Business Media, 2013
- Vitas, Krstev C., D. "Electronic edition of Serbian translation of Orwell's 1884 aligned with 7 languages by Duško Vitas, Cvetana Krstev". (1998a)
- Vitas, Nenadić G. Krstev C., D. "Electronic edition of Serbian translation of Plato's Republic aligned with 17 languages by Duško Vitas, Goran Nenadić, Cvetana Krstev". (1998b)

Интеграција српског језика у вишејезични речник Prolexbase

УДК 811.163.41'322.2

Цветана Крстев

cvetana@matf.bg.ac.rs

Универзитет у Београду

Филолошки факултет

Србија

Дени Морел

denis.maurel@univ-tours.fr

Универзитет у Туру

Француска

Душко Витас

vitas@matf.bg.ac.rs

Универзитет у Београду

Математички факултет

Србија

САЖЕТАК: Овај рад представља вишејезични речник властитих имена *Prolexbase*, посебно његов српски део. Представљена је комплексност властитих имена у српском језику, посебно одлике које се тичу њиховог превођења: правопис, деривација, флективне промене и дијалекатске варијације. Описује се модел базе *Prolex*, са посебним освртом на решења неопходна за интеграцију у њу српског језика (коришћење два писма, више нивоа деривације, постојање вишеструких облика). На крају се дају неки бројчани подаци који показују српски језик у бази *Prolex*.

КЉУЧНЕ РЕЧИ: властита имена, вишејезична база података, онтологија властитих имена, LMF формат, српски језик, *Prolexbase*.

РАД ПРИМЉЕН: 4. децембар 2018.

РАД ПРИХВАЋЕН: 14. децембар 2018.

1. Мотивација

Као и друге особености језика (неологизми, полилексемске јединице, идиоми и тако даље), властита имена могу бити узрок невероватних грешака. На пример, да ли би *Bush* требало превести на српски са *грм* (биљка) или *Буи* (лично име)? Да ли се *Casablanca* и *White House* односе на исто место? Влада опште уверење да се властита имена не могу преводити. У суштини, преводиоци користе све врсте преводилачких процеса – адаптацију, дослован превод итд. – када преводе текст са изворног на циљни језик (Lecuit et al., 2011).

Властита имена представљају изазов и за обраду природних језика, а посебно за задатке везане за *именоване ентитете*¹. Први задаци везани за именоване ентитете постављени у оквиру конференција о разумевању порука (Message Understanding Conferences MUC-6 и MUC-7) односили су се на попуњавање база података одговорима на питања као што су „кога је терориста напао?“, „где?“, „када?“ или „која фирма је преузела власништво друге фирме?“, „колики је удео?“, „по којој цени?“ и тако даље (Chinchor, 1997). Данас су изазови скоро супротни: ентитете из текста треба повезати са уносима у бази података (Hachey et al., 2013) јер, на пример, властита имена треба да постану једнозначна (видети за пример конференције о анализи текста (Text Analysis Conferences) (McNamee et al., 2010). За овакве задатке се често користи Википедија као и многе друге семантичке базе података: DBpedia (Auer and Lehmann, 2007), GeoNames, YAGO2 (Hoffart et al., 2012), BabelNet (Navigli and Ponzetto, 2012). Ове базе података представљају део система повезаних отворених података (Link Open Data system (LOD)) у коме властита имена заузимају посебно важно место.

Prolexbase је вишејезична релациона база података властитих имена (Maurel, 2008). Циљ Prolex базе је да помогне приликом превођења. Она укључује морфолошке, деривационе и семантичке релације. На пример, ако би требало превести реченицу *Београђанка ми је рекла да је Дунав прелеп* могло би бити корисно да се она прошири на следећи начин: *Женски [флексија] становник града [семантичко проширење] Београда [деривациона релација] у Србији [релација доступности] ми је рекла да је река Дунав [семантичко проширење] прелеп*. Вратићемо се на овај пример на крају овог рада.

Прву верзију Prolex базе која је укључивала осам језика (француски, немачки, енглески, италијански, холандски, пољски, португалски и шпански) је подржао француски пројекат *RNTL-Technolangue Project* (2003-2005). У ствари, у оквиру овог пројекта је конструисан модел базе података, остварена је висока покривеност за француски, док су други језици били слабије заступљени. У исто време је покренут пројекат *Egide Pavle Savic* (2004-2005) чији циљ је био укључивање српског језика у Prolex базу. Укључивање српског тима је било веома важно јер је оно омогућило да се боље разуме сложеност морфологије и деривације у моделу, који је до тада био сувише под утицајем француског

¹ Именовани ентитети се обично дефинишу референцијално или својом јединственошћу.

и енглеског језика. Други проблем је представљало коришћење два писма, ћириличног и латиничног. У овој првој верзији изабрано је незадовољавајуће решење за коришћење два писма: за српски језик су изграђена два блока, један који користи ћирилицу и други који користи латиницу.

Друга верзија Prolex базе коју је подржао пројекат *Hubert Curien Polonium* је донела добру покривеност за енглески и пољски (Savary et al., 2013). Српски део базе је био значајно унапређен у трећој верзији Prolex базе као резултат једномесечне посете проф. Цветане Крстев Универзитету у Туру. Током ове посете побољшана је покривеност српског језика, а раздвојена репрезентација за два писма је спојена у један блок. Такође је припремљен могући опис дијалекатских облика, екавског и ијекавског.

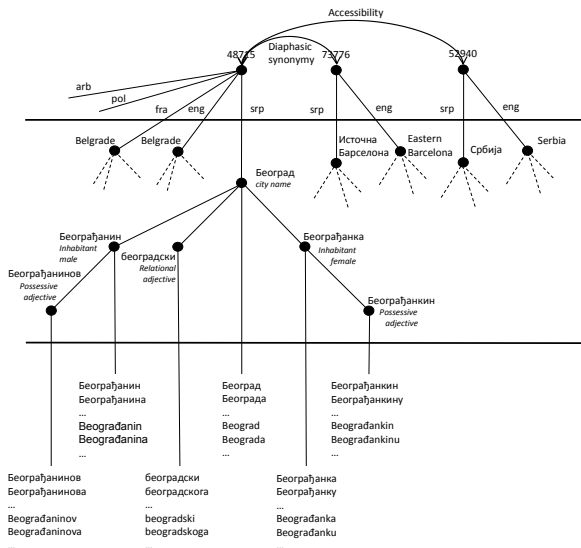
2. База Prolex

2.1 Модел Prolex базе

Пошто је база Prolex вишејезична база података потребан нам је модел који омогућава повезивање различитих појављивање властитих имена у разним језицима. Одлучили смо да дефинишемо лингвистичку класу властитих имена (и облика изведених деривацијом) као онтологију у смислу датом у (Gruber, 1995): „Концептуализација је апстрактан, поједностављен поглед на свет који желимо да представимо с неким циљем... Онтологија је експлицитна спецификација концептуализације“.

Средиште онтологије базе Prolex је *концептуално властито име, стожер*, који представља референта с одређене тачке гледишта. Примери су *Папа Фрањо* и *Хорхе Марио Бергољо* или *Београд* и *Источна Барселона*². Превођење преко стожера (Boitet, 1988) данас није уобичајено иако стожер може да буде профињен у неким језицима више него у другима. На пример, у пројекту Papillon (Mangeot, 2000), стожер за *rice* (*пиринач*) на енглеском има два финија стожера на јапанском, *raw rice* (*сирови пиринач*) и *cooked rice* (*кувани пиринач*). Како за *концептуално властито име* профињавање није потребно овај модел се може користити без проблема. У сваком језику стожер је повезан са јединственим скупом властитих имена, *пролексомом*. Овај скуп садржи властитито име, и по потреби његове псеудониме

² На интернету се *Београд* некад реферише као *Источна Барселона*



Слика 1. Пример: *Београд* у моделу базе Prolex

(алијасе) као и морфосинтаксички изведене облике (видети 2.3). Стожери представљају концептуални ниво модела док пролексеме представљају његов лингвистички ниво. Онтологију допуњавају још два нивоа: метаконцептуални ниво (типови и супертипови) испод кога је ниво примерака (облици власитих имена онако како се јављају у тексту). На слици 1 је модел илустрован на примеру властитог имена *Београд*.

2.2 Формат LMF

База Prolex је отворени ресурс који се користи под лиценцом LGPL-LR³. Формат за размену је инспирисан форматом за лексичко обележавање (Lexical Markup Format (LMF)) (ISO/TC 37/SC 4, 2007). На слици 2 су приказане LMF класе за репрезентацију модела базе Prolex. Овај формат представља избор класа из језгра LMF модела са деловима додатим из LMF проширења (пакети за морфологију, семантику,

³ <http://www.cnrtl.fr/lexiques/prolex/>

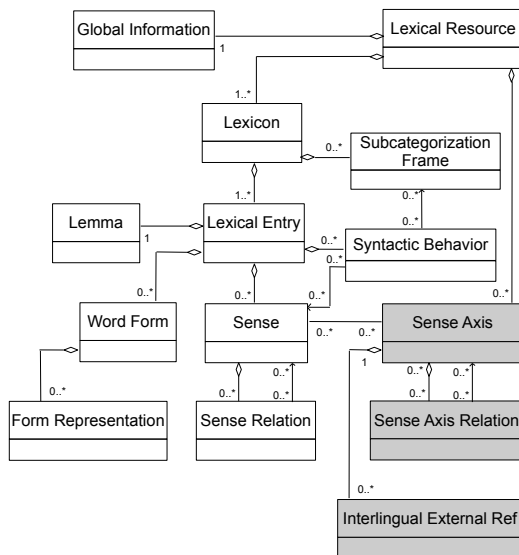
вишејезичну нотацију и синтаксу за потребе обраде природних језика). Вишејезични описи су представљени сивим кућицама. Цео ресурс је представљен класом *Lexical Resource* (лексички ресурс) с којом су повезане неке информације – код језика, писма, карактери које ресурс користи (класа *Global Information* (опште информације)). Ресурс садржи концептуални ниво (класа *Sense Axis* (оса значења)) и лингвистички ниво са више лексикона (класа *Lexicon* (лексикон)) који су једнојезични описи. Један од њих је српски лексикон. Лексички уноси су све леме пролексеме (властита имена, алијаси и деривациони облици) са одговарајућим облицима речи (сви примерци): класе *Lexical Entry* (лексички унос), *Lemma* (лема), *Word Form* (облик речи) и *Form Representation* (репрезентација облика). Ове леме су повезане са значењима која су стожери којима је додељена категорија, као описни придев или присвојни придев од имена мушког рода (класе *Sense* (значење) и *Sense Relation* релација значења)). Ови стожери су дефинисани у класи *Senses Axis* која спада у вишејезични део ресурса. Класа *Sense Axis Relation* (Релација осе значења) представља релацију између концептуалних властитих имена док класа *Interlingual External Ref* (међујезичка спољашња референца) представља типологије. Уочавају се и неке информације о класификационом контексту властитих имена (класа *Subcategorization Frame* (оквир за субкатегоризацију) и неке изузетне колокације (класа *Syntactic Behavior* (синтаксично понашање)). Ове класе се још увек не користе у српском блоку, а неки примери би били коришћење различитих предлога уз властита имена, на пример, *Србија је на Балкану и у Европи*.

2.3 Релације

Велики део базе Prolex чине релације између стожера (језички независне релације): синонимија, меронимија и доступност.

Релација синонимије, или прецизније, квази-синонимије или блискозначности, је релација између два стожера који се реферишу на исти референт за који постоје различите тачке гледања. Преводаца треба да изабере исправну тачку гледита, што није увек могуће. Разликујемо три различите тачке гледишта које одређују три дијасистематска својства (Coseriu, 1998):

- Дијахроне варијације зависе од времена: *Савезна Република Југославија* према *Државна Заједница Србија и Црна Гора*;



Слика 2. LMF шема базе Prolex

- Дијастратске варијације зависе од социокултурне стратификације: *Јосип Броз* према *Тито*;
- Дијафазне варијације зависе од употребе: *Београд* према *Источна Барселона*,

Релација меронимије, или партитивна релација, је релација укључивања. Пример укључивања за географска имена је: *Србија* је на *Балкану* који је део *Европе*, а пример временске релације је: бомбардовање Београда 6. априла 1941. године је део (догодио се током) Другог светског рата. Ову врсту релације проширујемо на друге домене, као што су економија, националности и тако даље.

Релација доступности (Ariel, 1990), или релација асоцијативности, значи да је властито име доступно преко неког другог властитог имена. У речницима, властита имена, за разлику од заједничких именица, немају дефиниције – оне су замењене релацијом са неким познатијим именом. С тога су ове релације ретко симетричне: у речницима се тако може прочитати да је *Арон* брат *Мојсија*, али се *Мојсије* не представља као брат *Арона* већ као предводник Јевреја. Према томе, *Арон* је доступан преко *Мојсија*, док је *Мојсије* доступан преко одговарајуће јеврејске приче. Разликујемо 12 оваквих релација:

- рођачке везе: *Арон* је брат *Мојсија*;
- главни град: *Београд* је главни град *Србије*;
- вођа: *Тито* је био политички вођа *Југославија*;
- оснивач: *Растко Немањић* је основао *Српску православну цркву*;
- следбеник: *Петар* је ученик *Исуса*;
- креатор: *Бранислав Нушић* је написао комедију *Госпођа министарка*;
- менаџер: *Ранко Жеравица* је био српски кошаркашки тренер који је водио *Југословенску кошаркашку репрезентацију*;
- становник: *Александар Вучић* борави у *Новом двору*;
- наследник: *Кнез Михаило Обреновић* је наследио *Кнеза Милоша Обреновића*;
- седиште: Седиште корпорације *Montinvest Beograd* се налази у *Београду*;
- супарник: *Партизан* је фудбалски ривал *Црвене звезде*;
- друг: *Мирко* је *Славков* најбољи друг и саборац.⁴

Језички зависне релације су фреквенција, је-алијас-од, је-дериват-од, колокација, контекст и епонимија.

Пролексема је скуп свих лема које су семантички повезане с властитим именом у једном језику. На пример, пролексеми *Београда* чини: Београд, београдски, Београђанин, Београђанка, Београђанинов, Београђанкин као што се види на слици 1. Три главне релације на језички зависном нивоу су фреквенција, је-алијас-од и је-дериват-од. Фреквенција која указује колико је име познато може да има три вредности: често се користи, не користи се често и ретко се користи. Данас се ова фреквенција може израчунати из повезаних отворених

⁴ Главни ликови из веома популарног југословенског стрипа о два партизанска дечака-курира.

података (LOD), углавном из Википедије (Elashter and Maurel, 2016). Алијаси су различите варијације властитог имена: краћи облици, скраћенице, акроними, разлике у писању, алтернативна транскрипција, дијатопски квази-синоними и објашњења. Што се тиче деривације, узимају се узбир само леме изведене морфосемантичким средствима (и леме изведене из њих). На пример, глагол *пастеризовати* који се односи на процес делимичне стерилизације је изведен из имена *Пастер* (Pasteur) али није с њим семантички повезан.

Релације колокације и контекста тичу се локалног коришћења властитих имена. У неким језицима, на пример у француском, испред имена државе често долази члан мушког или женског рода, при чему не постоји посебан разлог за избор рода: тако се каже, на пример, *la* (женски род) *France* и *le* (мушки род) *Montenegro*. Контекст је релација између властитог имена и типичних речи које га окружују. Контекст може да се односи на класификацију или доступност. MacDonald (1990) назива *екстерна структура* властитог имена класификациони контекст који је проширење именичке фразе (главни град, краљ, тренер, итд.). Класификациони контекст може да буде користан приликом превођења. На пример, *Сава* се преводи на енглески са *Sava River*. Контекст доступности је именичка фраза која уводи релацију доступности између два стојера. Она се може разумети као врста објашњења за властито име његовим повезивањем са неким добро познатим властитим именом. На пример, *Београд* се може превести са *Београд, главни град Србије*.

Релација епонимије се разликује од осталих релација: она говори да се превод не односи на властито име већ на заједничку именицу (антономазија), као у примеру *жилет* који у српском означава све врсте општрица за бријање, а не само оне које производи фирма Gillette, или на термине, какви су *Паркинсонова болест* или *Питагорина теорема*, или, пак, на идиоматску фразу каква је *све ми је равно до Косова* са значењем *сасвим ми је свеједно*.

2.4 Типологија

Метаконцептуални ниво се бави постојањем концепта и типологијом властитих имена.

Постојање концепта дели властита имена у три групе: (а) историјска која постоје или су постојала, као *Београд*; (б) религијска, чије постојање зависи од веровања, на пример *Архангел Михаил*, или (в) измишљена од стране аутора. У општем случају имена која спадају у последње две

наведене групе често захтевају да буду преведена, као што се *Snow White* са енглеског преводи на српски са *Снежана*.

Циљ типологије базе Prolex је класификација властитих имена. Дефинисали смо четири велике класе које одговарају примарном семантичком својству и назвали их *супертипови*): особе (*антропоними*), локације (*топоними*), конкретне ствари (*ергоними* – артефакти и имена дела) и догађаји (*прагмоними*). Дефинисали смо укупно тридесет типова који су представљени у табели 1. Ову типологију дефинише примарна релација хиперонимије успостављена између стожера и типа. Допунили смо је другом релацијом, секундарно релацијом хиперонимије, која је релација метонимије између типова, као што се види у табели 2.

властито име							
антропоним			ергоним	прагмоним	топоним		
индивидуално	колективно				астроним	зграда	територија
	друштво	ансамбл					
позната лич.	династија	друштво	објекат	несрећа	астроним	држава	
патроним	етноним	ансамбл	дело	празник	зграда	регион	
лично име		фирма	мисао	историја	геоним	наднационално	
псеудо-		институција	производ	манифестација	хидроним		
антропоним		организација	пловило	метеорологија	град		
					пут		

Табела 1. Типологија базе Prolex – примарна релација хиперонимије

3. Властита имена у српском језику

3.1 Писма

У Србији је употреба ћириличног писма прописана законом (Zakon, 2010, члан 1), док је коришћење латиничног писма дозвољено у посебним случајевима (саобраћајни знаци, називи улица, итд). Међутим, латинично писмо је из историјских и других разлога у широкој употреби и у Правопису српског језика (Пешикан et al., 1993, чланови 1–6) је дефинисано као равноправно са ћириличним писмом. Српско писмо, било ћирилично или латинично, има 30 слова, а између ова два скупа

типови	секундарни хипероними
држава регион наднационалан територија	колективни антропоним
град	колективни антропоним ергоним
зграда пут празник историја манифестација	ергоним
друштво ансамбл фирма група институција организација	ергоним топоним
пловило	топоним

Табела 2. Секундарна релација хиперонимије

постоји 1-1 пресликавање као што је приказано у Табели 3. Редослед слова у ћирилици и латиници је различит; слова у Табели 3 су приказана у ћириличном редоследу. У латиничном писму за српски се не користе нека од 26 слова енглеског алфабета⁵ и то Q, W, X и Y, али се нека слова користе са дијакритичким знацима: Č, Ć, Đ, Š и Ž, а нека су представљена као диграфи, односно, комбинације других слова: Lj, Nj и Dž. Диграфи се у електронским тектовима обично приказују помоћу кодова слова која га чине премда су у Unicode били уведени посебни кодови за ове симболе⁶. Треба имати на уму да велика слова диграфа Љ, њ и дж могу бити представљена на два начина: у првом је само прво слово диграфа велико слово – Lj, Nj и Dž – а у другом су оба слова диграфа велика слова – LJ, NJ и DŽ. Овај други начин се користи када

⁵ Слова описана ASCII кодом.

⁶ Видети кодну страну [Unicode Latin Extended-B](#)

је цела реч (или дужи текст) записан само великим словима. Ово је такође изражено у Unicode-у тако што за ове две репрезентације постоје различити кодови.

ћирилица	а А	ђ Ђ	ј Ј	н Н	с С	х Х
латиница	a A	đ Đ	j J	n N	s S	h H
ћирилица	б Б	е Е	к К	њ Њ	т Т	џ Џ
латиница	b B	e E	k K	nj Nj	t T	c C
ћирилица	в В	ж Ж	л Л	о О	ћ Ћ	ч Ч
латиница	v V	ž Ž	l L	o O	ć Ć	č Č
ћирилица	г Г	з З	љ Љ	п П	у У	џ Џ
латиница	g G	z Z	lj Lj	p P	u U	dž Dž
ћирилица	д Д	и И	м М	р Р	ф Ф	ш Ш
латиница	d D	i I	m M	r R	f F	š Š

Табела 3. Српско ћирилично и латинично писмо – редослед ћирилице одозго на доле и слева у десно.

3.2 Имена страног порекла

Властита имена страног порекла се, по правилу, у српском не пишу у изворном облику него се транскрибују. Ово начело се подједнако примењује како на лична тако и на геополитичка имена. Правопис српског језика ([Пешикан et al., 1993](#), чланови 101–180) допушта употребу изворног облика у текстовима на српском записаним латиничним писмом, али се ово ретко користи у пракси. Један од разлога је што транскрибовање поједностављује објављивање текста на оба писма и омогућава аутоматску конверзију текста из једног у друго писмо.⁷

У српском се праописна или уобичајена транскрипција користи како би се изговор у изворном језику прилагодио српском фонолошком

⁷ На пример, сви чланци на српској Википедији се могу прегледати и у ћирилично и у латиничном писму – видети, на пример, [Страницу Википедије о правописној транскрипцији у српском](#). Исту могућност конверзије пружају поједини новински портали, као на пример, [Politika](#).

систему. Правопис српског језика (Пешикан et al., 1993, articles 101–180) наводи транскрипциона правила за 27 језика укључујући ту и латински, стари и савремени грчки, јапански и кинески. Међутим, велики број властитих имена се не уклапа у ова правила, углавном због тога што је друкчији изговор прихваћен одавно или што друкчији облик више одговара српском језику и његовим морфолошким својствима. Неколико примера за ово је наведено у Правопису за географска имена: *Москва* (уместо *Масква*), *Волгоград* (уместо *Валгаград*), *Њујорк* (уместо *Њујок*) и *Лајпциг* (уместо *Лајпцих*) као и за лична имена: *Ганди* (уместо *Гандхи*) и *Стриндберг* (umesto *Стриндберј*). За нека страна географска имена, српско име не одговара ни изворном изговору ни транскрипцији као у примеру *Беч* за *Vienna*.

Вишечлана географска имена се по правилу транскрибују као полилексичка имена, као у примерима *Њу Хемпшир* (*New Hampshire*) и *Солт Лејк Сити* (*Salt Lake City*). Постоје ипак изузеци у односу на ово правило као у случају имена *Порторико* (*Puerto Rico*). Страна вишечлана географска имена које садрже као конституенте једну или више заједничких речи су понекада преведена, понекада делимично, а некада остају непреведена. На пример, *Rocky Mountains* се преводи као *Стеновите планине*, док је *Long Island* транскрибован као *Лонг Ајланд*. Исте заједничке речи су понекада преведене, а понекада само транскрибоване као у случају *Нови Јужни Велс* (*New South Wales*) према *Њу Делхи* (*New Delhi*).

За поједина страна имена у српском понекада постоје вишеструки облици као у примеру *Ком д'Ивоар* што представља транскрибовано име у званичној употреби за *Côte d'Ivoire* док његов преведени облик, *Обала слоноваче*, преовлађује у свакодневној употреби. Ово је често и у случајевима имена места са мешовитим становништвом као *Целовец* и *Клагенфурт* (*Klagenfurt*) (град у Аустрији) или оних локација чија су имена промењена из политичких разлога као што је то *Град Хо Ши Мин* (*Ho Chi Minh City*) уместо ранијег *Сажон* (*Saigon*).

Транскрипциона правила није увек јдноставно применити тако да су објављени додатни приручници који могу помоћи приликом писања страних имена као што су то *The transcription dictionary of English personal names* (Prčić, 1992) и *The English-Serbian dictionary of geographic names* (Prčić, 2004). Нажалост, информације у овим приручницима су понекада у супротности са Правописом: на пример, у (Prčić, 2004) транскрипција за *Rio de Janeiro* је *Рио де Жанејро* док Правопис

за исто име предлаже *Рио де Жанеиро* које није сасвим сагласно са транскрипционим правлима за португалски језик, али је одомаћено.

Имена организација показују особености у односу на друга властита имена. Она се чешће од других користе у оригиналном облику, а посебно акроними као што је *IBM* или *FBI*. Поред оваквих случајева, имена организација могу бити транскрибована као *Мајкрософт* (Microsoft) или преведена као *Организација за економску сарадњу и развој* (Organization for Economic Cooperation and Development). Штавише, за исту организацију су у употреби и транскрибовано и преведено име као у примеру *Британска телевизијска мрежа* и (ретко) *Бритиш бродкастинг корпорација* (British Broadcasting Corporation). Одговарајући акроними могу бити као у оригиналу: *BBC* или према изговору *Би-Би-Си* (Krstev et al., 2015).

3.3 Деривација

	Име	Становник Присвојни придев	Становница Присвојни придев	Придев
континент	Европа	Европљанин Европљанинов	Европљанка Европљанкин	европски
надрегион	Балкан	Балканац Балканчев	Балканка Балканкин	балкански
земља	Француска	Француз Французов	Францускиња Францускињин	француски
регион	Прованса	Провансалац Провансалчев	Провансалка Провансалкин	провансалски
град	Београд	Београђанин Београђанинов	Београђанка Београђанкин	београдски
део града	Дорћол	Дорћолац Дорћолчев	Дорћолка Дорћолкин	дорћолски

Табела 4. Имена становника и придеви изведени из појединих типова топонима у српском језику

Из већине географских властитих имена се могу извести друге именице и придеви⁸.

Имена становника или демоними се изводе из различитих географских имена као што су имена за континенте, надрегионе, земље, регионе, градове или делове града као што је приказано у табели 4. За нека имена овог типа није могуће извести име за становника као у случају топонима *Осло* (*Oslo*) и тада се користи израз *становник Осла*. Ако се из топонима може извести име становника, тада се, по правилу, може извести и име становнице, а за оба ова изведена облика се могу извести облици присвојних придева, као и релациони придев који се односи на полазни топоним. За *Осло* се ни такав придев не може извести. С друге стране, од неких имена становника се могу извести и други придеви као у примеру *Парижанин* → *Парижанинов* (који припада становнику Париза) → *парижански* (који се на начин понашања Парижанина) (насупротив *париски* (који се односи на Париз)). У појединим случајевима се могу извести и деминутиви из имена становника као, на пример, *Српче* и *Српчић*, који су деминутивни облици од *Србин*, а обично се односе на децу.

У појединим случајевима могу се извести два или чак три имена за мушког становника што даље даје вишеструка имена становница и облике присвојног и релационог придева. Пример су:

- двоструко име изведено из *Кореја*⁹:
 - *Корејац* (м), *Корејка* (ж), *Корејчев* (м присв.пр.), *Корејкин* (ж присв.пр.), *корејски* (пр.);
 - *Кореанац* (м), *Кореанка* (ж), *Кореанчев* (м присв.пр.), *Кореанкин* (ж присв.пр.), *кореански* (пр.);
- троструко име изведено из *Париз* (*Paris*)¹⁰:
 - *Парижанин* (м), *Парижанка* (ж), *Парижанинов* (м присв.пр.), *Парижанкин* (ж присв.пр.), *париски* и *паришки* (пр.);
 - *Парижлија* (м), *Парижлијка* (ж), *Парижлијин* (м присв.пр.), *Парижлијкин* (f poss.);
 - *Паризлија* (м), *Паризлијка* (ж), *Паризлијин* (м присв.пр.), *Паризлијкин* (ж присв.пр.).

⁸ Овде нећемо посматрати глаголе изведене из географских властитих имена као што је *пофранцузити се* (*понашати се као Француз или Францускиња*) како је објашњено у одељку 2.3.

⁹ Ови примери су према (Стијовић, 2016).

¹⁰ Ови примери су према (Стевановић, 1967).

За поједина вишечлана географска имена демоними и придружени придеви се изводе или слагањем чланова или употребом само једног од њих. Резултат је у оба случаја монолекемска реч као што је показано у табели 5¹¹. Међутим, за бројна вишечлана географска имена се не могу извести демоними и одговарајући придеви.

Српско и изворно име	Становник и становница	Придев
Кабо Верде (Cabo Verde)	Кабоверђанин Кабоверђанка	кабовердски
Буркина Фасо (Burkina Faso)	Буркинац Буркинка	буркински
Тринидад и Тобаго (Trinidad and Tobago)	становник Тринидада и Тобага становница Тринидада и Тобага	<i>описно</i>
Нови Сад	Новосађанин Новосађанка	новосадски
Бачко Ново Село	становник Бачког Новог Села становница Бачког Новог Села	<i>описно</i>

Табела 5. Имена становника и придеви изведени из вишечланих имане топонима на српском

Придеви се изводе и из других типова географских имена као што су хидроними или ороними. Пример за хидрониме су *дунавски* изведен из *Дунав* или *сенски* из *Сена*, а за орониме *алпски* из *Алпи*, *копаонички* из *Копаник*. За неке хидрониме и орониме се релациони придев не може извести као у примеру *Волга*. Ако се релациони придеви могу извести из вишечланих хидронима или оронима, онда су они монолексемске речи као у примерима *великоморавски* полазећи од *Велика Морава* и *старопланински* од *Стара Планина*.

Присвојни придеви могу бити изведени и из личних имена: имена, презимена и надимака. На пример, присвојни придев изведен из сваког од делова имена *Иво Лола Рибар* би били *Ивов*, *Лолин* и *Рибаров*. Из имена познатих личности се могу извести нове иенице и придеви.

¹¹ Примери су према (Стијовић, 2016).

На пример, из имена филозофа *Карла Маркса* генеришу се у српском изведенице: *марксизам* за доктрину, *марксиста* и *марксисткиња* за следбенике *марксизма*, *марксологија* за научну дисциплину, *марксолог* и *марксолошкиња* за истраживаче који пручавају *марксологију*, затим придеви *марксистички* и *марксолошки*. Многи од овако изведених придева и именица могу бити префиксирани нпр, префиксима *анти-*, *нео-*, *пост-*, итд. (Vitas and Krstev, 2013). Оваке изведенице нису узете у обзир у бази Prolex, као што је напоменуто у одељку 2.3.

Присвојни придеви се могу извести обично и из монолексемских имена организација. На пример, *Мајкрософтов* је присвојни придев изведен из *Мајкрософт*. Присвојни придеви се употребљавају и код акронима имена организација – у таквим случајевима се деривациони суфикс додаје на акроним полсе цртице, нпр. *IBM-ов*.

3.4 Граматичка својства

Властита имена у српском, као и именице и придеви изведени из њих, имају иста флективна својства као и заједничке именице и придеви.

Род географских имена, топонима, оронима и хидронима, може бити мушки, женски или средњи, а имена су променљива по падежима (седам различитих падежа). Број имена је непромењив и може бити или сингулар или плурал. Примери су дати у табели 6.

Географска имена су по правилу неаниматна премда постоји неколико збуњујућих примера: неколико градова у Србији носе имена знаменитих личности као што су *Јаша Томић* и *Алекса Шантић*. Ако се ова имена посматрају као неаниматна, онда би реченица *I travel to Jaša Tomić* била на српском *Путујем у Јаша Томић* која је неприхватљива, јер је Јаша Томић, као особа, обележен као аниматан¹².

Демоними изведени из географских имена су мушког рода (за становнике) и женског рода. Облик придева изведених из географских имена се мења према падежу, броју, роду и аниматности. Треба напоменути да присвојни придеви немају облике компаратива и

¹² Облик акузатива сингулара именица мушког рода зависи од аниматности: за неаниматне именице је он једнак облику номинатива, а за аниматне облику генитива. У овом примеру, предлог *у* захтева акузатив који је за (неаниматно) *Јаша Томић* једнак номинативу, док је за (аниматно) *Јаша Томић* једнак генитиву *Јашу Томића* (упоредити са реченицом *Милица се заљубила у Јашу Томића*).

тип	изворни или енглески	име	број	род
топоним ороним хидроним	Belgrade Olympos Danube	Београд Олимп Дунав	сингулар	мушки
топоним ороним хидроним	Karlovc Alpes Dardanelles	Карловци Алпи Дарданели	плурал	
топоним ороним хидроним	Athens Aconcagua Seine	Атина Аконкагва Сена	сингулар	женски
топоним ороним хидроним	Budějovice Divčibare Plitvice	Будјевице Дивчибаре Плитвице	плурал	
топоним ороним хидроним	Valjevo Pohorje Oranjerivier	Ваљево Похорје Орање	сингулар	средњи
топоним	Kaštela	Каштела	плурал	

Табела 6. Географска имена у српском различитог рода и броја

суперлатива, као ни релациони придеви осим изузетно као у примеру *Војводина је најевропскији део Србије*.

Српска лична имена и надимци могу бити мушког или женског рода, у сингулару су и мењају се по падежима. Српска презимена су мушког рода и, у општем случају се мењају и у броју и про падежима. Нека презимена, углавном страног порекла, се не мењају према броју због морфолошких запрека. Сложена правила слагања се примењују на пуно име у српском чији облик зависи од рода личног имена и редоследа личног имена и презимена – једно правило је да је презиме, када припада женској особи, непроменљиво (Gucul-Milojević, 2010). Женске особе се понекада обележавају обликом присвојног придева презимена у женском роду или именицом женског рода изведеном из придева моцијом рода. Примери су дати у табели 7.

Имена организација се мењају по падежима али се њихов род и број не мењају и зависе, у општем случају, од морфолошких

Облик	Презиме	Женски облик
номинатив сингулара	<i>Петровић</i>	<i>Петровићка</i> <i>Петровићева</i>
генитив сингулара		<i>Петровићеве</i>
номинатив плурала	<i>Петровићи</i>	<i>Петровићке</i> <i>Петровићеве</i>
Облик	Пуно мушко име)	Пуно женско име
номинатив сингулара	<i>Петар Петровић</i>	<i>Зорка Петровић</i> <i>Зорка Петровићка</i> <i>Зорка Петровићева</i>
генитив сингулара	<i>Петра Петровића</i>	<i>Зорке Петровић</i> <i>Зорке Петровићке</i> <i>Зорке Петровићеве</i>

Табела 7. Мушка и женска лична имена и њихова промена

својстава њиховог имена у случају моноксемских речи, а од свосјтава носеће речи у случају вишечланих речи. На пример, *Мајкрософт* је мушког, док је *Сорбона* женског рода. Вишечлано име организације **Универзитет** у Београду је мушког, *Београдска аутобуска станица* женског, а **Удружење спортских новинара Београда** средњег рода. Имена организација **Лекари без граница** и **Међународне мировне снаге** имају само облик множине ¹³.

3.5 Дијалекти

У српском се користе две стандардне варијанте изговора, екавска и ијекавска. Оне се разликују по рефлеку протословенске фонеме *јат*: у екавској варијанти она се замењује најчешће са *е*, док се у ијекавској замењује слоговима *ије* или *је*.

Ове варијанте немају много утицаја на властита имена јер их већина и не садржи рефлекс фонеме *јат*. Када је то ипак случај, име се обично

¹³ Носећа реч у овим вишечланим именима организација је наглашена црним слогом.

користи у само једној варијанти. На пример, у именима градова *Ријека* и *Ријека Црнојевића* заједничка именица се користи само у ијекавској варијанти – *ријека*, а не *река*). С друге стране, женско лично име настало из заједничке именице *вера/вјера* – постоји и у екавској *Вера* и у ијекавској варијанти *Вјера*. Међутим, лично име у једној варијанти изговора се неће мењати ако се нађе у тексту записаном у другој варијанти, то јест, оно је непроменљиво.

У полилексемским именима организација појављују се многе заједничке именице које могу бити у једној или другој варијанти изговора. Ове варијанте се онда одражавају и у именима организација у зависности од тога коју варијанту користи текст у коме се појављују, на пример, екавска варијанта *Светска банка* или ијекавска варијанта *Свјетска банка*.

4. Остварени резултати

4.1 Допринос српског језика моделу базе Prolex

Као што смо рекли у одељку 1., укључивање српског језика довело је до развоја бољег модела базе Prolex. Сарадња између истраживачких група са Универзитета у Туру и Универзитета у Београду је била плодна по много чему, али ми ћемо истаћи две најважније ствари: релација деривације и репрезентација облика.

Релација деривације. У одељку 3.3 смо представили сложеност деривационих правила српског језика, као што су могућности квази-систематског извођења из имена људи, као и из геополитичких имена (описни придеви и имена становника). На пример, (видети слику 1), из имена града *Београд* генерише се, као што је случај и у многим другим језицима *Београђанин* (мушки становник *Београда*), док се из *Београђанин* генерише *Београђанинов* (присвојни придев од мушког становника *Београда*). У енглеском и француском постоји само један ниво извођења: *Belgrade/Belgradian* у енглеском и *Belgrade/Belgradois* у француском. Први модел базе података није предвиђао релацију табеле *Derivative* са самом собом. Додали смо ту релацију у касније моделе и онда схватили да таква релација постоји и у француском: на пример, име награде, каква је Нобелова награда, дозвољава да се квази-систематски изведе глагол са значењем *доделити награду*, у овом примеру *nobeliser*,

док се из таквих глагола регуларно стварају други деривати, као *no-belisable* (особа која је могући изабраник Нобеловог комитета), и тако даље.

Репрезентација облика. Властита имена смештамо у моделу базе података Prolex у две табеле, *Prolexeme* (најдужи облик имена) и *Alias* (други облици). Међутим, у LMF репрезентацији (видети слику 2) ова разлика нестаје јер су сви алијаси равноправни уноси које повезује значење.

Поставило се питање да ли су име записано ћирилицом и исто име записано латиницом алијаси или не. Могло би изгледати чудно да је реч алјас самој себи! Али, ипак није тако. Прво смо размотрили могућност да дефинишемо две пролексеме за српски језик (за ћирилицу и латиницу), али овакво решење нарушава услов јединствености пројекције стожера за одређени језик. С тога смо прихватили друго решење које дефинише два лексикона, српски ћирилични лексикон и српски латинични лексикон. Коначно, у трећој верзији базе Prolex произведеној у Универзитету у Туру систематски смо додали под *Word Form* једну или више *Form Representations*. На пример, за *Београд* сада имамо:

```
<LexicalEntry partOfSpeech="noun">
  <Lemma>Београд</Lemma>
  <WordForm grammaticalGender="masculine"
    grammaticalNumber="singular"
    grammaticalCase="nominative"
    grammaticalAnimacy="nonAnimate">
    <FormRepresentation script="cyr1">
      Београд
    </FormRepresentation>
    <FormRepresentation script="latn">
      Beograd
    </FormRepresentation>
  </WordForm>
  ...
</LexicalEntry>
```

Пошто је такав избор начињен, додали смо за неке уносе и разликовање између екавског и ијекавског изговора (видети одељак 3.5), за шта смо користили исту LMF репрезентацију:

```

<LexicalEntry partOfSpeech="noun">
  <Lemma>Немачка</Lemma>
  <WordForm grammaticalGender="feminine"
    grammaticalNumber="singular"
    grammaticalCase="nominative"
    grammaticalAnimacy="nonAnimate">
    <FormRepresentation script="cyrл">
      Немачка
    </FormRepresentation>
    <FormRepresentation script="cyrл"
      geographicalVariant="ekavsk">
      Немачка
    </FormRepresentation>
    <FormRepresentation script="cyrл"
      geographicalVariant="ijekavsk">
      Нјемачка
    </FormRepresentation>
    <FormRepresentation script="cyrл"
      geographicalVariant="ijekavsk">
      Њемачка
    </FormRepresentation>
    <FormRepresentation script="latn">
      Nemačka
    </FormRepresentation>
  </WordForm>
  ...
</LexicalEntry>

```

Коначно, користили смо исти концепт *form representation* за неке различите варијанте писања, на пример у горњем примеру, *Њемачка* и *Нјемачка* (други начин се ретко користи), али и за неке друге случајеве, као што су разлике у транскрипцији, *Рио де Жанејро* и *Рио де Жанеиро* (видети одељак 3.2), или различити облици за исти скуп вредности граматичких категорија – презиме *Чехов* има три варијанте облика датива у сингулару: *Чехову*, *Чеховом* и *Чеховому*. Овај приступ смо применили и на друге језике елиминишући тако категорију алијаса *Variant*.

4.2 Имплементација базе Prolex

У табели 8 су дати неки подаци о имплементацији српског језика. Ручно смо унели пролексеме и повезали их са стожером из одабраног скупа француских пролексема, а додали смо и неке алијасе. Потом смо аутоматски генерисали имена изведена деривацијом и, наравно, све флективне облике пролексема, алијаса и изведених имена.

Serbian prolexemes	8 526
Serbian aliases	21
Serbian derivatives	920
Serbian instances	108 325
Serbian pivot relations	29 567

Табела 8. Имплементација српског језика

Овим бројевима бисмо могли да додамо и невероватан број примерака имена изведених из *Београд*.¹⁴ Да бисмо употпунили слику 1 са свим примерцима, морали би да додамо 626 облика...

Ако се вратимо на пример из одељка 1.: *Београђанка ми је рекла да је Дунав прелеп* сада добијамо:

Београђанка

female inhabitant (категорија деривације)

Belgrade (пролексема)

city (класификациони контекст)

Serbia (доступност)

capital (контекст доступности)

→ The female inhabitant of the city of Belgrade, capital of Serbia

ми је рекла да је

→ has told me that

Дунав

river (класификациони контекст)

→ the Danube River

прелеп

¹⁴ У поређењу с енглеским, па чак и француским!

→ is splendid

4.3 Закључак

Показали смо да је сложеност морфологије српског језика знатно допринела пројекту вишејезичног речника Prolex. Побољшања која су учињена да би се српски језик укључио у базу показала су се корисна и за друге језике из базе Prolex. Ова побољшања се посебно односе на решавање проблема деривације и на репрезентацију вишеструких облика. То је потврђено када су у базу укључени и неевропски језици, попут арапског, јер је интерна структура базе могла да их моделира. Овај рад је такође потврдио колико је важно да се у лингвистичке вишејезичне пројекте укључе не само блиски већ и разноврсни језици.

Захвалност

Аутори се захваљују Универзитету у Туру што је омогућио ово истраживање финансирањем једномесечног боравка проф. Цветана Крстев Универзитету. Аутори се такође захваљују фонду *Egide Pavle Savić* који је подржао пројекат којим је покренута сарадња између универзитета у Туру и Београду 2004. Део истраживања је подржало Министарство просвете, науке и технолошког развоја Републике Србије кроз пројекте 138006 и III47003.

Литература

- Ariel, M. *Accessing Noun Phrases Antecedents*, 1990
- Auer, S. and J. Lehmann. “What have Innsbruck and Leipzig in common? Extracting Semantics from Wiki Content”. У *ESWC 2007*, no. 4519, LNCS, 503–517. 2007
- Boitet, C. *Pros and cons of the pivot and transfer approaches to multilingual machine translation*, 93–106. 1988
- Chinchor, N. “Muc-7 Named Entity Task Definition”, 1997, URL http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/ne_task.html
- Coseriu, E. “Le double problème des unités dia-s”. У *Les Cahiers dia. Etudes sur la diachronie et la variation linguistique*, Université de Gent, Belgique, Vol. 1, 9–16. 1998

- Elashter, Mouna and Denis Maurel, “Estimer la notoriété d’un nom propre via Wikipedia”, *Y TALN 2016. Paris*, 2016, URL <https://jep-taln2016.limsi.fr/actes/>
- Gruber, T. R.. “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”. *Int. Journal of Human-Computer Studies* Vol. 43 (1995): 907–928
- Gucul-Milojević, Sandra. “Personal Names in Information Extraction”. *IN-FOtheca* Vol. 11, no. 1 (2010): 53a–63a
- Hachey, B, W Radford, J Nothman, M Honnibal and R Curran, J. “Evaluating entity linking with Wikipedia”, *Y Artificial Intelligence*, 194, 130–150. 2013
- Hoffart, J., F. M. Suchanek, K. Berberich and G. Weikum. “YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia”. *Artificial Intelligence Journal, Special Issue on Artificial Intelligence, Wikipedia and Semi-Structured Resources* (2012)
- ISO/TC 37/SC 4. *Language resource management - Lexical markup framework (LMF)*, 2007. <http://lirics.loria.fr/documents.html>
- Krstev, Cvetana, Duško Vitas and Ranka Stanković. “A Lexical Approach to Acronyms and their Definitions”. *Y Proceedings of 7th Language & Technology Conference, November 27–29, 2015, Poznań, Poland*. 2015
- Lecuit, Emeline, Denis Maurel and Duško Vitas. “A tagged and aligned corpus for the study of Proper Names in translation”. *Y Workshop Annotation and exploitation of parallel corpora, International Conference Recent advance in Natural Language Processing (RANLP 2011)*, 11–18. 2011, URL <http://aclweb.org/anthology/W11-43>
- MacDonald, D. *Internal and external evidence in the identification and semantic categorisation of Proper Names*, 21–39. 1990
- Mangeot, M. “Papillon Lexical Database Project: Monolingual Dictionaries & Interlingual Links”. *Y 7th Workshop on Advanced Information Network and System, Kasetsart University, Bangkok, Thailand*. 2000
- Maurel, D. “Prolexbase: A Multilingual relational Lexical Database of Proper Names”. *Y LREC 2008*, 334–338. 2008
- McNamee, P., H. T. Dang, H. Simpson, P. Schone and S. M. Strassel. “An evaluation of technologies for knowledge base population”. *Y LREC 2010*, 369–372. 2010
- Navigli, Roberto and Simone Paolo Ponzetto. “BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network”. *Artificial Intelligence* Vol. 193 (2012): 217–250
- Prčić, Tvrtko. *Transkripcioni rečnik engleskih ličnih imena [Transcription dictionary of English personal names]*. Nolit, 1992

- Prčić, Tvrtko, *Englesko-srpski rečnik geografskih imena [English-Serbian dictionary of geographic names]*. Zmaj, 2004
- Savary, A., L. Manicki and M. Baron. “Populating a Multilingual Ontology of Proper Names from Open Sources”. *Journal of Language Modelling* Vol. 1, no. 2 (2013)
- Vitas, Duško and Cvetana Krstev. “Derivational Morphology in E-Dictionaries of Serbian”. У *Proceedings of the 32nd International Conference on Lexis and Grammar, September 10–14, 2013, Faro, Portugal*. 2013
- Zakon, ур. *Zakon o službenoj upotrebi jezika i pisma [Law on Official Usage of Language and Script]*. Službeni glasnik Republike Srbije, 2010
- Пешикан, Митар, Јован Јерковић and Мато Пижурица, . *Правопис српскога језика [The Orthography of Serbian Language]*. Матица српска, 1993
- Стевановић, Михаило и др., ур. *Речник српскохрватскога књижевнога језика [Serbo-Croatian literary language dictionary]*. Матица српска, 1967
- Стијовић, Рада. “Званични пуни скраћени називи држава на српском и енглеском језику [Official and shorten names of countries in Serbian and English]”, 2016, internal report

Презентација српске књижевности, издаваштва и културе на сајмовима књига у свету

УДК 069.9(100):655.4(=163.41)"2018"

САЖЕТАК: Овај рад представља наступ Републике Србије на међународним сајмовима књига у свету, са аспекта културне размене. Универзитетској библиотеци „Светозар Марковић“ поверена је 2018. године организација представљања српског издаваштва, књижевности и културе на сајмовима књига у Лајпцигу, Техерану, Пекингу, Москви и Франкфурту. Посебна пажња посвећена је обележавању стогодишњице краја Првог светског рата, али и указивању на актуелне јубилеје наших класика. Међународна стручна и општа јавност је информисана о најважнијим савременим српским ауторима, од којих су неки били и специјални гости ових наступа, као и о модерним класицима српске литературе. Презентације српске културе на међународним сајмовима књига показала је да је присутност Србије на оваквим манифестацијама веома важна јер доприноси културној размени међу државама и народима и повећава видљивост Србије на светској културној мапи.

РАД ПРИМЉЕН: 21. новембар 2018.

РАД ПРИХВАЋЕН: 24. децембар 2018.

Василије Милновић

*Универзитетска библиотека
„Светозар Марковић“
Београд, Србија*

Према објављеном конкурс Министарства културе и информисања Републике Србије за Отворени позив – организација и реализација годишњег програма представљања српске књижевности и издаваштва на међународним сајмовима књига у 2018. години, Универзитетској библиотеци „Светозар Марковић“ је, захваљујући пројектном предлогу под називом *Цео век после краја* – презентација српске књижевности,

преводиштва, издаваштва и културе на међународним сајмовима књига у Лајпцигу, Техерану, Пекингу, Москви и Франкфурту, поверен тај одговоран задатак.

Универзитетска библиотека „Светозар Марковић“ (УБСМ) установа је од националног значаја и матична библиотека за све високошколске установе у Србији, која врши едукацију како библиотекара, тако и других стручних и научних профила који се образују на Београдском универзитету и многим другим универзитетима у Србији. Омогућавајући приступ записаном знању, подстичући и подржавајући проток информација, УБСМ учествује у педагошком, научно-истраживачком, стручном и уметничком раду српске академске јавности. Она је чувар писаног, научног и културног наслеђа Универзитета у Београду, а због своје дуге историје, чува и део културне баштине Србије од непроцењивог значаја. При томе, сам настанак УБСМ везан је за Први светски рат: као установа која је уништена у бомбардовању током Првог светског рата, садашња зграда је једна од три Карнегијеве библиотеке у Европи (Akund, 2011), а добар део почетног фонда обезбеђен је путем ратне репарације. Осим тога, УБСМ је данас постала главни добављач иновативног дигиталног садржаја и унапредила је своје пословање захваљујући богатим пројектним активностима. Као таква, представљала је заправо логичан и адекватан избор.

Ваља нагласити да је томе у прилог ишао и састав пројектног тима УБСМ, у коме су се налазила два доктора књижевних наука, један докторанд англистике и један мастер германистике, вођен од стране проф. др Александра Јеркова, управника УБСМ и еминентног професора српске књижевности XX века и савремене књижевности на Филолошком факултету Универзитета у Београду, који је био и уредник програма у оквиру овог пројекта. При томе, професор Јерков има богато издавачко искуство, као дугогодишњи уредник Просвете, Вулкана и других највећих домаћих издавачких кућа, као и завидно искуство када је реч о међународним сајмовима књига. На крају, за техничку реализацију пројекта је, у складу са поменути позивом надлежног министарства, била одговорна Привредна комора Србије.

Оно што је било заједничко свим наступима на поменути сајмовима књига била је жеља да се повежу културна историја Србије са савременом српском књижевношћу и културом, науком и уметношћу. Целина српског издаваштва, цео век после краја Првог светског рата, и траг који је он оставио, видела се тако што се најживљи

део наше књижевне продукције и културе постављао увек спрам књижевности и културе земље домаћина. Фокус наших наступа је, у складу са стогодишњицом Првог светског рата, обухватио искуство обнове државе и победе, али и културу примирја и помирења на којој је заснована савремена европска култура толеранције и демократије. Један другачији поглед на ствари, заснован на принципу сличности, данас се, пред многим изазовима савременог глобализма чини зато привилегованом формом. Према студији Лаку-Лабарта и Нансија „Књижевни апсолут“, примеса уметности и живота, пројектована као ризик како за друштво тако и за самог песника, повезана је првенствено са романтичарском поезијом (Lacoue-Labarthe et al., 1988). Идентичан, али обрнути процес усмерен ка изласку из корица књига у конкретну стварност, фигуративно говорећи, јесте главно настојање модернистичке парадигме. У том смислу, о авангардно-модернистичком настојању се може размишљати, следећи мисао Лаку-Лабарта и Нансија као о последици незавршеног пројекта романтизма и то превасходно оног немачких романтичара из Јене. Ако у овом контексту, примера ради, тумачимо замашно интересовање немачке културе за српску народну традицију у доба романтизма или константно интересовање српске културе за немачке културне вредности, што нарочито ескалира у доба предавангардне и авангардне српске културне праксе, онда тиме демонстрирамо могућност сагледавања двеју култура по принципу сличности, без обзира на чињеницу да је једна од њих репрезентативна у светским размерама. Јер уколико бисмо националну и глобалну културу замислили у координатном систему, замишљена крива би у антологјским делима или ствараоцима тзв. малих култура свакако морала бити увек при врху графикана, заједно са онима из тзв. великих култура.

Посебан акценат је зато стављен на јубилеј: 125 година од рођења великог српског писца Милоша Црњанског, као и на читав контекст модернистичке реинтерпретације традиције и разноликих утицаја који су долазили у српску књижевност и културу од XVIII века до епохе авангарде. У том смислу, наш штанд је на сваком сајму био платформа за културу разлике и сарадње, за једнако представљање националних мањина и српског културног корпуса изван територије Србије, равноправан положај свих књига и издавача, уз један додатни научни, академски акценат који је Универзитетска библиотека као део Универзитета у Београду донела.

Основни циљ пројекта *Цео век после краја* чиме се мисли на крај Првог светског рата, али асоцира и на крај века, помало и на искуство ратова уопште које ваља оставити за собом био је да се на репрезентативан, високо квалитетан начин презентује српска књижевност, преводилаштво, издаваштво и култура у целини, на међународним сајмовима књига у Лајпцигу, Техерану, Пекингу, Москви и Франкфурту. Примарна циљна група међу посетиоцима ових сајмова били су људи културе, уредници, агенти и издавачи, писци и преводиоци, научници и новинари, представници културне дипломатије. С обзиром на општи циљ јачања представљања наше културе и књижевности у свету и системску подршку развоју међународне сарадње и размене у области културе и уметности, циљ је био остављање снажног утиска на међународну стручну и културну јавност и на посетиоце уопште. Индиректна циљна група била је и домаћа стручна и општа јавност, која је информисана о предузетим активностима редовним извештајима, саопштењима за јавност и медијским прилозима. Стручна јавност је посебно била у фокусу пажње за време панела у којима су учествовали чланови српске делегације на сајмовима.

Представљање српске књижевности, издаваштва и преводилаштва по наведеном концепту подразумевало је указивање на међународни значај и универзалност врхунских дела националне књижевности, али и на њихове специфичности, нарочито с обзиром на утицаје и мотивска укрштања са делима из других књижевности. Ово посебно долази до изражаја у послератној авангардној парадигми српске књижевности и чињеници 1918. године, као години успона из ратног пепела, али и отварања српске културе ка великим књижевностима Европе и света што је и био фокус нашег концепта. Представљање српске књижевности и културе, у том смислу, на еминентним сајмовима књига, заправо значи стални труд на постављању српске културе у светске координате и има неупитан државни значај. Посебно важан сегмент презентације је традицијски лук који се у српској књижевности и култури успоставио од XVIII века, преко модернистичке реинтерпретације наслеђа и грађења новог лика књижевности у великом делу Милоша Црњанског (због чега је његов јубилеј био посебно срећна и важна околност), све до сасвим актуелне обраде европског културног наслеђа у савременој српској књижевној продукцији.

Визуелни идентитет штандова је имао непроменљиви део у свакој земљи – односио се на 100 година од примирја у Првом светском рату и 125 година од рођења Милоша Црњанског, што је била и

основа визуелног идентитета – и променљиви део који се односио на савремене књиге и ауторе који су се представљали на конкретном сајму, као и изложбени оквир с тим у вези. На пример, у Лајпцигу и Франкфурту је представљен први докторат о Доситеју Обрадовићу одбрањен у Лајпцигу, који се налази у фонду УБСМ, док је у Техерану представљена збирка оријенталних рукописа, такође из фонда УБСМ, од којих је девет на персијском језику. Такође, програмско ослањање на модернистички поетски контекст као полазна тачка за дијалог са савременом књижевношћу имао је своје утемељење у чињеници да је поезија, на различите начине, у Немачкој, Русији, Ирану и Кини посебно истакнут уметнички облик.

Истовремено, на штанду су увек били постављени монитори на којима су се смењивале фотографије писаца и књига, као и инсерти из позоришних, филмских и ТВ адаптација, на основу стваралаштва Милоша Црњанског и других модернистичких стваралаца, али и савремене српске књижевности. Изложбена поставка подразумевала је излагање капиталних издања, новијих књижевних, теоријских, историографских и монографских дела на српском језику, као и излагање дела која су последњих година преведена на језик земље у којој се сајам одржавао (на пример, превод књиге Горана Петровића „Атлас описан небом“ на фарси), али и монографије са темом Србије. Другим речима, излагале су се књиге свих важнијих издавача, књиге пристигле по позиву упућеном издавачким кућама и установама културе које имају релевантну издавачку делатност. Такође, излагане су и књиге на мањинским језицима.

У стручној јавности је већ одавно антиципирано да ће истраживачи све више информације битне за њихов истраживачки рад обезбеђивати путем дигиталне технологије (Newtown, 2000). Другим речима, способност приступа и употребе дигиталних технологија постаје неопходан аспект савремене научне праксе. Академске библиотеке су управо она места где истраживачи откривају нове информације и где се развија и унапређује дигитална писменост и обученост истраживача (Julien et al., 2017). Улога такве библиотеке је да постане својеврсни информациони пункт где се може експериментисати са новим технологијама и где се могу усвајати нови концепти истраживања (Jerkov et al., 2015). Управо због тога, а у складу са концептом наступа и чињеницом да је за њих била задужена највећа академска библиотека на овим просторима, на сајмовима у Лајпцигу и Франкфурту, представљени су и интерактивни екрани (*Magic Box*) мањих и већих

димензија, који су недавна инвестиција УБСМ.¹ Инсистирајући на академском приступу темама, УБСМ је запажено место посветила учешћу професора и стручњака из датих области, јер УБСМ је као део Универзитета у Београду приступила овом пројекту са жељом да се на сајмовима, поред осталог, афирмише научна подршка и удео у стварању научних дела, вредних издања и изузетних пројеката из области науке и културе, на савремен и технолошки атрактиван начин. Ове презентације публика је веома добро пропратила на оба значајна сајма.

На сваком сајму су се одржавали панели у којима су учествовали чланови српске делегације. Дакле, учесници сајамских програма били су: уредник програма, преводилац на језик земље организатора сајма, један или два координатора сајамског наступа из УБСМ (који су истовремено и стручни за већину тема панела) и гости: писци, књижевни историчари, теоретичари, критичари, уредници издавачких кућа и културни посленици који су стручни и компетентни да говоре о темама панела или који су имали дело које реферише о тим темама или који имају објављене публикације и радове на језику земље организатора сајма.

Тако су, примера ради, будући да су у фокусу сајма у Лајпцигу били писци из мањинских заједница, наши гости на штанду Србије, кроз сарадњу са организацијом преводилаца књижевности југоисточне Европе „Традуки“, били српски писци Ото Хорват и Енес Халиловић. Они су, истовремено, били и гости самог сајма књига у Лајпцигу и у склопу својих активности имали веома посећене панеле и књижевне вечери. У Техерану, где смо били и почасни гост сајма, од писаца гости су били Горан Петровић, Дејан Алексић и Мухарем Баздуљ, али и шира делегација руководиоца најзначајнијих српских институција културе, предвођена министром културе и информисања Републике Србије, директором Народне библиотеке Србије и председником Матице српске, док су у Пекингу гостовали Владимир Кеџмановић и проф. др Драган Бошковић. У Москви су осим књижевника – Владана Матијевића, Драгана Великића и Николе Вујчића, и књижевног преводиоца Наталије Ненезић, делегацију чинили и представници сајма књига у Београду, као и Града Београда, док су у Франкфурту – највећем од свих сајмова, који је више од других оријентисан на бизнис – осим књижевника одомаћених у немачком књижевном контексту (Стеван Тонтић и проф. др Бојана Стојановић Пантовић), били присутни и издавачи (поред осталих: Бора

¹ Магија на српском штанду (на вебу)

Бабић, Ален Бешић, Ненад Шапоња, Радован Влаховић, Дејан Алексић итд.).

Неке од стандардних тема наших панела су биле:

- 1918: крај и почетак (Лајпциг, Техеран, Пекинг, Москва),
- Доба великих писаца: од Црњанског до данас (Техеран, Пекинг, Москва),
- Путовање мотива: утицаји европских књижевности на традицијски лук: српска култура XVIII века – модернистичка реинтерпретација српске традиције – савремено читавање наслеђа (Лајпциг, Москва, Франкфурт),
- Нови хоризонти дигитализације културног наслеђа (Франкфурт),
- Мањинска књижевност у срцу националне литературе (Лајпциг).

За наступ Србије на међународним сајмовима књига израђивао се комуникациони материјал у штампаном и електронском издању: брошура *Цео век после краја: програм представљања Србије* (на конкретном сајму), каталог *Милош Црњански: велики непознати европски писац*, као и два електронска каталога: *Каталог издавача* и *Избор преведених књига* (у последњих пет година, на језик земље у којој смо се представљали). Такође, на штанду је увек био доступан промотивни материјал и сувенири прикупљени у сарадњи са Туристичком организацијом Србије.

Уочи одласка на конкретни сајам, посредством аташеа за штампу у нашим амбасадама, на језику земље организатора, тамошњим медијима најављивала су се наша представљања са основним информацијама и детаљима. Дописницима наших медија упућивали су се лични позиви да посете наш штанд и програме. За време конкретног сајма, домаћим медијима су се свакодневно слале информације, најаве догађаја и извештаји. Најбитније информације и фотографије са сајамских презентација постављане су на веб-страницу Министарства културе и информисања Републике Србије и РТС-а, са којим УБСМ има уговор о сарадњи. Ове информације су такође биле дневно ажуриране на веб-сајту и фејсбук страници УБСМ.

Верујући у значај наступа Србије на наведеним сајмовима, УБСМ је припремала, у складу са могућностима, и додатне активности, које су допринеле квалитету промоције наше културе и књижевности. Тако је, на пример, у Москви организована посета споменику Милораду Павићу у дворишту московске Библиотеке светске књижевности, док је у Техерану уприличен сусрет иранских и српских писаца.

Представљање српског издаваштва, књижевности и културе је пут да нас свет боље упозна. Важно је представити српску књижевност, издаваштво и преводилаштво, указати на међународни значај и универзалност врхунских дела националне књижевности, али и на њихове специфичности, нарочито с обзиром на утицаје и мотивска укрштања са делима из других књижевности. С обзиром на циљ конкурса на који се односио пројекат *Цео век после краја* – развој међународне сарадње у области књижевности, преводилаштва и издаваштва – резултат овог пројекта несумњиво ће утицати на јачање капацитета актера из области културе и уметности на територији Републике Србије и на конкретан развој међународне сарадње и културне размене, али водиће и њиховој пуној интеграцији у концепт „отворене науке“ (Fresa, 2014) и транспарентности информација.

Значај овог прегнућа може се сагледати тек учешћем у њему. Стога би, у закључку, неопходно било направити вишегодишњу стратегију наступа Србије на овим изузетно значајним манифестацијама, али и издвојити већи буџет за учешће, како би њихов утицај био силовитији и учинковитији по државу Србију и њену културу. Искуство УБСМ, у том смислу, може бити од велике помоћи.

Литература

- Akund, Nadin. “Delatnost Karnegijevih zadužbina na Balkanu posle Prvog svetskog rata: Univerzitetska biblioteka u Beogradu, 1919-1926”. *INFOteka* Vol. 12 , no. 1 (2011): 3–12
- Jerkov, Aleksandar Adam Sofronijević and Dejana Kavaja. “Smart and Sustainable Library: Information Literacy Hub of a New City”. *Information Literacy: Moving Toward Sustainability* (2015): 22–30
- Fresa, Antonella. “Digital Cultural Heritage Roadmap for Preservation”. *Journal Of Humanities & Arts Computing: Journal Of Digital Humanities* no. 8 (2014): 3–12
- Julien, Heidi, Melissa Gross and Don Latham. “Survey of Information Literacy Instructional Practices in US Academic Libraries”. *College & Research Libraries* (2017): 17–2014
- Lacoue-Labarthe, Philippe and Jean-Luc Nancy. *The Literary Absolute: the Theory of Literature in German Romanticism*. New York: State University of New York Press, 1998

Newtown, Leonard. "Data-logging in practical science: research and reality".
International Journal of Science Education Vol. 22, no. 12 (2000): 1247–
1259

Платформа за отворену науку – обавезност публиковања у отвореном приступу у Републици Србији

УДК 027.8(497.11):371.38]:004.738.5

Весна З. Абадић
vesnaa@kg.ac.rs

Марија М. Гордић
mgordic@kg.ac.rs

Универзитетска библиотека
Крагујевац, Србија

САЖЕТАК: Пратећи рад универзитетске јавности, уочавају се предности умреженог дигиталног окружења, којим свако може брже и лакше приступити оствареним резултатима истраживања. Масовна комуникација, без повреде кодекса академског интегритета, довела је до појаве нове парадигме: отворене науке (ОН). Библиотечко-информациони системи у Србији увелико раде на одрживом развоју дигиталних репозиторијума активно учествујући у пројектима Министарства НППС. BE-OPEN (енг. *Boosting Engagement of Serbian Universities in Open Science*) као Ерасмус+ структурни пројект из области изградње капацитета у високом образовању, окупио је највеће државне универзитете. Усвајањем “Платформе за отворену науку”, универзитети су у обавези да оформе институционалне платформе, као предуслов примене принципа отворене науке.

КЉУЧНЕ РЕЧИ: отворена наука, платформа за отворену науку, BE-OPEN.

РАД ПРИМЉЕН: 15. август 2018.

РАД ПРИХВАЋЕН: 22. новембар 2018.

1. Пракса отвореног приступа у Србији

Универзитетске библиотеке у Србији са мрежом академских библиотека чине библиотечко-информационе системе универзитета.

Више од читаве деценије, библиотеке чланице истог система залажу се за публикување у отвореном приступу и имају кључну улогу у одржавању инфраструктуре дигиталних репозиторијума (zak, 2011). Дефиницију отвореног приступа даје Питер Субер – „литература у отвореном приступу је дигитална, онлајн, бесплатна и ослобођена ограничења по питању копирајта и лиценцирања“ (Suber, 2016). Отворени приступ даје слободу целокупној научној заједници да размењује идеје и резултате научноистраживачког рада. Универзитети у Београду, Нишу и Крагујевцу су 2011. године, у оквиру TEMPUS пројекта „Нови библиотечки сервиси на универзитетима Западног Балкана“, као дигитални репозиторијум, успоставили систем PHAIDRA, преузет са Универзитета у Бечу.¹ Без обзира на промоције и одржане едукације на поменутиим универзитетима, сами истраживачи нису депоновали велики број објеката. Највећи број дигиталних колекција су поставили сами библиотекари. Масовнију примену PHAIDRA добија изменом Закона о високом образовању² који уводи обавезност депоновања одбрањених докторских дисертација.

Отворена наука³ (енгл. *Open Science*) представља принцип који промовише и креира слободан приступ научним знању и резултатима научних истраживања, без правних, технолошких или друштвених ограничења. Као нови термин у академској заједници, могло би се рећи да вуче корене још од краја 16. и почетка 17. века (David, 2008), када је одређена заједница људи препознала потребу за узајамном комуникацијом у оквиру исте научне области. Овакав вид ренесансе у науци потиснуо је устаљени интровертан начин опхођења појединаца у оквирима научноистраживачког рада и започео нову „револуцију“ у академским круговима (David, 2008). „Дељење“ података довело је до пораста „кооперативних ривалитета“ у откривању новог знања (Da-

¹ PHAIDRA дигитални репозиторијум универзитета у Бечу (приступљено 26. 07. 2018)

² Чланом 30. став 9. Закона о високом образовању („Сл. гласник РС“, бр. 76/2005, 100/2007 - аутентично тумачење, 97/2008, 44/2010, 93/2012, 89/2013, 99/2014, 45/2015 - аутентично тумачење, 68/2015 и 87/2016) прописано је да је Универзитет дужан да установи дигитални репозиторијум у којем се трајно чувају електронске верзије одбрањених докторских дисертација, заједно са извештајем комисије за оцену дисертације, подацима о ментору и саставу комисије и подацима о заштити ауторских права, као и да све наведене податке учине јавно доступним.

³ Отворена наука (приступљено 26. 07. 2018)

vid, 2004). Ова друштвена појава омогућила је израду многобројних лако доступних извора података, па тако и лакшу комуникацију истраживача. Популарност и неопходност праћења и коришћења свих аспеката отворене науке, приступа информацијама и изворима података у сталном је порасту (McKiernan, 2016), иако се може рећи да овај концепт још није потпуно усвојен. Данас, захваљујући електронској комуникацији, помоћу отворене науке можемо пропратити све сегменте једног истраживачког процеса, почевши од методологије рада, преко информација о коришћеној апаратури и техничким средствима, све до резултата истраживања. Овим својеврсним „акцелератором знања“ (Wolleff et al., 2011) постиже се већа транспарентност науке.

Отворена наука, као шири појам од самог публикувања у отвореном приступу, заснива се на више принципа:

- отворени подаци добијени истраживањима,
- отворени код,
- творено рецензирање, научна комуникација и методологија,
- транспарентно праћење резултата научног рада у циљу даље евалуације, употребом широког спектра индикатора, за све врсте истраживачких резултата.

Европска комисија је у априлу 2018. године донела „Препоруке за платформе отворене науке“ (енгл. *Open Science Policy Platform Recommendations*) у којима је назначено осам приоритета који подстичу њен развој (Commission, 2017). Акценат је на отвореним подацима, њиховој заштити и могућности да буду поново употребљени (Kanjilal and Das, 2015).

Јасно је да овакав подухват захтева и комплексну техничку инфраструктуру, компатибилну са постојећим системима и обученост свих учесника у процесу. Обавезност публикувања у отвореном приступу (радова и резултата истраживања) насталим у пројектима које финансирана Европска уније, донела је Европска комисија у виду „Смерница за научно публикување у отвореном приступу и отворени приступ истраживачким резултатима“ (енгл. *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research data in Horizon 2020*) у оквиру програма ХОРИЗОНТ 2020 (H20, 2017). Наши истраживачи који су учествовали у међународним пројектима депоновали су своје радове на различите платформе или мреже (Zenodo, Research Gate...) јер није постојала законска регулатива за ову област.

2. Реализација BE-OPEN пројекта

Академска јавност је свесна оправданости потребе публикавања у отвореном приступу; постоје појединачни примери добре праксе института из Србије који годинама уназад имају институционалне или тематске репозиторијуме. Чињеница да се буџет Републике Србије намењен развоју науке не може поредити са земљама Европске уније био је мотив да се додатне могућности траже и развијају у оквиру пројектног финансирања Европске комисије и преузимају готова решења. Свесни чињенице да се ова област мора формално уредити, највећи универзитети у Србији су се окупили у BE-OPEN,⁴ Еразмус+ пројекту (област К2 – изградња капацитета у високом образовању). У њему учествује шест државних универзитета, као и Министарство просвете, науке и технолошког развоја Републике Србије. Пројекат се реализује у периоду од октобра 2016. до октобра 2019. године.

Главни циљ пројекта је управо развој услова за имплементацију пуних принципа отворене науке, што ће се реализовати у следећим корацима:

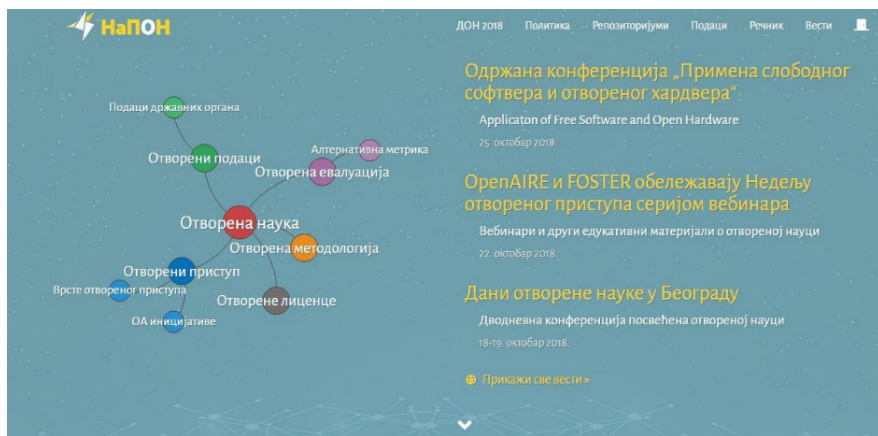
- развој националних и институционалних правних аката и смерница (платформи);
- имплементација институционалних дигиталних репозиторијума на свим универзитетима у Србији, као и Националног портала отворене науке;
- јачање индивидуалних компетенција организовањем семинара, конференција и радионица за све потенцијалне кориснике репозиторијума;
- интеграција путем успостављања Националног портала отворене науке која ће обезбедити трансфер технологија и знања са академске заједнице на индустрију и ширу јавност и истовремено пружити аналитичке податке за анализу резултата истраживача.

У првој години реализације пројекта дата је исцрпна анализа постојећег стања; извештаји о усвојеној легислативи и текућој пракси отворене науке у Србији доступни су на сајту пројекта, у оквиру радног пакета WP-1.⁵

Почетком новембра 2017. године креирана је веб-страница Портала отворене науке која ће представљати чвориште отворене науке у

⁴ Насловна страна BE-OPEN пројекта (приступљено 28. 7. 2018)

⁵ Радни пакет WP-1 (приступљено 20. 10. 2018)



Слика 1. Почетна страна Националног портала отворене науке

Србији, а доступна је на адреси <http://www.open.ac.rs/>.⁶ НАПОН – Национални портал отворене науке треба да повеже све постојеће и будуће дигиталне репозиторијуме појединачних институција, податке државних органа и обезбеди полазну тачку за све информације у пољу отворене науке.

3. Платформа за отворену науку

У складу са циљевима, Министарство је 9. јула 2018. године усвојило „Платформу за отворену науку“ (pla, 2017) намењену свим учесницима у научноистраживачкој делатности и односи се на резултате истраживачких пројеката и програма финансираних у целини или делимично из буџета Републике Србије. Реализација принципа отворене науке, која подразумева пуну заштиту етичких норми, ауторских права и права интелектуалне својине, одвијаће се у четири правца:

- отворени приступ научној литератури,
- доступност података прикупљених у научним истраживањима,
- транспарентност научне комуникације и методологије и

⁶ Национални портал отворене науке (приступљено 28. 10. 2018)

- развој дигиталне инфраструктуре која ће омогућити остваривање наведених циљева.

Министарство прописује норму да интегрални текст објављених резултата изворно буде у отвореном приступу, и то најкасније 12 месеци од дана објављивања у области природних, медицинских и технолошких наука, или 18 месеци за истраживања у оквиру друштвених и хуманистичких наука. Метаподатке научних публикација, одмах након публиковања, неопходно је депоновати у институционални или национални репозиторијум. Обавеза депоновања у репозиторијуме односиће се и на радове у часописима, монографијама и зборницима радова који се већ налазе у отвореном приступу. У случају да су радови претходно били објављени у часописима комерцијалних издавача, који су и носиоци права коришћења, морају се поштовати одредбе уговора који је аутор потписао са издавачем. Постоји могућност да се депонује објављена верзија (уз дозволу издавача) и уз поштовање ембарго периода, или рецензирана верзија која је прихваћена за објављивање. Корисно је да аутори пре слања рукописа провере издавачку политику часописа за који су се одлучили на порталу SHERPA/RoMEO⁷ који на једном месту обједињује издавачке политике различитих часописа.

Што се тиче научног издаваштва, уобичајена пракса у свету је да аутори плаћају издавачу надокнаду за трошкове објављивања у отвореном приступу (енгл. *APC – Article Processing Charges или BPC – Book Processing Charges*). Како би се ауторима изашло у сусрет, најављена је могућност да ови трошкови буду ставка у буџету пројеката које финансира Министарство.

Метаподаци морају бити увек јавно доступни, а јединствени идентификатори попут DOI броја дигиталног објекта (енгл. *Digital Object Identifier*) и ORCID (енгл. *Open Researcher and Contributor ID*) идентификатора истраживача биће стандардни елементи сета метаподатака. ORCID представља јединствени међународни идентификатор истраживача, повезан са афилијацијом, који је постао неопходан елемент у публиковању у бројним страним часописима. Темељан и истрајан приступ библиотекара високошколских библиотека на формирању комплетних библиографија истраживача, њиховом повезивању са одговарајућим ORCID идентификаторима, омогућиће повезивање података из узајамне библиографске базе са објектима у отвореном приступу.

⁷ Портал SHERPA/RoMEO

Што се тиче примарних података прикупљених у истраживањима, постављало се питање оправданости њиховог архивирања. Могућност да се обезбеди трајно чување, у машински читљивим форматима и доступност у оквиру интероперабилне дигиталне платформе, како би поново послужили у другим истраживањима или накнадним проверама, представља довољан разлог за овај подухват. Након архивирања, дефинише се степен доступности података. Метаподаци морају увек бити видљиви. Будуће институционалне платформе прописале услове депоновања примарних података, сврстане у три групе доступности, у складу са законским или етичким ограничењима:

- затворени подаци,
- подаци доступни дефинисаној групи истраживача или
- у потпуности јавно доступни.

Адекватан третман истраживачких података обухвата више корака, почевши од прикупљања, обраде до анализе података изабраном статистичком методом, дефинисања начина чувања и омогућавања поновне употребе од стране других истраживача. У зависности од врсте података (на пример, да ли су укључени подаци о личности) зависиће и њихова будућа заштита. Размена података између различитих истраживачких група, у различитим дисциплинама, може омогућити употребу података у новим истраживањима и на тај начин донети уштеде. Захтеви да подаци буду приступачни (*Findable*), доступни (*Accessible*), интероперабилни (*Interoperable*) и виšekратно употребљиви (*Reusable*), једним именом се називају FAIR принципи истраживачких података.

Као и већ постојећи универзитетски дигитални репозиторијуми у Србији, тако и будуће дигиталне платформе морају обезбедити интероперабилност, што подразумева могућност аутоматског преузимања података у складу са међународним протоколом OAI-PMH (енгл. *Open Archive initiative Protocol for Metadata Harvesting*) и структурирану шему метаподатака у складу са Dublin Core⁸ стандардом. До сада у Србији нису развијени репозиторијуми за депоновање истраживачких података, па је у недостатку институционалних ресурса препорука Министарства да истраживачи своје податке депонују у одговарајуће међународне тематске репозиторијуме.

⁸ [Dublin Core metadata иницијатива](#) (приступљено 2. 8. 2018)

4. Законске одредбе

На све депоноване објекте стандардно ће се примењивати машински читљиве лиценце Креативне заједнице – *Creative Commons* (CC)⁹ које дефинишу услове под којима се дело или резултати истраживања могу даље користити.

За пуну примену ових одредби, усвојен је период од шест месеци, за који универзитети и институти, као самосталне јединице, морају усвојити своје институционалне платформе за отворену науку.

У склопу завршних одредби текста „Платформе за отворену науку“ наводи се да ће Министарство пратити поштовање свих наведених норми, јер ће од резултата праћења зависити будућа финансирање нових пројеката и осталих делатности под ингеренцијама Министарства. У оквиру својих правних аката – институционалних платформи, установе ће имати обавезу да дефинишу ниво обавезности публиковања у отвореном приступу, што се може верификовати на различите начине, на пример у поступку вредновања приликом избора у звање. С обзиром на то да се ниво обавезности усваја од тренутка усвајања законске регулативе, питање је како ће се у репозиторијумима наћи научни радови настали пре законске обавезе. Могуће аутоматско преузимање раније објављених публикација, претходно депонованих у другим репозиторијумима, уз преузимање доступних података из цитатних база и других отворених архива, допуниће корпус отворене науке у Србији.

Постоји велики број радова истраживача из Србије у отвореном приступу који су депоновани на бројним порталима и агрегаторима широм света. Није ретка појава да се код извесног броја радова не види афилијација аутора, или она није ажурирана. Додатним ангажовањем је омогућено повезивање тих радова са институционалним репозиторијумима, што ће у крајњој линији имати утицај на ширење српске научне мисли, као и на цитираност аутора и укупну позицију сваког универзитета на некој од лествица за рангирање.

5. Закључак

Из свега наведеног, са сигурношћу се може закључити да ће реализација пуних принципа отворене науке допринети бољој дисеминацији и видљивости научне продукције, њеном адекватном

⁹ *Creative commons ауторство 3.0 Србија* (приступљено 2. 8. 2018)

вредновању и већој искоришћености резултата научних процеса. Подстицање истраживачких процеса, повећање видљивости резултата и могућност да се употребе више пута (Swan, 2016) доноси многе предности широкој јавности. Систематизованост у праћењу резултата научног рада, поштовање одредби ауторских и сродних права у коришћењу јавно доступних података, њихова евалуација у пројектном одлучивању и сл. постићи ће се усвајањем бројних унификованих процедура у раду. Узимајући у обзир све аспекте које треба задовољити, јасно је да је у питању читав низ техничких, правних, организационих елемената и да је потребно време да се нотификују пуни ефекти примене. Универзитети и институти у Србији имали су различите праксе и искуства у овом пољу, па је један од резултата и уједначеност процедура, док свест о важности овог процеса свакако постоји.

Будућа инфраструктура окупиће све актере у процесу креирања научних резултата: установе, истраживаче, издаваче, рецензенте, библиотекаре са једне стране и финансијере и кориснике резултата научне заједнице са друге стране (Министарства, привредни субјекти) и допринети ефикаснијој повезаности науке и привреде, али и науке и целокупног друштва путем развоја иновативних услуга и производа. Библиотекари имају активну улогу у овом процесу, почевши од едукације и помоћи у вези са процедурама депоновања, контроле метаподатака, одржавања система, до надзора у примени „Платформе за отворену науку“, у складу са институционалним регулативама које ће се усвајати, па су стога академске библиотеке кључне у посредовању у процесима отворене науке.

Литература

- “Закон о библиотечко-информационој делатности”, 2011
- “H2020 Programme: Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020”, 2017, accessed on 07/28/2018, <http://www.mpn.gov.rs/wp-content/uploads/2018/07/Platforma-za-otvorenu-nauku.pdf>
- “Платформа за отворену науку”, 2017. Приступљено 30.7. 2018, <http://www.mpn.gov.rs/wp-content/uploads/2018/07/Platforma-za-otvorenu-nauku.pdf>
- Commission, European. “OSPP-REC: Open Science Policy Platform Recommendations”, 2017, accessed on 07/28/2018, <https://ospp-rec.europa.eu/>

- [//ec.europa.eu/research/openscience/pdf/integrated_advice_opspp_recommendations.pdf#view=fit&pagemode=none](https://ec.europa.eu/research/openscience/pdf/integrated_advice_opspp_recommendations.pdf#view=fit&pagemode=none)
- David, Paul A. "Understanding the emergence of open science institutions: functionalist economics in historical context". *Industrial and Corporate Change* Vol. 13, no. 4 (2004): 571–589. Приступљено 30. 10. 2018, <https://academic.oup.com/icc/article-abstract/13/4/571/718486?redirectedFrom=fulltext>
- David, Paul A. "The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution". *Capitalism and society* Vol. 3, no. 2 (2008). Приступљено 30. 10. 2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2209188>
- Kanjilal, Uma and Anup Kumar Das. "Introduction to open access", 2015, Преузето 30. 7. 2018, <http://unesdoc.unesco.org/images/0023/002319/231920E.pdf>
- McKiernan, Erin C. "Point of view: How open science helps researchers succeed", 2016, Приступљено 30. 10. 2018, <https://elifesciences.org/articles/16800>
- Suber, Piter. "Otvoreni pristup", 2016
- Swan, Alma. "Policy guidelines for the development and promotion of open access", 2016, Приступљено 30. 10. 2018, <http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>
- Woelfle, Michael, Todd Oliario Piero and H. Matthew. "Open science is a research accelerator". *Nature Chemistry* Vol. 3 (2011): 745–748, Приступљено 30. 10. 2018, <https://www.nature.com/articles/nchem.1149>

Фокус на отвореној науци у Универзитетској библиотеци „Светозар Марковић“

Ивана Гавриловић
ivana.gavrilovic@gmail.com

Универзитетска
библиотека
„Светозар Марковић“
Београд, Србија

РАД ПРИМЉЕН: 22. новембар 2018.

РАД ПРИХВАЋЕН: 1. децембар 2018.

У Универзитетској библиотеци „Светозар Марковић“ у Београду, 12. новембра 2018. године одржана је једнодневна међународна радионица под називом „Фокус на отвореној науци (Поглавље VIII).“¹ Радионица је реализована у сарадњи Библиотеке са компанијом Scientific Knowledge Services а уз подршку удружења LIBER.² Радионица је намењена библиотекаrimа и научним радницима, и осма је у низу истоимених поглавља одржаних у Европи.³

Након уводног излагања проф. др Александра Јеркова, директора Библиотеке домаћина ове радионице, кроз четири сесије се разматрало о многим релевантним темама које се тичу отворене науке данас. Акценат је био пре свега на улози библиотека, посебно академских, на пољу отворене науке; на институционалној, регионалној и међународној размени искустава; на неопходности ширења свести о важности ове теме; партнерству; начинима финансирања; као и на стратегијама заговарања, а све у циљу примене најбоље праксе.

Модератор радионице у Београду био је др Тиберијус Игнат.⁴ Компанија чији је он директор бави се унапређењем рада европских библиотека кроз имплементацију нових технологија ради остваривања што већег потенцијала за пружање, размену и коришћење информативних и дигиталних услуга. Сви присутни на радионици

¹ *Focus on Open Science (Chapter VIII).*

² *Ligue des Bibliothèques Européennes de Recherche – Association of European Research Libraries.*

³ Претходиле су радионице у Будимпешти, Љубљани, Барселони, Гдањску. Наредна је у Бечу.

⁴ Tiberius Ignat, директор компаније *Scientific Knowledge Services*.

имали су прилику да учествују у раду интерактивно, путем друштвених мрежа или директно.

У првом блоку предавања говорили су др Пол Ајрис,⁵ један од највећих ауторитета данас на овом пољу, и Ванеса Праудмен,⁶ директорка европског сектора удружења SPARC. Они су указали на значај концепта отворене науке, важност демократизације знања и неопходност доступности научних радова широј заједници.

Др Пол Ајрис је представио Платформу⁷ Европске комисије која се састоји од низа препорука, са посебним акцентом на осам кључних области отворене науке: 1) награде и подстицаји, 2) индикатори истраживања и алтметрија наредне генерације, 3) будућност научне комуникације, 4) интегритет истраживања, 5) вештине и образовање, 6) учешће грађанства у научно-истраживачким процесима, 7) подаци FAIR⁸ стандарда, као и 8) коришћење јединственог виртуелног окружења⁹ од стране истраживача. Такође је указао на модел објављивања радова који је развио Лондонски универзитетски колеџ¹⁰ оснивањем *Open Access Press*-а, чији је он извршни директор. Оснивањем отвореног издаваштва омогућено је да научно-истраживачки радови са Лондонског универзитетског колеџа буду лако, једноставно и брзо доступни широм света. Посебно је истакао потребу за подизањем нивоа свести у научним заједницама, променом застарелих образаца размишљања и преиспитивањем политике самих универзитета.

Ванеса Праудмен је указала на потребу пружања подршке истраживачима и студентима, у виду тражења нових модела финансирања, развијања нових алата за управљање научним подацима, као и на потребу комуникације и сарадње између држава како би се идеја отворене науке уопште одржала и даље развијала у будућности.

⁵ Paul Ayris (University College London); наслов излагања: *Be the change that you wish to see in the world (Mahatma Gandhi): Universities and Cultural Change to Deliver Open Science*.

⁶ Vanessa Proudman (*Scholarly Publishing and Academic Resources Coalition – SPARC*); наслов излагања: *It's high time to rethink how we pay for Open Infrastructure; it's high time to act*.

⁷ Open Science Policy Platform.

⁸ Акроним настао од следећих термина: Findability (претраживост) Accessibility (приступачност), Interoperability (компатибилност) и Reusability (могућност поновне употребе).

⁹ European Open Science Cloud.

¹⁰ London University College.

У другом блоку предавања, тема је разматрана првенствено из перспективе универзитета и истраживача. Говорили су др Игнаси Лабастида и Хуан¹¹ и др Јанис Цаконас,¹² представник LIBER-а. Др Игнаси Лабастида и Хуан истакао је да универзитети морају променити матрицу по којој управљају научним информацијама. Информације се данас складиште, дисеминирају, објављују и користе на другачији начин него у прошлости, а и сам ток научног истраживања се такође одвија на другачији начин. Универзитети морају трансформисати своју политику уз поштовање принципа отворене науке. Јанис Цаконас је говорио о стратегији под називом Мапа отвореног пута.¹³ Удружење LIBER, као највећа мрежа истраживачких библиотека у Европи, донело је нову стратегију активности и предложило је низ практичних корака које библиотеке треба да предузму: промовисање принципа отворене науке, ширење свести о њеној важности у друштву, континуирана едукација особља у погледу употребе нових алата и сервиса, као и сарадња и партнерство са сродним установама на локалном и међународном нивоу.

Миро Пушник,¹⁴ директор Централне техничке књижнице у Љубљани и један од домаћина прве радионице одржане у Љубљани 2015. године, поделио је са присутнима своје искуство у вези са имплементацијом отворене науке у Словенији. Указао је на улогу коју академске библиотеке имају у трансформацији парадигме научно-истраживачког процеса, истичући значај изградње добре инфраструктуре и важност едукације библиотекара, као битних актера ових промена.

Централна идеја др Адама Софронијевића,¹⁵ заменика директора Универзитетске библиотеке „Светозар Марковић“, заснива се на промени односа савременог друштва према институцијама, првенствено научно оријентисаним (универзитети, библиотеке, научни центри, консалтинг агенције, министарства и сл.). Фокус оваквих институција помера се

¹¹ Ignasi Labastida i Juan (Universitat de Barcelona); наслов излагања: *Facing the Open Science challenges from a university perspective*.

¹² Giannis Tsakonas (University of Patras, Greece); наслов излагања: *A new kind of dialogue : Open Science as a discourse topic between libraries, researches and society*.

¹³ Open Road Map.

¹⁴ Miro Pušnik (University of Ljubljana); наслов излагања: *Encouraging Open Science Literacy: For years of Focus on Open Science workshops*.

¹⁵ Adam Sofronijević (University of Belgrade); наслов излагања: *Open Science: The Remedy for a Post-truth World*.

са развоја науке на бизнис, а оне саме губе своју основну улогу „заштитника истине“. Постају затворене за шири аудиторијум, а њихов рад је скоро нетранспарентан. Грађани су са једне стране изгубили веру у институције, а са друге им је на располагању обиље информација које су често непроверене, чак и нетачне. Управо из ових разлога, институције морају трансформисати своју свест кроз концепт отворене науке, како би вратиле изгубљено поверење људи и усмериле даљи развој науке у одговарајућем смеру.

Open Science је покрет који се заснива на отворености, сарадњи и размени. Покрет је усмерен на развој науке за добробит човечанства. Библиотеке, пре свега академске, морају препознати значај своје улоге у транзицији и пронаћи своје место у њој. Поред добре воље, неопходно је познавање и поштовање права, али и обавеза које прате научни рад. Потребно је и стално учење, упознавање са новим алатима, као и повезивање са сродним институцијама у циљу размене добре праксе. *Focus on Open Science* радионице имају улогу да дају смернице и упутства, али и да истакну све предности овог покрета: лакши и бржи приступ научним информацијама, већа транспарентност научног рада, економски аспект, отварање институција према грађанству и многе друге. Оне дају препоруке на који начин промена друштва и његових установа треба да се одвија, али исто тако упозоравају на замке и проблеме који прате транзицију ка отвореној науци: ограничења у погледу ауторских права, питање евалуације великог броја научних радова у отвореном приступу, повећање административних послова научних радника приликом објављивања радова (које су традиционално обаљали издавачи) и сл. Радионица у Београду је свакако у томе успела.

Приказ летње школе „ESSLLI 2018“

Бранислава Шандрих
branislava.sandrih@fil.bg.ac.rs

Филолошки факултет
Универзитет у Београду
Србија

РАД ПРИМЉЕН: 5. децембар 2018.

РАД ПРИХВАЋЕН: 18. децембар 2018.

1. О школи

Европска летња школа логике, језика и информација, ESSLLI 2018 (енгл. *European Summer School in Logic, Language and Information*) одржана је 6–17. августа 2018. године у Софији, престоници Бугарске. Није изненађујућа беспрекорна организација и ове године, будући да је школа одржана по тридесети пут. Претходне школе одржане су у Тулузи (Француска), Болцану (Италија), Барселони (Шпанија), Тибингену (Немачка), Диселдорфу (Немачка) итд. За организацију су одговорне следеће установе: Софијски универзитет „Св. Климент Охридски“, Институт за информације и комуникационе технологије, Бугарска академија наука, као и Бугарско удружење за рачунарску лингвистику. Настава је била одржавана у просторијама ректората Универзитета. Позив учесницима у виду постера може се видети на Слици 1.

2. Организација

Програмски одбор чинили су:

- Председавајућа Лаура Калмејер, Универзитете у Диселдорфу, Немачка;
- Ко-председавајућа Галија Ангелова, Институт за информационе и комуникационе технологије при Бугарској академији наука (ИИКТ-БАН);
- (Језик и израчунавање) Ноа Гудмен, Станфорд, САД и Барбара Пленк, Универзитет у Гронингену, Холандија;



Слика 1. Постер школе

- (Језик и логика) Марта Абрусан, Национални центар за истраживања, Информатички истраживачки центар, Тулуз и Институт „Жан Никод“, Париз, Француска и Роберт Левин, Државни универзитет у Охају, САД;
- (Логика и израчунавање) Војчех Јамрога, Пољска академија наука, Варшава, Пољска.

Организациони одбор чинили су:

- Петја Осенова, Универзитет у Софији, Бугарска (СУ);
- Кирил Симов, ИИКТ-БАН;
- Галија Ангелова, ИИКТ-БАН;
- Светла Бојчева, ИИКТ-БАН;
- Владислава Григорова, ИИКТ-БАН;
- Иван Којчев, СУ;
- Ивелина Николова, ИИКТ-БАН;
- Цветомира Пашова, СУ;

- Александра Соскова, СУ;
- Ирина Темникова, СУ;
- Анели Кременска, СА;nelly Kremenska (У
- Георги Јечев, СУ.

Уз још двадесетак помагача, овај уходани тим задужен је за беспрекорну организацију школе ESLLII 2018.

3. Настава

Школа нуди готово 50 курсева. Један курс траје седам и по сати (односно, пет радних дана по сат и по времена). Сваки од курсева класификован је у једну од наредних три категорија:

- курсеви језика и логике;
- курсеви логике и израчунавања;
- курсеви језика и израчунавања.

Слично, курсеви су разврстани по нивоу, те у оквиру сваке категорије постоје основни, уводни и напредни курсеви. Сва настава одржава се на енглеском.^{1,2} Један полазник може похађати највише осам курсева (четири „слота“ у по пет дана).

Типичан радни дан у оквиру ове летње школе организован је на следећи начин:

09.00–10.30 Настава

10.30–11.00 Пауза за кафу

11.00–12.30 Настава

12.30–14.00 Пауза за ручак

14.00–15.30 Настава

15.30–15.50 Пауза за кафу

15.50–16.50 Студентска излагања

17.00–18.30 Настава

18.30–19.00 Пауза за кафу

19.00–20.00 Вечерње предавање

¹ Преглед прве седмице (на вебу)

² Преглед друге седмице (на вебу)

Поред разноврсних курсева, могуће је похађати и радионице. Радионице подразумевају наставу више оријентисану ка практичним применама. Прве седмице, организоване су радионице са следећим називима:

- Вишезначност: перспектива репрезентације и одлучивања (ориг. *Ambiguity: Perspectives on Representation and Resolution*)
- Повезивање формалне и концептуалне семантике (ориг. *Bridging Formal and Conceptual Semantics*)
- Анотирање у дигиталној хуманистици: како (рачунарски) лингвисти могу да пруже допринос (ориг. *Annotation in Digital Humanities (an-nDH): How Can Linguistics/ Computational Linguistics Help with Annotation in DH*)

Друге седмице, радионице су покривале:

- Обрада природних језика у времену великих скупова података, дубоког учења и **Post Truth?** (ориг. *NLP in the Era of Big Data, Deep Learning, and Post Truth*)
- Појам квантитета у контексту језика и мисли (ориг. *Quantity in Language and Thought*)

Свакодневно, сат времена издвојено је и за студентска излагања. Ту су полазници у петнаестак минута могли, по жељи, да представе неко своје истраживање и евентуално позову остале полазнике на сарадњу.

Током викенда између радних седмица, одржана је и конференција *Formal Grammar 2018*.³

Поред свих поменутих, стручних активности, ни друштвене активности нису занемарене. Након пријемне вечере првог дана школе, за викенд, организоване су екскурзије. Прва екскурзија водила је у Пловдив, други по величини град у Бугарској, који датира још из VI века пре нове ере. Наредног дана полазници су могли да обиђу Манастир светог Ивана рилског из X века нове ере, лоциран у Рилским планинама. На крају прве седмице, сви полазници окупили су се још једном и провеселили уз музику уживо. Друге седмице, полазници су чак заиграли и у фудбал против предавача. Преглед свих активности може се погледати **на вебу**.

³ Конференција (на вебу)

4. Одабрани курсеви

У овом одељку биће укратко описано неколико одабраних курсева према избору аутора.

4.1 Напредне регресионе методе за лингвисте

Предавач Мартин Вилинг (Универзитет у Гронингену, Холандија) одржао је овај курс у првој седмици школе. Курс је обухватао преглед регресионих метода са посебним освртом на практични аспект у језику *R*. Курс је почео предавањем о вишеструкој регресији. Након тога, два предавања обухватила су Гаусову и логистичку регресију мешовитих ефеката, које узимају у обзир структурну варијабилност иначе присутну у подацима.⁴ Последња два предавања овог курса пружила су темељни увод у генерализовано моделирање адитива, што је снажан метод за анализу нелинеарних образаца у подацима. Овај приступ је нарочито користан када је потребно анализирати податке о временским серијама (као што су ЕЕГ, праћење очију или артикулаторни подаци).

4.2 Вишечлане лексеме под лупом

Друге седмице, предавачи Карлос Рамиш (Аик Марселје Универзитет, Француска), Агата Савари (Универзитет Франсоа Рабелаис Тоурс, Француска) и Ајлин Вилавићенћио (Универзитет у Есексу, Уједињено Краљевство и Рио Гранде де Сул државни универзитет у Бразилу) одржали су изврстан курс о једној о горућих тема у рачунарској лингвистици. Циљ овог практичног курса било је пружање детаљног увода у вишечлане лексеме, покривањем у различитим језицима. Обухватао је теоријске основе, својства и смернице за њихову примену, могуће сценарије за њихов рачунски третман и технике предвиђања идиоматичности. Практичне вежбе пружиле су прилику да се користе различити алати за ручну и програмску обраду вишечланих лексема. Курс је намењен студентима и истраживачима рачунарске лингвистике који желе да анализирају и интегришу вишечлане лексеме у своје рачунарске алате и језичка истраживања.

⁴ На пример, експерименти у лингвистици често укључују учеснике који одговарају на више ставки. Ова структура мора бити доведена у модел како би се спречило прекомерно поуздане (тј. прениске) *p*-вредности.

4.3 Пробабилистичко моделовање и Бајесова анализа података у експерименталној семантици и прагматици

Курс су одржали Мајкл Франке (Универзитет у Тибингену, Немачка) и Мајкл Хенри Теслер (Станфорд, САД) у другој седмици школе. На почетку, предавачи су поставили питање: како успостављени теоријски појмови доводе до предвиђања која се могу проверити емпиријски и шта се може научити из експерименталних података о теоретским варијаблама од интереса? Овај курс обрадио је одговоре на постављена питања увођењем теоријског пробабилистичког моделирања у вези са Бајесовом анализом података као помоћног скупа алата за учење података посматрања кроз објектив теоријског модела. Предавачи су представили основе Бајесове анализе података и пробабилистичког моделирања кроз серију конкретних студија случајева у семантици и прагматици природног језика.

4.4 Специјализација векторског простора речи

Иван Вулић (Кембриџ, УК), у другој седмици наставе, упознао је полазнике са најновијим методама за изградњу векторских простора речи специјализованих за низ различитих апликација у области обраде природних језика. Модерне технике репрезентације углавном се темеље на дистрибуционој хипотези „реч се познаје по контексту“ (енг. *You shall know a word by the company it keeps*), јер се заснивају на информацијама о заједничком појављивању у великим корпусима, али и на другим врстама информација. Два су главна приступа за изградњу оваквих векторских простора речи:

- Ненадгледане методе које уче из сирових текстуалних корпуса користећи контекст речи;
- Приступи засновани на експлоатацији екстерних, расположивих ресурса, за кодирање екстрахованих информација у дистрибуционе векторске просторе.

Предавач је пружио детаљни преглед предложених метода и дискутовао о најбољим праксама за њихову унутрашњу и апликативну евалуацију.

5. Додатне могућности

Организациони одбор школе полазницима сваке године нуди и различите видове финансијске подршке. Стипендије могу да, потпуно или делимично, покривају путне трошкове, трошкове смештаја, али и само похађање школе. При одабиру стипендиста, предност имају полазници који желе да излажу у некој од студентских сесија, студенти са изврсним просеком и научним постигнућима, али и они који не могу сами да издвоје финансијска средства.

При регистрацији првог дана, сваки полазник добио је, поред сертификата, и један образац са празним пољима за потпис. Наиме, организатори ESSLLI школе охрабрују образовне институције да курсеве бодују равноправно са курсевима у оквиру својих програма. За установе које су то прихватиле, студенти могу након одслушаног курса, од предавача да затраже потпис. У зависности од броја курсева који су полазници похађали, зависи и обим ЕСПБ поена који ће добити од своје установе. Предлог организатора је да се два одслушана курса бодују као 1 ЕСПБ бод, четири курса 2 ЕСПБ бода, а шест или више курсева као 3 ЕСПБ бода. На крају је одлука матичне установе студента како ће бодовати ангажовање.

Наредна школа одржаће се у 5–16. августа 2019. у престоници Летоније, Риги.

Конференција COBISS 2018 – 30 година система

Емина Чано Томић
emina.cano.tomic@nb.rs
Народна библиотека
Србије
Београд, Србија

РАД ПРИМЉЕН: 12. децембар 2018.

РАД ПРИХВАЋЕН: 14. децембар 2018.

Конференција COBISS 2018 одржана је од 27. до 29. децембра у просторијама Института информацијских знаности у Марибору. Прва оваква конференција одржана је 1998. године, а овогодишња је у знаку својеврсног јубилеја – тридесет година од успостављана система узајамне каталогизације на просторима бивше Југославије. Институт информацијских знаности је 1988. године на основу пројекта СНТИЈ (Систем научне и технолошке информације Југославије) започео развој система узајамне каталогизације, из кога су проистекли библиотечко-информациони систем COBISS и мреже COBISS.SR¹ и COBISS.net.² Током ове три деценије многобројни стручњаци из Србије и са осталих простора бивше Југославије били су укључени у развој овог система.

Први дан конференције био је намењен библиотекарима из Словеније, док су други и трећи посвећени гостима из Албаније, Босне и Херцеговине, Бугарске, Црне Горе, Косова, Македоније, Словеније и Србије – држава чланица мреже COBISS.net У поздравном обраћању Алеш Бошњак, директор ИЗУМ-а, најавио је значајне промене у систему, откривши да се планира прелазак на COBISS4. У питању је стратешка одлука да се у току наредних неколико година цео постојећи систем пребаци на платформу са отвореним приступом. Централна тема Конференције била је школска библиотека у глобалној библиографској информационој мрежи. Ова тема је пригодно обележила завршетак пројекта укључења свих школских библиотека у Словенији у COBISS. Учесници из Словеније истакли су предности које то овим библиотекама доноси, а гост из иностранства Дороти Вилијамс, професорка на Универзитету Роберт Гордон у Шкотској, на основу

¹ COBISS.SR

² COBISS.net

својих дугогодишњих истраживања осврнула се на значај школских библиотека у процесу учења.

У излагању Давора Шоштарића (ИЗУМ) имали смо прилике да се упознамо са искуствима библиотечке заједнице Словеније у примени ГДПР уредбе о заштити података о личности. У Србији је та уредба донета још 2016. године, а после предвиђеног периода прилагођавања ступила је на снагу у мају 2018. године. За сада су у Србији ретке институције и организације које су успеле да прилагоде свој начин пословања захтевима ГДПР-а. Уредба прописује начин пописа, категоризације и кодирања информација о корисницима које су дефинисане као њихови лични подаци, па је било занимљиво чути искуства из Словеније, где се уредба већ дужи време примењује. Ове године су у Србији почеле припреме за финалну имплементацију система нормативне контроле личних имена у систему COBISS.SR. Испред ИЗУМ-а тај пројекат је координирала Гордана Мазих, а испред нашег националног COBISS Центра за узајамну каталогизацију Републике Србије координатор је био начелник центра Милорад Вучковић. У свом излагању Гордана Мазих је приказала како је изгледао читав процес: од одређивања опсега иницијалне базе, дефинисања каталогских правила и рефералних извора, дефинисања језика и писма нормативног записа, као и критеријума за иницијално повезивање у првој фази формирања иницијалне базе, до тестирања програмске опреме, исписа и иницијалног повезивања и едукације каталогизатора за рад у систему са нормативном контролом у финалној фази имплементације. На пословима формирања и редакције иницијалне базе за нормативну контролу имена радили су стручњаци библиотека оснивача ВБС од 2013, а током ове године они су лиценцирани за предаваче и успешно су извели обуку свих каталогизатора у нашем систему за рад у систему са нормативном контролом личних имена.

На конференцији је представљен и COBISS+, најновија апликација за претраживање информација библиотечких каталога преко интернета, а која ће почетком године заменити апликацију COBISS/OPAC која је уз неколико унапређивања коришћена још од 1997. године. Апликација је једноставна за употребу, навигација је логична и интуитивна, што интерфејс чини удобним за коришћење. Прилагођена је мобилним уређајима и уређајима са екранима на додир, а сада постоји и могућност подешавања интерфејса (Мој профил), могућност пријављивања са више идентитета, фасетно ограничавање погодака, и много других новина.

Последњи дан конференције, посвећен чланицама COBISS.net мреже, започео је представљањем извештаја о раду националних COBISS центара у протекле две године, да би потом уследила презентације новог решења за упис чланова библиотеке преко интернета и прегледа системских решења која се користе за заштиту мреже и података у ИЗУМ-у. Представљена је нова генерација заштитних зидова, решења за откривање системске рањивости (ВМ) и Веб апликације (ВАС), као и решење за откривање непознатих (zero day) напада.

Матјаж Залокар (ИЗУМ) представио је затим пројекат израде COBISS-ове опште листе предметних одредница – СИГ, којим је мрежа COBISS.SI ове године имплементирала нормативну базу предметних одредница. Рад на овој нормативној бази трајао је 18 година, а угледали су се на нормативне базе предметних одредница Конгресне библиотеке (SHLC) и Француске националне библиотеке (Rameau). Матјаж Залокар је 2005. године у Народној библиотеци Србије на позив центара ВБС-а представио рад на овом пројекту док је он још био у зачетку, тако да је сада било занимљиво видети како изгледа коначна имплементација.

У току конференције одржан је и редован састанак директора COBISS.net-a, као и више састанака представника COBISS.SR и ИЗУМ-а, на којима је, поред осталог, договорено укључивање COBISS+ апликације у наш систему почетком следеће године, као и имплементација нормативне контроле у марту 2019.

Све презентације доступне су на веб-страници Конференције.³ ИЗУМ је и овог пута суфинансирао трошкове превоза и преноћишта за одређен број учесника из свих земаља чланица система COBISS.net и као свих претходних година био одличан домаћин.

³ Веб-страница Конференције

УПУТСТВО ЗА АУТОРЕ

Сви радови у часопису *Инфотека* објављују се на енглеском и на српском језику у истом издању. Аутори треба да доставе своје радове на једном од ова два језика; тек након обавештења о прихватању рада, очекује се да аутор пошаље превод (за српске ауторе; за све друге ауторе часопис ће обезбедити превод са енглеског језика на српски језик). Осим у штампаном издању, сви радови се објављују и онлајн у отвореном приступу.

КАТЕГОРИЗАЦИЈА РАДОВА

За документе прихваћене за објављивање и који подлежу рецензији примењује се следећа категоризација у часопису:

1. Научни чланци:
 - Оригинални научни рад (са претходно необјављиваним резултатима сопствених истраживања научном методом);
 - прегледни рад (који садржи оригиналан, детаљан и критички приказ истраживачког проблема или подручја у коме се допринос аутора показује и аутоцитатима);
 - претходно саопштење (оригинални научни рад прелиминарног карактера и мањег обима);
 - научна расправа и осврти на одређену тему заснована на научној аргументацији.
2. Стручни чланци који представљају искуства корисна за унапређење професионалне праксе
3. Информативни прилози могу бити:
 - уводници и коментари;
 - прикази књига, рачунарског програма, база података, стандарда и сл.;
 - научног догађаја, јубилеја.

Радови класификовани као научни морају имати бар две позитивне рецензије. Ставови Редакције не морају се подударати са ставовима у објављеним радовима. Рад се не може прештамповати нити објавити под сличним насловом нити у измењеном облику.

ЕЛЕМЕНТИ РУКОПИСА

За научне и стручне радове, потребно је доставити следеће податке:

1. Радови не би требало да буду дужи од 15 страна А4 формата, писаних врстом слова Times New Roman величине 12pt. У случају дужих радова, потребно је ступити у контакт са уредницима часописа.
2. Имена и презимена свих аутора требало би да буду написана оним редом којим ће се појавити у објављеном раду.
3. Пуно име једног или више аутора, без назнака титула и диплома, треба да буде достављено уз e-mail адресу, као и уз пун, званичан назив институције којој припадају (у сложеним организацијама се наводи укупна хијерархија назива, одозго надоле).
4. Наводи се и датум слања рада.
5. Аутори треба да предложе категорију свог рада, премда коначну одлуку о категоризацији доноси главни уредник.
6. Информативни сажетак, НЕ ДУЖИ ОД 200 РЕЧИ, који језгровито представља суштину рада, циљ истраживања и примењене методе и пружа главне закључке, треба да буде достављен уз рад. Сажетак треба написати на оба језика који се користе за објављивање часописа. У сажетку, аутори треба да користе термине који се као стандардни често користе за индексирање и претраживање чланака.
7. Аутори треба да доставе најмање 3 и не више од 10 кључних речи, одвојених зарезима, које означавају главне појмове представљене у раду. Списак кључних речи треба доставити на оба језика који се користе за објављивање часописа.
8. Ако рад произилази из мастер рада или докторске тезе, аутори треба да доставе назив тог рада или тезе, као и датум када су рад или теза предати и назив надлежне институције.
9. Ако рад представља резултат учешћа аутора у неком пројекту или програму, аутори би требало да укажу захвалност институцији која је била задужена за финансирање у посебном одељку под називом „Изјаве захвалности” на крају рада, пре одељка под називом „Литература“. У истом том одељку требало би навести имена особа које су помогле у изради рада.
10. Ако је неки рад представљен на конференцији, али није изашао у зборнику радова са те конференције, то такође треба нагласити у засебној белешци.
11. Аутори могу да користе фусноте, док употреба напомена на самом крају текста није дозвољена; ипак, треба избегавати употребу предугачких фуснота. Аутори могу да убаце додатке у своје радове.
12. Реферисани материјал треба да буде излистан у одељку под називом „Литература“ на крају рада. У списку реферисаних радова аутори треба да наведу све информације које су неопходне за проналажење радова на које се реферише. У овом одељку треба да се наведу све библиографске јединице реферисане у раду – такође, овде не треба да се појави ниједна библиографска јединица која у раду није реферисана.

ПРАВИЛА УРЕЂИВАЊА ПРИХВАЋЕНИХ РАДОВА

1. Препоручује се и подржава слање радове припремљених коришћењем пакета L^AT_EX, коришћењем стила часописа (стил часописа и сви потрени пакети могу се преузети са сајта часописа). Аутори који нису упознати са коришћењем овог пакета треба да припреме свој рад у програму Word (формати .docx, .doc, .rtf или .txt). Аутори који користе програм Word на треба на посебан начин да га форматирају – пребацивање у L^AT_EX ће обавити редакција часописа.
2. Радови писани на српском језику треба да буду откуцани **ЋИРИЛИЧНИМ** писмом јер ће на том писму бити и штампани. Изузетак су једино делови текста за које је коришћење другог писма, на пример латиничног, погодније. Све врсте писама треба представити користећи Unicode UTF-8 кодни распоред.
3. Наслов рада не би требало да буде написан великим словима. Дужина наслова треба да буде у разумним границама – пожељно је да то буде мање од 100 карактера. За све наслове, аутори треба да доставе скраћене верзије тих наслова које ће бити коришћене у заглављима.
4. Искошена слова (Italic) је могуће користити за истицање у тексту, док се подебљана слова (Bold) или искошена подебљана (Italic bold) могу користити ако је то неопходно. Употребу подвучених слова (Underlined) треба избегавати. Молимо да не истичете целе реченице, нити целе пасусе.
5. Рад може бити подељен у одељке и пододељке, али не треба користити више од два нивоа наслова одељака. Сви одељци ће бити на одговарајући начин нумерисани. Додаци треба да дођу на крај рада, а ако их има више и они ће бити означени. Ако се користе листе, не треба користити више од два нивоа угњеждености.
6. Сви пасуси треба да буду раздвојени једним празним редом (једним притиском на дугме Enter).
7. Код припреме табела и слика треба водити рачуна да се радови штампају у А5 формату, тако да треба избегавати сувише широке табеле. Све илустрације треба припремити у неком од формата који сажимањем не губе на квалитету, на пример .png, .tif или .jpg и у резолуцији од најмање 300 dpi.
8. Молимо ауторе да, ако је то могуће, додају везу на екран са кога је неки снимак екрана узет. Препоручујемо да аутори приликом чувања снимка екрана, или дела екрана, користе Zoom опцију претраживача или неког другог програма. За дијаграме који су произведени помоћу програма Excel, молимо да доставите оригинални .xls документ.
9. Све табеле, илустрације, дијаграме и фотографије треба припремити као засебне датотеке, црнобеле за штампање, а у боји за онлајн верзију. Наслови испод табела, илустрација, дијаграма или фотографија треба да остану у

тексту. Свака датотека треба да носи исто име као главни текст, уз опис врсте материјала коме је додат редни број у тексту. На пример, датотека у којој се налази четврта по реду слика у раду под називом „Пример“ треба да буде именована Пример_слика_4.

10. Молимо да додате све потребне датотеке у којима су објашњени неки посебни аспекти форматирања Вашег рада.
11. Уколико се у раду појављују URL адресе веб страница на чији се садржај текст реферише, њих би требало наводити у фуснотама, уз обавезно навођење датума када је страници приступљено.

БИБЛИОГРАФИЈА И ЦИТИРАЊЕ

1. Коришћену литературу навести на крају текста, у оквиру ненумерисаног одељка Литература. Овај одељак треба да садржи сва дела наведена у тексту, и ништа више од тога. Реферисану литературу не треба наводити у фуснотама у оквиру самог текста.
2. Сва дела треба сложити по азбучном реду имена аутора, уредника или издавача (ако аутор није наведен), а уколико се јавља више дела истог аутора њих треба поређати у хронолошком редоследу.
3. За израду референци користити Чикаго стил – (Chicago Style) (www.chicagomanualofstyle.org).
4. Наводити пуне, а не скраћене наслове часописа или акрониме.
5. Сви аутори, без обзира да ли предају рукопис припремљен коришћењем L^AT_EX-а или Word-а, припремиће све библиографске јединице из одељка литература и у BibTeX према шаблонима који су наведене уз примере на сајту Инфотеке (<http://infoteka.bg.ac.rs/index.php/sr/upu-s-v-z-u-r>).