

Импресум

ЗА ИЗДАВАЧА:

проф. др Александар Јерков

Универзитетска библиотека „Светозар Марковић“

Филолошки факултет, Универзитет у Београду

office@unilib.rs

ИЗДАВАЧ:

Филолошки факултет, Универзитет у Београду

Универзитетска библиотека „Светозар Марковић“

Заједница библиотека универзитета у Србији

УРЕДНИЦИ ОВОГ БРОЈА:

др Александра Тртовац

др Бранислава Шандрих Тодоровић

aleksandra@unilib.rs

branislava.sandrih@fil.bg.ac.rs

ТЕХНИЧКИ УРЕДНИК:

др Александра Тртовац

Универзитетска библиотека „Светозар Марковић“

aleksandra@unilib.rs

УРЕДНИК ЕЛЕКТРОНСКОГ ИЗДАЊА:

др Јелена Гавриловић

Универзитетска библиотека „Светозар Марковић“

andonovski@unilib.rs

УРЕЂИВАЧКИ ОДБОР:

проф. др Александра Вранеш, проф. др Александар Јерков, проф. др Цветана Крстев, 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 Светлана 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)
ISSN 2217-9461 (е-издање)

Београд, год. 22, бр. 1, јун 2022.

ИЗРАДА ВЕБ ПОРТАЛА ЧАСОПИСА:

др Јелена Гавриловић

Универзитетска библиотека „Светозар Марковић“

ЛЕКТОР ЗА ЕНГЛЕСКИ ЈЕЗИК:

Свјетлана Ћелић

Универзитетска библиотека „Светозар Марковић“

ДИЗАЈН И ПРИПРЕМА ЗА ШТАМПУ:

др Бранислава Шандрих Тодоровић

Филолошки факултет, Универзитет у Београду

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 ИнфоШека

Булевар краља Александра 71, 11000 Београд

+381 11 3370-211

infothecajdh@gmail.com

ШТАМПА:

Мамиго плус

Београд

Часопис излази два пута годишње

Садржај

Научни радови

Унапређене верзије програма „Ка минималним паровима” Д. Алексић и Ј. Мркела	7
Примена ТХМ алата за анализу корпуса просторних планова М. Милинковић	32
Дигитални извори о историји ваздухопловства Јб. Мацура	52

Стручни радови

Европеанин модел података: кључ за сарадњу са Европеаном Т. Домазет, Н. Петровић и Т. Бутиган Вучај	73
Мултимедијални пројекат „Добродошли у мој најдражи кутак” В. Ивановић, М. Чекетић, Ј. Стојичић и Д. Галић	88

Прикази

„Дигитална трансформација и књижнице у посебним околностима” – приказ конференције О. Кринуловић и В. Петровић	99
Приказ београдске радионице European Language Grid Б. Башарагин	103

Обука за читање на даљину – Exploring EL- TeC: Use-Cases for Information Extraction and Analysis	
A. Марковић	107

Унапређене верзије програма „Ка минималним паровима”

УДК 81'322.2

САЖЕТАК: Програм КаМП налази парове речи које су на сегменталном нивоу говора неподударне само по двама одабраним факторима (Deza and Deza 2016, 215) дужине бар по 1, нпр. *пѣт* ~ *пѣћ*, *фѣлма* ~ *фѣрма*, *истѣризовати* ~ *мајѣризовати*, *пѣсничкѣ* ~ *полѣтичкѣ*. Предмет рада су брже варијанте КаМП-а са побољшаним сортирањем и са суплементарним модом.

КЉУЧНЕ РЕЧИ: фонетика, фонологија, обрада природног језика, корпусна лингвистика, Python.

РАД ПРИМЉЕН: 11. јануар 2022.

РАД ПРИХВАЋЕН: 11. мај 2022.

Данило Алексић

danilo.aleksic@fil.bg.ac.rs

Универзитет у Београду

Филолошки факултет

Београд, Србија

Лазар Мркела

lazar.mrkela@metropolitan.ac.rs

Универзитет Метрополитан

Факултет информационог

технолошког

Београд, Србија

1. Увод

Према (Bugarski 2003, 128), минимални парови су парови „kod kojih se dve po značenju odelite reči formalno razlikuju samo u jednoj fonemi“, нпр. *бѣс* ~ *чѣс*.¹ Програм Ка минималним паровима (Алексић and Шандрих 2021) у српском корпусу налази парове речи узајамно формално различитих само по задатим поднискама², при чему занемарује прозодију и то да ли су слова велика или мала. Корпус треба да буде кодиран схемом UTF-8. Осим задатих подниски, у „речима“ могу

1. Ивић (1961–1962, 75) помиње прозодијске системе „у којима се јавља на хиљаде минималних парова речи одн. облика диференцираних искључиво прозодијским контрастима [курзив – Д. А.]“. Такви парови су *вѣла* ~ *вѣла*, *лѣза* (генитив од *лѣз* ‘срећка’) ~ *лѣза* и др.

2. Уопште важи да је свака ниска подниска саме себе (Partee, Meulen, and Wall 1993, 433; Singh 2009, 33) и да дужина ниске може бити 1 (Partee, Meulen, and Wall 1993, 432; Python 2021b). Дакле, подниском не би било погрешно називати (I) ниску `"ima"` из регуларног израза `\b\S*ima\S*\b` посматрану у односу на упарену ниску `"ima"`, ни (II) карактер `"a"`.

бити (I) карактери од "A" до "Z", од "a" до "z" и од "Č" до "ž" у одговарајућим табелама Unicode-а и (II) цртице у медијалном положају.

Садржај улазне датотеке	Задате подниске	Ниска за излаз
"Klima-uređaji pre klima-uređaja"	"a", "i"	"klima-uređaja ~ Klima-uređaji" (или "Klima-uređaji ~ klima-uređaja")
"α-čestica, α-čestice, α-čestici"	"a", "e"	"čestica ~ čestice"
"α-čestica, β-čestica"	"α", "β"	"α-čestica ~ β-čestica"

Табела 1. КаМП: примери улаза и излаза

Презентовани програм може бити од користи професорима српског као страног језика и лингвистима (Алексић and Шандрих 2021, 574–75).

Област	Задате подниске	Утилизација
Настава српског као страног језика	"с", "č"	Темељ питања у вежби: „Ц или ч? 1) Шта би ти урадио, драги читао_e? Наше новине ће поштовати своје читао_e. 2) [...]“
Дериватологија	"auto", "samo"	Подаци о конкуренцији сегмената <i>ауто</i> - и <i>само</i> -.

Табела 2. КаМП: примери утилизације

Овом приликом аутори објављују и коментаришу унапређене верзије КаМП-а које су изградили, КаМП 2 и КаМП 2.1.³

3. В. Додатак 1.

КаМП 2 и КаМП 2.1 алати су из домена обраде природног језика ако се она схвати „у широком смислу“, на тај начин да подразумева „било коју врсту рачунарске манипулације природним језиком“ (Bird, Klein, and Loper 2009, ix). Наиме, дотични програми не акцентују ни мали број свих српских речи, него српски језик процесирају површно. Будући да су донекле подешени за претрагу обимних корпуса, нови КаМП-ови су скромни прилози и корпусној лингвистици ако се она дефинише нпр. као „рачунарска потпомогнута анализа врло опсежних збирки транскрибованих говорних јединица или писаних текстова“ (McEnergy and Hardie 2012, i).

КаМП 2 и КаМП 2.1 кодирани су у Python-у 3.8.2 (Python 2021a). Python је интерпретирани виши програмски језик опште намене (Pajankar 2020, 52). Овај језик је „и елегантан и прагматичан, и једноставан и моћан“; „погодан је за почетнике у програмирању и одличан за стручњаке“ (Martelli, Ravenscroft, and Holden 2017, ix). Python постаје све популарнији, и у 2017. години постао је најпопуларнији језик на свету према магазину IEEE Spectrum (Shovic and Simpson 2021, 1). Python је најшире употребљавани програмски језик у обради природног језика (Antić 2021, vii). Може се очекивати да Python буде спор у поређењу са компајлираним језицима, али је он бржи ако се у обзир не узме само колико траје извршење кода него и колико траје развијање кода (Unpingco 2021, 2). Python је у касним 80-им годинама прошлог века створио холандски програмер Гвидо ван Росум (Cicolani 2021, 41; Rajagopalan 2021, 1).

2. Сродни ресурси

У (Алексић and Шандрих 2021, 569) наведена су четири алата за налажење минималних парова одн. „фонолошких суседа“ (Mairano and Calabrò 2016, 258) која су направљена пре КаМП-а. Њима се могу прибројати програмски пакет за Python 3 Minpair (PyPI 2021) и кратки програм на Python-у 2.7 са странице (Stack Overflow 2021a).

Minpair по најмање двама задатим „вокалским фонолошким елементима“ тражи „минималне парове (и минималне скупове) [само једносложних — Д. А.] речи из америчког енглеског“, (A) `defaultdict`-ом групишући те речи према прикљученим транскрипцијама у којима је (B) задате вокале, помоћу регуларног израза и функције `enumerate()`, претходно заменио тачком.

Приступ А има начелну паралелу у КаМП-у 2.1, али (према Додатку 2) нешто ефикаснију. КаМП 2.1 речи упарује помоћу стандардног речника (в. Додатак 1).

Приступ Б има начелну паралелу у КаМП-у 2 и КаМП-у 2.1, али (према Додатку 3) много ефикаснију. КаМП 2 и КаМП 2.1 задате елементе специјалном ниском замењују помоћу метода `str.replace()` и `str.format()` (в. Додатак 1).

Унос из <code>smudict-a</code>	Торка за груписање ако су задати вокали "AE" и "OW"
("cat", ["K", "AE1", "T"])	("K", ".", "T")
("coat", ["K", "OW1", "T"])	

Табела 3. Minpair: пример улаза и торке за груписање

```

1 # Minpair: примери употребе 1 и 2
2 import minpair
3 print(minpair.vowel_minpair(["AO", "ER"])[12:13])
4 # Излаз: [{ 'AO': 'saw', 'ER': 'sir' }]
5
6 print(minpair.vowel_minpair(["AA", "AO", "EH"])[6:7])
7 # Излаз: [{ 'AO': 'dawn', 'EH': 'den', 'AA': 'don' }]

```

Могу се задати и врста речи или врсте речи.

```

1 # Minpair: примери употребе 3, 4 и 5
2 import minpair
3 print(minpair.generator(pos=["ADV"]).vowel_minpair(
4     ["AH", "EH"]))
5 # Излаз: [{ 'AH': 'once', 'EH': 'whence' }]
6
7 print(minpair.generator(
8     pos=["ADJ", "VERB"]).vowel_minpair(
9     ["AE", "IH"][:1]))
10 # Излаз: [{ 'AE': 'bad', 'IH': 'bid' }]
11
12 print(minpair.generator(
13     pos=["ADJ", "VERB"]).vowel_minpair(
14     ["AE", "IH"][22:23]))
15 # Излаз: [{ 'AE': 'sang', 'IH': 'sing' }]

```

Извор(е) речи није могуће задати. Minpair „зависи од неколиких корпуса са платформе NLTK“: „*brown*, *cmudict*, *universal_tagset* и *words*“.

Кодом понуђеним на страници (Stack Overflow 2021a) упарују се ниске које се разликују по једном карактеру, а исте су дужине. Ниске морају бити унутар нпр. листе, али се пред њих не постављају никакви природнојезички услови (није нужно да оне буду из одређеног језика, односно на одређеном писму, па ни да буду сачињене од алфабетских карактера). Током извршења кода, свака задата ниска пореди се карактер по карактер са сваком задатом ниском иза ње. Ако се поклапају сви карактери осим једног, дотични пар ниски се исписује.

```
1 # (Stack Overflow 2021a)
2 for n1, word1 in enumerate(wordlist):
3     for word2 in wordlist[n1+1:]:
4         if len(word1)==len(word2):
5             ndiff=0
6             for n, letter in enumerate(word1):
7                 if word2[n]!=letter:
8                     ndiff+=1
9             if ndiff==1:
10                 print word1, word2
11
12 """Програм са странице (Stack Overflow 2021a): пример употребе 1
13 (додао Д. А.)
14 Улаз: ["kula", "kule", "kuli", "kulom"]
15 Излаз:
16 kula kule
17 kula kuli
18 kule kuli
19 """
```

Ниједан карактер осим оног што уноси разлику не може у првом члану излазног пара бити мало, а у другом велико слово и обрнуто.

```
1 """Програм са странице (Stack Overflow 2021a): примери употребе 2 и 3
2 Улаз: ["kula", "kulE", "kuli", "kulom"]
3 Излаз:
4 kula kulE
5 kula kuli
6 kulE kuli
7
8 Улаз: ["kula", "Kule", "kuli", "kulom"]
9 Излаз:
```

```

10 kula kuli
11 ""

```

Када је у улазној листи прва ниска била дуплирана, у излазу се јавио дуплиран пар.

```

1 ""Програм са странице (Stack Overflow 2021a): пример употребе 4
2 Улаз: ["kula", "kula", "kule", "kulom"]
3 Излаз:
4 kula kule
5 kula kule
6 ""

```

	Minpair	(Stack Overflow 2021a)	КаМП 2 и КаМП 2.1
Излаз су парови ниски различитих по било ком карактеру на датој позицији.	×	✓	×
Диференцијални елементи се задају.	✓	×	✓
Број задатих диференцијалних елемената не мора бити 2.	✓		×
Задати диференцијални елементи не морају бити вокали.	×		✓
Речи не мора диференцирати низ од само једне фонеме одн. од само једног карактера.	×	×	✓
Улаз бира корисник.	×	✓	✓
Улаз не мора бити већ токенизован.		×	✓

Табела 4. КаМП 2 / КаМП 2.1 спрам сродних алата

3. Важније новине у КаМП-у 2 и КаМП-у 2.1

У припремном делу алгоритма формирају се торке за поредбу речи, нпр. ("Avali", "avali", "vli"), ("Požeškom", "požeškom", "požškom"), ("sekretarijata", "sekretarijata", "skrtrijt"). Први члан торке за поредбу јесте реч која садржи једну или обе задате подниске. Други члан је први члан пребачен у мала слова. Трећи члан је други члан у ком је бар једна потврда прве или друге задате подниске замењена ниском "v"⁴. Једној ексцерпираној речи могу одговарати једна реч са заменом задатих подниски (в. Табелу 5) или, (I) када се задате подниске преклапају (Lothaire 2005, 7) или (II) када је једна задата подниска права подниска (Böckenhauer and Bongartz 2007, 24) друге задате подниске, више речи са заменом задатих подниски (в. Табелу 6).

Ниске за поредбу		
Ексцерпирана реч	Ексцерпирана реч пребачена у мала слова	Ексцерпирана реч пребачена у мала слова уз замењивање задатих подниски
"Knjiga"	"knjiga"	"knjigv"
"knjigu"	"knjigu"	"knjigv"
"sveska"	"sveska"	"sveskv"
"SVESKU"	"svesku"	"sveskv"
"računaljku"	"računaljku"	"rčnlgk"

Табела 5. КаМП 2 / КаМП 2.1: примери ниски за поредбу (ако су задате подниске "a" и "u")

У главном делу алгоритма формиране торке се пореде. КаМП 2 генерише све могуће двочлане комбинације торки са првом задатом подниском и торки са другом задатом подниском и онда прескаче нежељене комбинације. Могуће двочлане комбинације овај програм добија тако што израчунава Декартов производ⁵ поменутих двеју група

4. Она је одабрана јер је упадљива и релативно неуобичајена. Те су особине, наравно, присутне и код многих других ниски.

5. На пример, Декартов производ скупа ниски {"број", "улица"} и скупа ниски {"МЕСТО", "ОПШТИНА"} јесте скуп уређених парова

Ексерпирана реч	Ниске за поредбу	
	Ексерпирана реч пребачена у мала слова	Ексерпирана реч пребачена у мала слова уз замењивање једне задате подниске
"ONA"	"ona"	"on ∇ "
"Onima"	"onima"	"on ∇ "
"Onima"	"onima"	"onim ∇ "
"onimima"	"onimima"	"onim ∇ "
"onimima"	"onimima"	"onimim ∇ "

Табела 6. КаМП 2 / КаМП 2.1: примери ниски за поредбу (ако су задате подниске "a" и "ima")

торки. КаМП 2.1 од торки са другом задатом подниском образује хеш-мапу (табелу) и проверава да ли се у њој срећу речи са заменом узете из торки са првом задатом подниском. Кључ мапе је реч са заменом, док је вредност мапе мапа речи од којих се добија тај исти кључ. – КаМП 2 и КаМП 2.1 исписују оне парове ексерпираних речи чији се чланови (I) разликују када се пребаце у мала слова и (II) подударају по речима са заменом, дакле парове речи које се разликују само по задатим поднискама. На пример, ако су задате подниске "a" и "u", а у улазној датотеци је само ниска "Knjiga, knjigu, sveska, SVESKU", КаМП 2 и КаМП 2.1 неће исписати ниске "Knjiga ~ SVESKU" и "knjigu ~ sveska", пошто ниска "knjig ∇ " није једнака нисци "svesk ∇ ", него ће исписати ниске "Knjiga ~ knjigu" и "sveska ~ SVESKU".

КаМП 2 и КаМП 2.1 имају (A) мод у ком игноришу разлике између великих и малих слова али фаворизују ниске малих слова и (Б) мод у ком би нпр. ексерпиране ниске "vitraž" и "Vitraž" обрадили као засебне речи (в. Табелу 7). Разлог је следећи оправдани коментар из (Алексић and Шандрих 2021, 574): „Поставља се само питање важности величине слова.“ На пример, у настави српског као страног језика властите именице некад имају приоритет над невластитим речима. Име *Чак* (*Бери, Норис...*) погодно је за вежбу изговора са фотографијама; каква би фотографија дочарала значење непроменљиве

ниски {"број", "МЕСТО"}, ("број", "ОПШТИНА"), ("улица", "МЕСТО"), ("улица", "ОПШТИНА")}.

речи *чак*? У моду Б, КаМП 2 и КаМП 2.1 из корпуса POL исписују не само пар "čak ~ žak" него и пар "Čak ~ Žak" (уз "Čak ~ ŽAK" итд.).⁶

Садржај улазне датотеке	"EUPRAVE, eUprave, euprave, EUPRAVA, eUprava, euprava"
Ниска за излаз КаМП-а	"EUPRAVA ~ EUPRAVE" (или "EUPRAVE ~ EUPRAVA")
Ниска за излаз КаМП-а 2 и КаМП-а 2.1 у моду А	"euprava ~ euprave"
Ниске за излаз КаМП-а 2 и КаМП-а 2.1 у моду Б	"euprava ~ euprave", "euprava ~ eUprave", "euprava ~ EUPRAVE", "eUprava ~ euprave", "eUprava ~ eUprave", "eUprava ~ EUPRAVE", "EUPRAVA ~ euprave", "EUPRAVA ~ eUprave", "EUPRAVA ~ EUPRAVE"

Табела 7. КаМП и КаМП 2 / КаМП 2.1: (не)разликовање великих и малих слова (ако су задате подниске "a" и "e")

Функција `segmentacija_korpusa()`⁷ реорганизована је. Она корпус више не учитава помоћу бесконачне `while`-петље,⁸ него, по угледу на пример препорученог начина за позивање функције до стражарске вредности из (Hettinger 2021, 12.27 и даље), помоћу `for`-петље, која је „брза и лепа“.

6. Велико почетно слово не гарантује да "Čak" у POL-у може бити име (због реченица као "Čak sam pronašao i kurca."). Доказе да може пружа накнадна претрага ("Čak Blekvel", "Čak Dejli", "Čak Noris"...).

7. Посреди је генераторска функција која, да би се штедела радна меморија, улазни корпус сегментира „тако да се не цепају речи“ (Алексић and Шандрих 2021, 572, 581).

8. Уп. код унет 14. маја 2021. у одговор од 11. јуна 2015. са странице (Stack Overflow 2021b).

КаМП нађене парове сортира према Unicode-кодним позицијама простих слова, док КаМП 2 и КаМП 2.1 нађене парове сортирају по позицијама простих слова у нискама `mali_alfabet` и `veliki_alfabet` (в. Табелу 8).

```

1 # Ниске за сортирање у КаМП-у 2 и КаМП-у 2.1
2 mali_alfabet = "- ~abcčćdđefghijklmnopqrsštuvwxyzž"
3 veliki_alfabet = "- ~ABCČĆDĐEFGHIJKLMNOPQRSŠTUVWXYZŽ"

```

Улазна листа	[	"nota ~ note", "đaka ~ đake", "Bač ~ Beč"	]
Улазна листа сортирана онако како КаМП сортира парове	[	"Bač ~ Beč", "nota ~ note", "đaka ~ đake"	]
Улазна листа сортирана онако како КаМП 2 и КаМП 2.1 сортирају парове	[	"Bač ~ Beč", "đaka ~ đake", "nota ~ note"	]

Табела 8. КаМП и КаМП 2 / КаМП 2.1: сортирање парова

Истина, и нови КаМП-ови сортирајући парове користе Unicode-кодне позиције, али само код карактера којих нема у нискама за сортирање (в. Табелу 9).

Пар	Листа за сортирање	Порекло броја
α	945	Unicode
-	0	Ниска
z	32	mali_alfabet
r	23	
a	3	
č	6	
e	10	
n	19	
j	15	
e	10	
	1	
~	2	
	1	
β	946	Unicode
-	0	Ниска
z	32	mali_alfabet
r	23	
a	3	
č	6	
e	10	
n	19	
j	15	
e	10	

Табела 9. КаМП 2 / КаМП 2.1: пример листе за сортирање

КаМП 2 и КаМП 2.1 сортирају чланове сваког пара пре него што их споје у ниску за излаз (нпр. ["knjigu", "Knjiga"] $\rightarrow$ ["Knjiga", "knjigu"]).

4. Брзина извршавања

Брзина је мерена у Python-у 3.8.2, на Manjaro Linux-у, рачунаром са процесором i5-11600K и два DDR4-3200 CL16 SDRAM-а од по 16 GB и на корпусу P0L, који „броји око 117.900.900 речи из 223.308 текстова са сајта *Политика*“ (Алексић and Шандрих 2021, 575).

Задате подписке	Брзина у секундама (просек пет сукцесивних мерења)				
	КаМП 2		КаМП 2.1		
	КаМП	Мод А	Мод В	Мод А	Мод В
"č", "d"	334	199	246	67	66
"dž", "đ"	135	81	85		
"nadm", "supersuper"	6639	65,79	65,51	66,38	66,09
"ir", "zir"	143	86	90	67 (!)	66
"ma", "va"	2153	1257	1516		

Табела 10. КаМП, КаМП 2 и КаМП 2.1: брзина извршавања

5. Кратка оцена ефикасности КаМП-а 2⁹

Функција која проналази парове у КаМП-у 2 заснована је на Декартовом производу две листе. Овај начин представља елегантно решење у смислу прегледности и комплексности кода, али због квадратног понашања није довољно ефикасан у случају листи са великим бројем елемената. Проблем се лако уочава из експерименталних резултата, где се примећује значајно дуже време извршавања у случају подписки које се чешће појављују (ma-va) (уп. Табелу 10).

9. Одељке 5 и 6 написао је Л. Мркела. Остале одељке (без оних двеју реченица из Одељка 3 што се тичу искључиво КаМП-а 2.1), Додатак 2 и Додатак 3 написао је Д. Алексић.

6. Даљи рад

У реалним условима, који могу захтевати да се овај програм покреће на слабијим рачунарима, само читавање корпуса и издвајање речи које садрже тражене подниске може да траје превише дуго. На пример, читавање корпуса POL.xml на једном старијем лаптоп рачунару (Acer Aspire 3, Intel Quad Core N3710, 4GB RAM) траје и до приближно 15 минута. Предлог је да се дода и опција креирања речника од корпуса који би се сачувао на диску. Ова обрада корпуса би се извршила само једном, а касније би се речник користио за проналажење парова за нове подниске.

Следећи могући корак у скраћивању времена извршавања јесте паралелизација претраге. Данас и слабији рачунари имају више „језгара“ у оквиру процесора (на пример, рачунар из претходног пасуса има 4 језгра). Зато је могуће неке делове кода паралелно извршавати и тиме додатно убрзати програм. Један предлог за једноставну паралелизацију јесте подела једне од листи речи на n делова, а затим обрада тих делова на различитим процесорима (језгрима) паралелно. Како је и секвенцијална верзија са хеширањем већ веома ефикасна, прво треба решити проблем са спорим читавањем корпуса, па тек онда размишљати о даљим убрзањима.

7. Закључак

У поређењу са КаМП-ом, КаМП 2 и КаМП 2.1 за краће време постижу више — налазе практично исте парове и нађене парове и речи унутар нађених парова још сортирају по бољем методу.

КаМП 2.1 је у већини испитаних случајева био неспорно бржи од КаМП-а 2.

Додатак 1. КаМП 2 и КаМП 2.1¹⁰

```

1  """КаМП 2.1 је модификована верзија КаМП-а 2. КаМП 2 је
2  пак модификована верзија КаМП-а.
3  Упаривање у функцијама КаМП_2_1_a() и КаМП_2_1_b()
4  допринос је Ј. Мркеле, а остатак кода (са функцијама
5  ekscep(), obrada_reči_1() и obrada_reči_2()) допринос је
6  Д. Алексића.
7  """
8
9
10 def main():
11     from functools import partial
12     from itertools import product
13     import re
14     import sys
15
16     sys.stdout.reconfigure(encoding="utf-8")
17     """В. (Алексић and Шандрих 2021, 580)."""
18     prvo_slovo = "ma".casefold()
19     drugo_slovo = "va".casefold()
20     preklapanje = False
21     razl_vel_i_mal_slova = False
22     sort_znak = chr(1114111)
23     mali_alfabet = "- abcčćddefghijklmnopqrsštuvwxyzž"
24     veliki_alfabet = "- ABCČĆDDEFGHIJKLMNOPQRSŠTUVWXYZŽ"
25
26     def spajanje_niski(*niske):
27         return "".join(niske)
28
29     def segmentacija_korpusa(
30         korpus, veličina=8192, separator="\n"):
31         """Уп. (581).
32         Функција враћа делове корпуса задате
33         величине по задатом сепаратору.
34         """
35         ostatak = ""
36         for komad in iter(

```

10. Главни извор била је званична документација, са сајта (Python 2021a). Други извори су индицирани упућивањем на линкове из Литературе, упућивањем на одговарајућа места у (Алексић and Шандрих 2021) и упућивањем на једно место у овом раду.

```

37         partial(korpus.read, veličina), ""):
38             """B. (Hettinger 2021, 12.27 и даље)."""
39             komad = spajanje_niski(ostatak, komad)
40             if separator in komad:
41                 delovi = komad.rsplit(separator, 1)
42                 """B. (W3Schools 2021)."""
43                 yield delovi[0]
44                 ostatak = delovi[1]
45             else:
46                 ostatak = komad
47         if ostatak:
48             yield ostatak
49
50     def prvo_mala_slova_1(reč):
51         """Кључ за сортирање речи који фаворизује речи
52         сачињене од малих слова.
53         """
54         if reč.islower():
55             return "!"
56         elif reč.istitle():
57             return sort_znak
58         else:
59             for slovo in reč:
60                 if slovo.isupper():
61                     reč = reč.replace(slovo, sort_znak)
62             return reč
63
64     def prvo_mala_slova_2(reč):
65         """Кључ за сортирање нађених парова. На врху
66         листе ће бити парови који садрже мање великих слова.
67         """
68         if reč.islower():
69             return "!"
70         else:
71             reč = reč.replace(" ~ ", "")
72             if reč.istitle():
73                 return sort_znak
74             else:
75                 for slovo in reč:
76                     if slovo.isupper():
77                         reč = reč.replace(
78                             slovo, sort_znak)
79             return reč

```

```

80
81 def indeksiranje_za_listu(reč):
82     """Уп. (Stack Overflow 2021c).
83     Кључ за сортирање по задатом редоследу.
84     """
85     lista_za_sort = []
86     for slovo in reč:
87         if slovo in mali_alfabet:
88             lista_za_sort.append(
89                 mali_alfabet.index(slovo))
90         elif slovo in veliki_alfabet:
91             lista_za_sort.append(
92                 veliki_alfabet.index(slovo))
93         else:
94             lista_za_sort.append(ord(slovo))
95     return lista_za_sort
96
97 def prosta_zamena_slova(reč):
98     """У речима се задате подписке замењују
99     специјалном ниском.
100    """
101     if prvo_slovo in reč and drugo_slovo not in reč:
102         reč_sa_zamenom = reč.replace(
103             prvo_slovo, "\u23B2")
104     elif prvo_slovo not in reč and drugo_slovo in reč:
105         reč_sa_zamenom = reč.replace(
106             drugo_slovo, "\u23B2")
107     elif prvo_slovo in reč and drugo_slovo in reč:
108         reč_sa_zamenom = reč.replace(
109             prvo_slovo, "\u23B2")
110         reč_sa_zamenom = reč_sa_zamenom.replace(
111             drugo_slovo, "\u23B2")
112     return (reč_sa_zamenom,)
113
114 def složena_zamena_slova(reč, slovo):
115     """У речима се задате подписке замењују
116     специјалном ниском.
117     Покривају се случајеви када између
118     задатих подписки има преклапања.
119     В. (Алексић and Шандрих 2021, 580--81).
120     """
121     izlazni_skup = set()
122     reč_za_obrađu = reč.replace(slovo, "{}")

```

```

123     for kombinacija in product(
124         [slovo, "\u23B2"],
125         repeat=reč_za_obradu.count("{}")):
126         reč_sa_zamenama = reč_za_obradu.format(
127             *kombinacija)
128         if reč_sa_zamenama != reč:
129             izlazni_skup.add(
130                 reč_sa_zamenama)
131     return izlazni_skup
132
133 def zamena_slova(reč):
134     if not preklapanje:
135         return prosta_zamena_slova(reč)
136     else:
137         skup = set()
138         skup.update(
139             složena_zamena_slova(reč, prvo_slovo),
140             složena_zamena_slova(reč, drugo_slovo))
141         return skup
142
143 def tokenizacija():
144     """Корпус се претвара у речник речи издвојених помоћу
145     регуларног израза. Избегнута је употреба врло сложених
146     регуларних израза у случајевима када су задате
147     подниске дуже (в. Табелу 10).
148     Уп. Одељак 6.
149     """
150     rečnik = {}
151     with open(r"../POL.xml",
152             "r", encoding="utf-8") as korpus:
153         komadi = segmentacija_korpusa(korpus)
154         for komad in komadi:
155             pogoci = re.findall(
156                 "[A-Za-zČ-ž-\u00ad]+", komad)
157             ""B. (573--74).""
158             for pogodak in pogoci:
159                 reč = pogodak.strip("-")
160                 if "\u00ad" in reč:
161                     reč = reč.replace("\u00ad", "")
162                 ""B. (581).""
163                 rečnik[reč] = reč.casefold()
164     return rečnik
165

```

```

166 def ekscerpt(slovo):
167     """Из речника добијеног од корпуса узимају се речи
168     које садрже задату поднику.
169     """
170     return (ključ
171             for ključ, vrednost
172             in rečnik_od_korpusa.items()
173             if slovo in vrednost)
174
175 def dekart(lista_1, lista_2):
176     """Елиминишу се нежељени парови из Декартовог
177     производа обрађених речи.
178     """
179     return (
180         (*sorted([b, e]), a, d)
181         for (a, b, c), (d, e, f)
182         in filter(
183             lambda torka: torka[0][2] == torka[1][2]
184                        and torka[0][1] != torka[1][1],
185             product(lista_1, lista_2, repeat=1)))
186
187 def obrada_reči_1(gen):
188     """Враћају се торке у којима су речи, речи пребачене
189     у мала слова и речи са заменама.
190     Ова функција се позива када се жели занемарити
191     разлика између великог и малог слова.
192     """
193     lista_torki = []
194     brojač = set()
195     lista_reči = sorted(
196         list(gen), key=prvo_mala_slova_1)
197     for reč in lista_reči:
198         reč_malim_slovima = rečnik_od_korpusa[reč]
199         if reč_malim_slovima not in brojač:
200             brojač.add(reč_malim_slovima)
201             for reč_sa_zamenom in замена_slova(
202                 reč_malim_slovima):
203                 lista_torki.append(
204                     (reč, reč_malim_slovima,
205                     reč_sa_zamenom))
206     return lista_torki
207
208 def obrada_reči_2(gen):

```

```

209     """Враћају се торке у којима су речи, речи пребачене
210     у мала слова и речи са заменама.
211     Ова функција се позива када се НЕ ЖЕЛИ занемарити
212     разлика између великог и малог слова.
213     """
214
215     lista_torki = []
216     for reč in gen:
217         reč_malim_slovima = rečnik_od_korpusa[reč]
218         for reč_sa_zamenom in замена_slova(
219             reč_malim_slovima):
220             lista_torki.append(
221                 (reč, reč_malim_slovima,
222                  reč_sa_zamenom))
223     return lista_torki
224
225 def КаМП_2_a():
226     """Завршна обрада у КаМП-у 2 ако се жели занемарити
227     разлика између великог и малог слова.
228     """
229     бројач = set()
230     for torka in dekart(
231         обрада_reči_1(ekscerp(
232             прво_slovo)),
233         обрада_reči_1(ekscerp(
234             друго_slovo))):
235         if (torka[0], torka[1]) not in бројач:
236             бројач.add((torka[0], torka[1]))
237             коначни_skup.add((torka[2], torka[3]))
238     lista_parova = [
239         " ~ ".join(sorted(list(torka),
240                             key=индексирање_за_listu))
241         for torka in коначни_skup]
242     lista_parova.sort(key=индексирање_за_listu)
243     for par in lista_parova:
244         print(par)
245     print("\n\tБРОЈ ПАРОВА:")
246     print("\t\t", len(lista_parova))
247
248 def КаМП_2_b():
249     """Завршна обрада у КаМП-у 2 ако се НЕ ЖЕЛИ занемарити
250     разлика између великог и малог слова.
251     """
252     коначни_skup = {(torka[2], torka[3])

```

```

252         for torka in dekart(
253             obrada_reči_2(ekscerp(
254                 prvo_slovo)),
255             obrada_reči_2(ekscerp(
256                 drugo_slovo))))
257 lista_parova = [
258     " ~ ".join(sorted(list(torka),
259                         key=indeksiranje_za_listu))
260     for torka in konačni_skup]
261 lista_parova = list(set(lista_parova))
262 lista_parova.sort(key=prvo_mala_slova_2)
263 lista_parova.sort(key=indeksiranje_za_listu)
264 for par in lista_parova:
265     print(par)
266 print("\n\tBROJ PAROVA:")
267 print("\t\t", len(lista_parova))
268
269 def KaMP_2_1_a():
270     """Упаривање речи и завршна обрада у KaMP-y 2.1
271     ако се жели занемарити разлика између великог
272     и малог слова.
273     """
274     lista1 = obrada_reči_1(ekscerp(
275         prvo_slovo))
276     lista2 = obrada_reči_1(ekscerp(
277         drugo_slovo))
278     mapa = {}
279     for x in lista2:
280         if x[2] not in mapa:
281             mapa[x[2]] = {}
282         mapa[x[2]][x[0]] = x[1]
283     for torka in lista1:
284         result = mapa.get(torka[2])
285         if result is not None:
286             for k, v in result.items():
287                 if (torka[1] != v and (k, torka[0])
288                     not in konačni_skup):
289                     konačni_skup.add((torka[0], k))
290     lista = []
291     for par in konačni_skup:
292         par = list(par)
293         par.sort(key=indeksiranje_za_listu)
294         izlaz = spajanje_niski(par[0], " ~ ", par[1])

```

```

295         lista.append(izlaz)
296     lista.sort(key=indeksiranje_za_listu)
297     for par in lista:
298         print(par)
299     print("Broj parova: ", len(lista))
300
301     def КаМП_2_1_b():
302         """Упаривање речи и завршна обрада у КаМП-у 2.1
303         ако се НЕ ЖЕЛИ занемарити разлика између великог
304         и малог слова.
305         """
306         lista1 = obrada_reči_2(ekscerp(
307             prvo_slovo))
308         lista2 = obrada_reči_2(ekscerp(
309             drugo_slovo))
310         mapa = {}
311         for x in lista2:
312             if x[2] not in mapa:
313                 mapa[x[2]] = {}
314             mapa[x[2]][x[0]] = x[1]
315         for torka in lista1:
316             result = mapa.get(torka[2])
317             if result is not None:
318                 for k, v in result.items():
319                     if (torka[1] != v and (k, torka[0])
320                         not in konačni_skup):
321                         konačni_skup.add((torka[0], k))
322         lista = []
323         for par in konačni_skup:
324             par = list(par)
325             par.sort(key=indeksiranje_za_listu)
326             izlaz = spajanje_niski(par[0], " ~ ", par[1])
327             lista.append(izlaz)
328         lista.sort(key=prvo_mala_slova_2)
329         lista.sort(key=indeksiranje_za_listu)
330         for par in lista:
331             print(par)
332         print("Broj parova: ", len(lista))
333
334     if (prvo_slovo[-1:] == drugo_slovo[:1]
335         or prvo_slovo[:1] == drugo_slovo[-1:]
336         or (prvo_slovo in drugo_slovo
337             or drugo_slovo in prvo_slovo)):

```

```

338     preklapanje = True
339     konačni_skup = set()
340     rečnik_od_korpusa = tokenizacija()
341     if razl_vel_i_mal_slova:
342         KaMP_2_1_b() # KaMP 2 позива функцију KaMP_2_b().
343     else:
344         KaMP_2_1_a() # KaMP 2 позива функцију KaMP_2_a().
345
346
347 if __name__ == "__main__":
348     main()

```

Додатак 2. Minpair vs. KaMP 2.1: брзина упаривања

```

1  """Приступ из Minpair-a.
2  """
3  from collections import defaultdict
4
5  mapa_1 = defaultdict(lambda: {})
6  for x in lista_2:
7      mapa_1[x[2]][x[0]] = x[1]
8
9  """Приступ из KaMP-a 2.1.
10 """
11 mapa_2 = {}
12 for x in lista_2:
13     if x[2] not in mapa_2:
14         mapa_2[x[2]] = {}
15     mapa_2[x[2]][x[0]] = x[1]
16
17 """Улаз је била листа торки типа ("subsidiaries",
18 "subsidiaries", "subsidi.ri.s"), направљена од
19 речи са "a" и/или "e" из smudict-a.
20
21 Приступ из KaMP-a 2.1 показао се око 7% бржим
22 у Python-у 3.8.2 на систему описаном у Одељку 4.
23 Поређени су просеци 500 сукцесивних мерења
24 (55 ms : 51 ms).
25 """

```

Додатак 3. Minpair vs. КаМП 2 / КаМП 2.1: брзина замењивања

```
1 """Приступ из Minpair-a.
2 """
3 vowels_regex = re.compile(r'^(?:%s)' % '|'.join(vowels))
4 matches = [vowels_regex.search(phone) for phone in word]
5 list_with_repl = []
6 for i, character in enumerate(word):
7     for j, match in enumerate(matches):
8         if i == j:
9             if match:
10                 list_with_repl.append(".")
11             else:
12                 list_with_repl.append(character)
13 string_with_repl = ".".join(list_with_repl)
14
15 """Приступ из КаМП-а 2 и КаМП-а 2.1.
16 """
17 reč_sa_zamenom = reč.replace(
18     prvo_slovo, ".")
19 reč_sa_zamenom = reč_sa_zamenom.replace(
20     drugo_slovo, ".")
21
22 """Улаз су биле речи са "a" и/или "e" из
23 studict-a.
24
25 Приступ из КаМП-а 2 и КаМП-а 2.1 показао се
26 око 95% бржим у Python-у 3.8.2 на систему
27 описаном у Одељку 4. Поређени су просеци
28 500 сукцесивних мерења (472 ms : 23 ms).
29 Међутим, мора се истаћи да у Minpair-у кôд
30 за замењивање добија листу, а враћа торку
31 (нпр. ["L", "UW", "S"] → ("L", ".", "S")),
32 док је у КаМП-у 2 и КаМП-у 2.1 и улаз и излаз
33 кода за замењивање -- ниска
34 (нпр. "teorijska" → "t❧orijsk❧").
35 """
```

Литература

- Antić, Zhenya. 2021. *Python Natural Language Processing Cookbook: Over 50 recipes to understand, analyze, and generate text for implementing language processing tasks*. Birmingham: Packt Publishing.
- Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media.
- Böckenhauer, Hans-Joachim, and Dirk Bongartz. 2007. *Algorithmic Aspects of Bioinformatics*. Berlin: Springer.
- Bugarski, Ranko. 2003. *Uvod u opštu lingvistiku*. 2nd ed. Beograd: Čigoja štampa.
- Cicolani, Jeff. 2021. *Beginning Robotics with Raspberry Pi and Arduino: Using Python and OpenCV*. 2nd ed. Berkeley, CA: Apress.
- Deza, Michel Marie, and Elena Deza. 2016. *Encyclopedia of Distances*. 4th ed. Berlin: Springer.
- Hettinger, Raymond. 2021. “Transforming Code into Beautiful, Idiomatic Python.” Accessed August 21, 2021. <https://www.youtube.com/watch?v=OSGv2VnC0go>.
- Lothaire, M. 2005. *Applied Combinatorics on Words*. Cambridge: Cambridge University Press.
- Mairano, Paolo, and Lidia Calabrò. 2016. “Are minimal pairs too few to be used in pronunciation classes?” In *La fonetica nell'apprendimento delle lingue: Phonetics and language learning*, edited by Renata Savy and Iolanda Alfano, 255–268. Milano: Officinaventuno.
- Martelli, Alex, Anna Ravenscroft, and Steve Holden. 2017. *Python in a Nutshell*. 3rd ed. Sebastopol, CA: O'Reilly Media.
- McEnery, Tony, and Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Pajankar, Ashwin. 2020. *Raspberry Pi Computer Vision Programming: Design and implement computer vision applications with Raspberry Pi, OpenCV, and Python 3*. 2nd ed. Birmingham: Packt Publishing.
- Partee, Barbara H., Alice ter Meulen, and Robert E. Wall. 1993. *Mathematical Methods in Linguistics*. Dordrecht: Kluwer Academic Publishers.

- PyPI. 2021. “minpair 0.1.3.” Accessed October 26, 2021. <https://pypi.org/project/minpair>.
- Python. 2021a. Accessed August 21, 2021. <https://www.python.org>.
- Python. 2021b. “Text Sequence Type — `str`.” Accessed August 21, 2021. <https://docs.python.org/3/library/stdtypes.html#text-sequence-type-str>.
- Rajagopalan, Gayathri. 2021. *A Python Data Analyst's Toolkit: Learn Python and Python-based Libraries with Applications in Data Analysis and Statistics*. Berkeley, CA: Apress.
- Shovic, John C., and Alan Simpson. 2021. *Python All-in-One For Dummies*. 2nd ed. Hoboken, NJ: John Wiley & Sons.
- Singh, Arindama. 2009. *Elements of Computation Theory*. London: Springer.
- Stack Overflow. 2021a. “Finding (phonological) minimal pairs with python.” Accessed August 31, 2021. <https://stackoverflow.com/q/26157361>.
- Stack Overflow. 2021b. “Lazy Method for Reading Big File in Python?” Accessed August 31, 2021. <https://stackoverflow.com/q/519633>.
- Stack Overflow. 2021c. “Sorting string values according to a custom alphabet in Python.” Accessed October 26, 2021. <https://stackoverflow.com/q/26579392>.
- Unpingco, José. 2021. *Python Programming for Data Analysis*. Cham: Springer.
- W3Schools. 2021. “Python String `rsplit()` Method.” Accessed October 26, 2021. https://www.w3schools.com/python/ref_string_rsplit.asp.
- Алексић, Данило, and Бранислава Шандрих. 2021. “Аутоматска експерција парова речи за учење изговора у настави српског као страног језика.” *Српски језик: студије српске и словенске* 26 (1): 567–584. ISSN: 0354-9259. <https://doi.org/10.18485/sj.2021.26.1.32>. <http://doi.fil.bg.ac.rs/pdf/journals/sj/2021-1/sj-2021-26-1-32.pdf>.
- Ивић, Павле. 1961–1962. “Број прозодијских могућности у речи као карактеристика фонолошких система словенских језика.” *Јужнословенски филолог* 25: 75–113.

Примена ТХМ алата за анализу корпуса просторних планова

УДК 811.163.41'322.2

Милена Милинковић

milenam@iaus.ac.rs

Институт за архитектуру

и урбанизам Србије

Београд, Србија

САЖЕТАК: У раду су представљена квантитативна текстометријска истраживања *PPTXM* корпуса просторних планова у ТХМ окружењу. Захваљујући примени развијених метода и технологија у области корпусне лингвистике и алата *SrpNER* за обележавање именованих ентитета, указано је на различите лингвистичке и статистичке карактеристике текстуалног дела планских документа. Коришћењем ТХМ алата током истраживања, добијени су резултати који су омогућили увид у учесталост појављивања различитих врста речи и именованих ентитета, како у целом корпусу, тако и њихову распоређеност и прогресију по партицијама овог корпуса.

КЉУЧНЕ РЕЧИ: ТХМ, текстометрија, именовани ентитети, дигитални корпуси, просторни планови..

РАД ПРИМЉЕН: 1. мај 2022.

РАД ПРИХВАЋЕН: 1. јун 2022.

1. Увод

Квантитативни приступ анализе текстова подразумева уочавање њихових статистичких карактеристика заснованих на унапред дефинисаним критеријумима класификације и квантификације (Вјекіћ et al. 2012). Класичан приступ квантитативне анализе текстова углавном је захтевао сарадњу већег броја истраживача, али и напоран и дуготрајан рад (Utvic 2013), што је изискивало значајне материјалне и временске ресурсе (Mehl and Gill 2010). Са појавом рачунара и

развојем савремених рачунарских технологија овакве анализе су умногоме олакшане. Без обзира да ли је у питању класичан или рачунарски подржан приступ, за било какву лингвистичку анализу неопходно је формирање адекватних и репрезентативних корпуса, како би истраживања вршена на њима могла да се примене и на друге текстове који задовољавају унапред постављене критеријуме. У лингвистици корпус се може описати као „колекција текстова за које се претпоставља да су репрезентативни за дати језик, дијалект или други подскуп језика, а који ће се користити за лингвистичку анализу“ (Francis 1975, 15). Оваква дефиниција може се применити на све корпусе, независно од тога који је носач информација у питању. Међутим, са развојем рачунарске лингвистике и формирањем електронских корпуса прихватљивија је дефиниција која под корпусом подразумева „колекцију аутентичних машински читљивих текстова који представљају репрезентативни узорак конкретног језика или језичког варијетета“ (McEnergy, Xiao, and Tono 2006, 5). Електронски корпуси су подједнако као и преелектронски корпуси, намењени за језичка истраживања, али су за разлику од њих, прилагођени аутоматској или полуаутоматској обради и анализи.

Постоје различите поделе корпуса у зависности од обима, од броја језика који су у њега укључени, од намене, домена, од носача и других карактеристика (Porović and Vitas 2003). Једна од њих је подела на корпусе опште намене и специјализоване, тј. доменске корпусе. Корпус савременог српског језика – *СрпКор* представља корпус опште намене. Његов развој започет је 1981, а најновија верзија *СрпКор2021*, објављена је 2021. године са преко 600 милиона речи. У његовом саставу налазе се и различити доменски корпуси формиран током година. Доменски корпус кулинарства (Витас and Крстев 2016; Витас 2018; Vitas 2019, 2022), доменски корпус законских текстова Народне скупштине Републике Србије (Васиљевић 2014), доменски корпус из области библиотекарства (Трговац 2017) и доменски корпус из области рударства (Обрадовић et al. 2017) само су неки од специјализованих корпуса који су током година развијали чланови Друштва за језичке ресурсе и технологије – ЈеРТех.¹ Последњи доменски корпус формиран у оквиру Друштва, крајем 2021. године је доменски корпус просторног планирања. На подскупу овог корпуса вршена су истраживања алатима *SrpNER* и *TXM* која ће бити представљена у наредним поглављима.

1. Друштво за језичке ресурсе и технологије – ЈеРТех

2. Алат ТХМ

ТХМ² је софтвер отвореног кода (Heiden 2010) чија је примена омогућена за три различита оперативна система: 64-битни Windows, Mac OS X и 64-битни Linux. Постоје две верзије овог софтвера. Десктоп верзија захтева инсталацију ТХМ софтвера после чега је кориснику омогућено да на сопственом рачунару увезе свој корпус и изврши жељене анализе, док за верзију ТХМ софтвера за веб инсталација није потребна већ корисник преко претраживача приступа различитим онлајн корпусима.

Да би текстови могли да се увезу у ТХМ окружење неопходно је да буду у одређеном формату и у адекватном кодном распореду. ТХМ алатом могуће је анализирати документа у TXT формату чистог текста или XML (*eXtensible Markup Language*) формату. Први формат мора да буде у препорученом UTF-8 кодном распореду, а други, поред наведеног кодног распореда мора да буде и у складу са TEI (*Text Encoding Initiative*)³ упутствима. Степен сложености анализа, које је могуће спровести у ТХМ окружењу, зависи од нивоа репрезентације текстова. На текстовима у TXT формату могуће је спроводити само основне анализе будући да имају најнижи ниво репрезентације. За разлику од њих, текстови кодирани у XML-TEI формату могу имати мање или више богату репрезентацију, што директно утиче на комплексност анализа које је могуће применити на текстове.

ТХМ алат пружа корисницима најразличитије технике за статистичку обраду текстова, а његово радно окружење је изразито прегледно и лако за коришћење па је стога погодно за употребу како почетницима тако и стручним лицима која се баве статистичким истраживањима. С обзиром на то да је у ТХМ окружењу могућа хијерархијска организација текстуалних објеката на којима се истраживање спроводи, анализе је могуће вршити и на целом корпусу и на издвојеном поткорпусу или партицији јер су корпуси подељени на саставне јединице: текстуалне, структуралне и лексичке. Текстуалну чине сви текстови који се у њему налазе обележени неопходним метаподацима. Структуралне јединице корпуса су поглавља, пасуси, реченице и сл. Најнижи ниво хијерархијске организације чине лексичке јединице, тј. речи од којих се текст и састоји (Јаџиновић 2019). Резултати задатих упита приказују се као конкорданце, табеле или графикони.

2. What is TXM? (Шта је ТХМ?)

3. TEI (*Text Encoding Initiative*).

3. Корпус РРТХМ

За потребе овог истраживања формиран је корпус РРТХМ⁴ који чине текстови просторних планова и представља део доменског корпуса из области просторног планирања. Планска документација чини специфичну врсту докумената карактеристичних за област просторног планирања. Постоје четири врсте просторних планова који се израђују на територији државе. То су Просторни план Републике Србије, Регионални просторни план, Просторни план јединица локалне самоуправе и Просторни план подручја посебне намене (Службени гласник РС **52/2021**). Без обзира о ком плану је реч, после доношења одлуке о изради плана, основно је утврдити обухват подручја плана, урадити анализу тренутног стања, а потом у складу са условима и смерницама који су прецизирани у планским документима вишег реда, приступити његовој изради. Цео овај процес захтева сарадњу просторних планера и научника и стручњака из многих других области (архитектуре, урбанизма, демографије, екологије, социологије и др.). У почетним фазама израде плана припрема се елаборат за рани јавни увид како би се шири јавност упознала са сврхом и циљевима израде плана, са потенцијалним концепцијама и решењима развоја конкретног подручја и понуђеним планом заштите животне средине. Потом следе припремне активности које подразумевају прибављање подлога за израду графичког дела плана и прибављање услова и података неопходних за израду нацрта планског документа који садржи текстуални и графички део. Нацрт плана подлеже стручној контроли која траје 15 дана, а након поступања у складу са извештајем о обављеној стручној контроли план се излаже на јавни увид. Током ове фазе израде плана заинтересована физичка и правна лица подnose евентуалне примедбе на плански документ искључиво у писаном облику. По завршетку израде планови иду на усвајање у надлежне институције, а потом се текстуални део плана објављује у одговарајућем службеном гласилу (Службени гласник РС **32/2019**).

Поменути корпус се састоји од шест просторних планова, четири плана подручја посебне намене, једног регионалног просторног плана и једног просторног плана јединице локалне самоуправе, а то су: *Регионални просторни план Златиборског и Моравичког управног округа*, *Просторни план подручја посебне намене Националног парка „Ђердап“*, *Просторни план подручја посебне намене слива*

4. Корпус просторних планова формиран за анализе у ТХМ окружењу.

акумулације „Грлице“, Просторни план подручја посебне намене слива акумулације „Телије“, Просторни план подручја посебне намене међународног водног пута Е-80 – Дунав (Паневропски коридор VII) и Просторни план општине Књажевца. Једини просторни план који није укључен у овај корпус је Просторни план Републике Србије, међутим, имајући у виду да су наведеним плановима покривена различита подручја на територији Србије, као и да су њима покривене све разноликости географских карактеристика Србије, може се рећи да је наведени корпус репрезентативан.

4. Текстометрија алатом ТХМ

Текстометрија је методологија анализе текстуалних података чији су развој започели француски теоретичари Пјер Гиро (Pierre Guiraud), Шарл Милер (Charles Muller), Жан-Пол Бензекри (Jean-Paul Benzécri) и др., и своју примену је нашла у квантитативним лингвистичким истраживањима (MacMurray and Leenhardt 2011), али и у многим другим хуманистичким и друштвеним наукама (Јаџиновић 2019). Текстометријском анализом, која је примењена на овај корпус, биће указано на различите лингвистичке и статистичке карактеристике текстуалног дела планских документа. Од када је основана ова метода 80-их година двадесетог века, захваљујући богатој палети алата и софтвера који се користе најпре за обраду, а потом и анализу природних језика, као и низу међународних стандарда за структурирање података и текста, она представља много више од својеврсног бројања речи (Pincemin, Heiden, and Decorde 2020). Различити видови текстометријске анализе омогућени су ТХМ софтвером. Неки од примера текстометријске анализе у ТХМ окружењу спроведени су на једном од доменских корпуса који је развијен у Друштву за језичке ресурсе и технологије. У питању је SrpELTeC корпус⁵ који је у том тренутку садржао 21 прозно дело на српском језику објављено у периоду 1840–1920. године (Trtovac, Milnović, and Krstev 2021). Мотив његове израде било је укључење текстова корпуса у вишејезичну збирку европских књижевних текстова (енгл. *European Literary Text Collection*). Добијени резултати су пружили увид у велику разноликост анализа коју је могуће вршити ТХМ алатима, истовремено указујући на специфичности употребе различитих врста речи у зависности од пола

5. Завршена колекција доступна на [SrpELTeC](#)

аутора и учесталости употребе појединих лема у текстовима SrpELTeC корпуса (Jaćimović 2019).

5. SrpNER и обележавање именованих ентитета

Један од многобројних задатака обраде природних језика је и препознавање именованих ентитета (*Named Entity Recognition* – NER). Ова врста обраде и анализе текста се односи на препознавање имена особа, организација, локација као и различитих нумеричких израза, укључујући проценте, новац, време и датуме. У скорије време технологија препознавања именованих ентитета примењује се и на екстраховање назива догађаја, производа, наслова књига и обележавање електронских адреса. Развој овог сегмента обраде природног језика траје већ скоро три деценије (Maurel, Friburger, and Eshkol-Taravella 2014), а за препознавање именованих ентитета у српском језику, у оквиру Друштва за језичке ресурсе и технологије, већ дуги низ година развијан је SrpNER систем. Начин рада и коришћења система за препознавање именованих ентитета у српском језику детаљно је описала проф. др Цветана Крстев (Krstev et al. 2014).

Пре било каквих текстометријских анализа у корпусу PPTXM обележени су именовани ентитети алатом SrpNER. За потребе овог истраживања обележавани су геополитички појмови (`top.deogr`, `top.dr`, `top.geo`, `top.gr`, `top.hyd`, `top.reg`, `top.supreg`, `top.ul`), демоними (`demonum`) и називи организација (`org.com`, `org.gen`, `org.pol`, `org.rel`). Да би алатом TXM могла да се врше истраживања у корпусу, биле су неопходне извесне корекције које су подразумевале да се из етикета за обележавање именованих етикета, тачка замени доњом цртом, тј. да се етикете преименују тако што ће етикета `org.com` бити замењена етикетом `org_com`, етикета `top.gr` етикетом `top_gr` и тако редом. Једина етикета која је остала непромењена је етикета за обележавање демонима (табела 1). Сем наведеног било је потребно елиминисати угнежђене етикете именованих ентитета, које производи обележавање алатом SrpNER, а које нису подржане у TXM окружењу. Анотација организације ЈКП „Водовод“ Зајечар алатом SrpNER извршена је на следећи начин:

```
<org.com>
  <org.gen>JKP „Vodovod“</org.gen>
  <top.gr>Zaječar</top.gr>
</org.com>
```

У оквиру спољних етикета које се користе за обележавање комерцијалних организација `<org.com>` и `</org.com>` налазе се и угнежђене етикете за опште организације `<org.gen>` и `</org.gen>` којима је обележен *ЈКП „Водовод“* и етикете за градско насеље или насељена места `<top.gr>` и `</top.gr>` којима је обележен *Зајечар*. Уклањањем угнежђених етикета пун назив организације је обележен етикетом `<org.com>`.

ТХМ етикете	објашњење етикета	Пример
demonym	именице којима је означено становништво, етничке групе и придеви изведени из географских назива	<i>београдски</i>
org_com	комерцијалне организације	<i>ЈКП „Водовод“</i>
org_gen	опште организације (организације које нису на други начин разврстане)	<i>Зајечар Институт за архитектуру и урбанизам Србије</i>
org_pol	политичке организације	
org_rel	религијски објекти (одн. организације)	<i>Црква Светог Марка</i>
top_deogr	део градског насеља или насељеног места	<i>Дедиње</i>
top_dr	држава	<i>Србија</i>
top_geo	географска обележја (низије, планине, брда, висоравни...)	<i>Авала</i>
top_gr	градско насеље или насељено место	<i>Београд</i>
top_hyd	Хидроними (реке, језера, извори...)	<i>Дунав</i>
top_reg	регион унутар државе	<i>Подунавље</i>
top_supreg	наддржавни регион	<i>Европа</i>
top_ul	улице или уопште градске локације	<i>Кнез Михаилова улица</i>

Табела 1. ТХМ етикете са објашњењима и примерима

6. Текстометријска обрада и анализа корпуса РРТХМ у ТХМ окружењу

Први корак по увозу корпуса у ТХМ окружење подразумевао је аутоматску сегментацију, токенизацију, лематизацију и на крају етикетирање врстом речи (Јаџиновић 2019). Ово је омогућено алатом

TreeTagger⁶ интегрисаним у ТХМ софтвер који повезује сваку реч у тексту са лемом и придруженим морфосинтаксичким категоријама. Овако аотиран корпус је погодан за анализу претраживачем “Corpus Query Processor” (CQP) који уместо карактера као јединицу претраге подразумева корпусну реч што пружа могућност за различита фразеолошка истраживања.

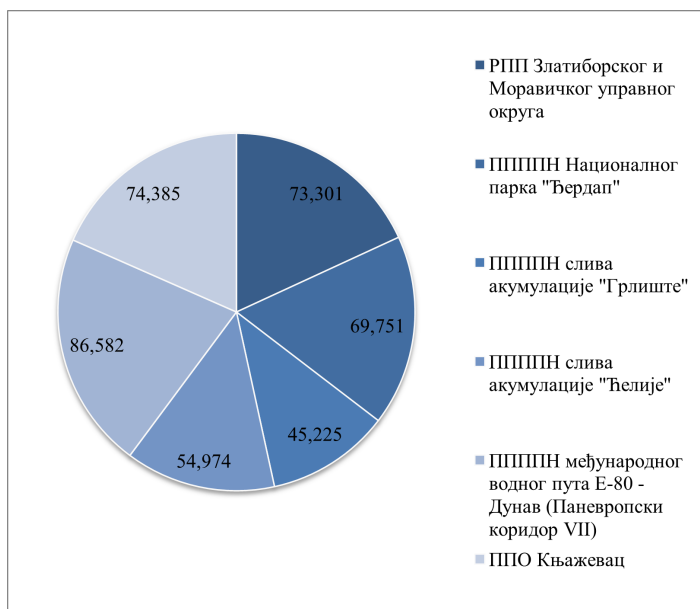
Број пасуса и реченица може се добити коришћењем CQP упита, који као резултат дају листе конкорданци. Упит за пасусе /region[p] даје као резултат 11.895 конкорданци, док се за упит /region[s] који се односи на реченице добијају 17.063 линије конкорданци.

Токенизацијом текстова утврђено је да у корпусу РРТХМ постоји 487.213 токена (табела 2). Од тог броја токена појавних облика има 32.565, а различитих лема има 14.903. Када се од укупног броја токена одузму интерпункцијски знаци, долази се до податка да се корпус састоји од 404.218 речи. Број речи као ни број токена нису подједнако распоређени по партицијама овог корпуса (слика 1, табела 2).

Назив просторног плана	Обухват плана	Димензије партиција на основу броја токена
Регионални просторни план Златиборског и Моравичког управног округа	9.184 km ²	90.724
Просторни план подручја посебне намене Националног парка „Бердап“	1542 km ²	83.058
Просторни план подручја посебне намене слива акумулације „Грлиште“	400 km ²	53.907
Просторни план подручја посебне намене слива акумулације „Ђелије“	935 km ²	66.892
Просторни план подручја посебне намене међународног водног пута Е-80 – Дунав (Паневропски коридор VII)	4.536 km ²	104.284
Просторни план општине Књажевац	1.202 km ²	88.348

Табела 2. Основне географске и језичке карактеристике корпуса РРТХМ

6. TreeTagger - a part-of-speech tagger for many languages



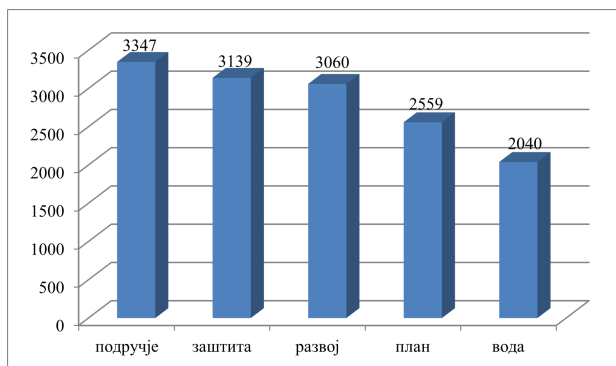
Слика 1. Екстрахован број речи по партицијама у корпусу RPTXM.

Корпус RPTXM, како је већ напоменуто, састоји се од шест планова (табела 2). Територијални обухват планова се знатно разликује од плана до плана, као и димензије текстова мерене бројем токена. Територије обухваћене плановима се крећу од 935 km² до 9.184 km², а димензије текстова од 53.907 до 104.284 токена. Међутим, уочљиво је да величина територије обухваћена планом није пропорционална димензији планских текстова, што се може видети на примеру *РПП Златиборског и Моравичког управног округа* којим је покривено 9.184 km² територије и чији план садржи 90.724 токена и *ППППН међународног водног пута E-80 – Дунав (Паневропски коридор VII)* чија територија је по површини готово дупло мања, а текст плана садржи око 14.000 токена више. Сличан пример је и са *ППППН Националног парка „Ђердап“* и *ППО Књажевац*. Први од наведених планова је по обухвату територије већи, а други садржи већи број токена.

Следећи корак у анализи корпуса представљао је утврђивање учесталости појављивања знакова интерпункције и различитих врста речи у текстовима. То је постигнуто употребом морфолошких етикета (енгл. *Part of Speech tags*) за српски језик. На основу добијених резултата

srpos	F	srpos	F
N (именица)	153.857	DET (детерминатори)	7.624
PUNCT (знак интерпункције)	82.995	ADV (прилог)	7.367
ADJ (придев)	78.039	AUX (помоћни глаголи)	6.991
ADP (предлог)	42.125	PART (речца)	5.917
CCONJ (напоредни везник)	33.368	X (остало)	3.878
PROPN (властита имена)	27.500	SCONJ (зависни везник)	2.261
VERB (глагол)	17.865	PRON (заменица)	893
NUM (број)	15.873	INTJ (узвик)	660

Табела 3. Листа учесталости свих заступљених вредности позиционог атрибута **srpos** у РРТХМ корпусу



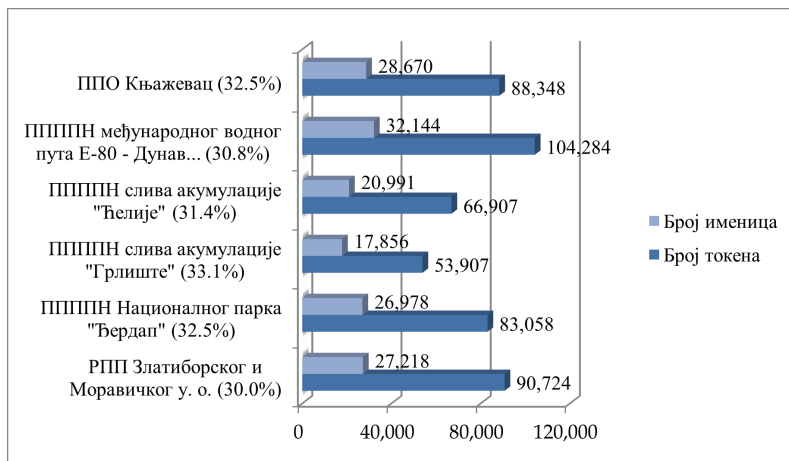
Слика 2. Приказ именица чија је фреквентност у корпусу већа од 2.000

(табела 3) може се видети је да је фреквентност именица највећа, за њима следи учесталост појављивања знакова интерпункције, а затим су наведене апсолутне фреквентности осталих врста речи. Поред укупног броја именица, упитом [**srpos**="NOUN"] утврђено је да постоји 110.587 појавних облика, као и 5.204 различите леме.

Истим упитом установљено је и које именице су најучесталије у целом корпусу, а пет најфреквентнијих је приказано на слици 2 на којој се види да је учесталост именица *подручје*, *заштита* и *развој* у корпусу већа од 3.000 и да се оне чешће користе у плановима

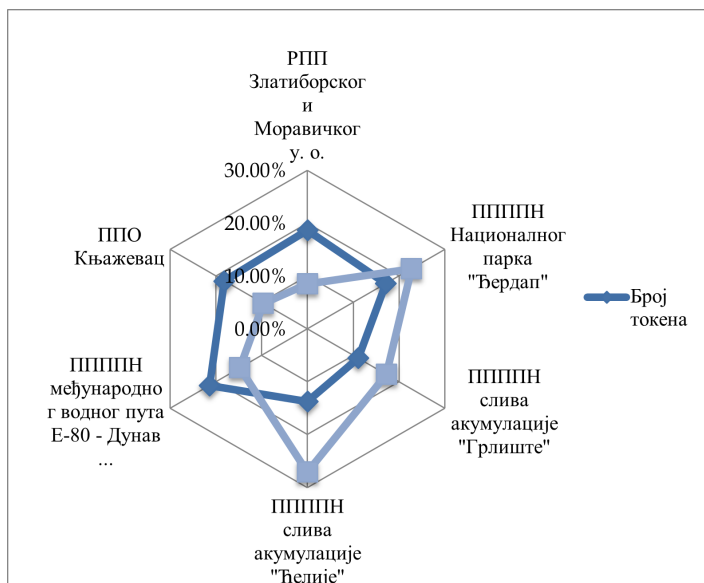
Као што је већ речено у оквиру сваке партиције постоји одређен број токена. На исти начин могуће је утврдити колико је нека од врста речи заступљена у сваком од текстова РРТХМ корпуса. Ако се посматра однос броја именица и токена по партицијама корпуса, може се закључити да је тај однос поремећен само када су у питању *РПП*

Златиборског и Моравичког управног округа и ППО Књажевац. По броју токена РПП Златиборског и Моравичког управног округа налази се на другом, а по броју именица на трећем месту, док је у случају ППО Књажевац ситуација супротна. Остали планови су задржали пропорционалан однос заступљености токена и именица (слика 3).



Слика 3. Однос заступљености токена и именица по партицијама корпуса RPTXM.

Ситуација је нешто другачија када је у питању однос токена и заменица у партицијама RPTXM корпуса. Укупан број токена, као што је већ утврђено, у целом корпусу је 487.213, а укупан број заменица добијених упитом [srpos = "PRON"] је 893. Заступљеност токена по партицијама није пропорционална заступљености заменица. На слици 4 је приказан однос заступљености токена и заменица у сваком од шест текстова просторних планова. Највећа учесталост употребе заменица је у ППППН слива акумулације „Ђелије“, мада је број токена у овом плану већи једино од броја токена у ППППН слива акумулације „Грлиште“. Са друге стране ППППН међународног водног пута Е-80 – Дунав (Паневропски коридор VII), као највећа партиција рачуната на основу броја токена, по процентуалној заступљености заменица налази се тек на четвртном месту. ППО Књажевац и ППППН Националног парка „Ђердап“, по броју токена заузимају треће, односно четврто место са

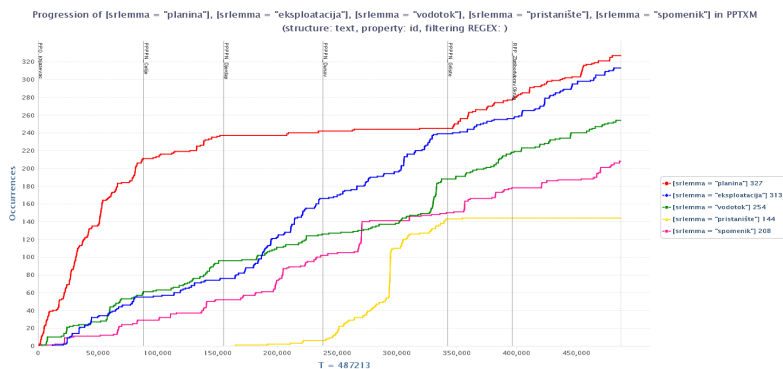


Слика 4. Однос заступљености токена и заменица по партицијама корпуса RPTXM.

разликом од свега 5.000 токена. Међутим, у *ППППН Националног парка „Ђердап“* заступљеност заменица је готово два и по пута већа него у *ПППО Књажевац*.

Досадашње анализе представљају основни облик статистичке обраде текстова, а алати у ТХМ окружењу омогућавају много сложеније анализе различитих облика учесталости и заступљености у текстовима. Оно што је карактеристично за ТХМ окружење је чињеница да је у корпусу поред укупног појављивања одређене врсте речи и фреквентности појединачних облика омогућено пратити и прогресију и кумулативну фреквенцију различитих врста речи кроз саставне делове корпуса и кроз цео корпус.

На графикану прогресије, приказаном на слици 5, дати су примери за пет именица: *планина*, *експлоатација*, *водоток*, *пристаниште*, *споменик*. Црвеном линијом је приказана кумулативна фреквентност именице *планина*. Приметно је да се она најчешће јавља у *ПППО Књажевац*, док је њено појављивање у *ППППН Националног парка „Ђердап“* (5) и *ППППН међународног водног пута Е-80 – Дунав* (Паневропски коридор VII) (3) на нивоу статистичке грешке. У



Слика 5. Приказ прогресије пет именица у текстовима РРТХМ корпуса; партиције редом одговарају ППО Књажевца, ППППН Телије, ППППН Бердап, ППППН Дунав, ПППН Грлиште, РПП Златиборски и моравички округ.

преостала три план, именица *планина*, најређе је употребљавана у ППППН слива акумулације „Телије“, нешто чешће у ППППН слива акумулације „Грлиште“, а најчешће у РПП Златиборског и Моравичког управног округа. Оваква учесталост именице *планина* по партицијама РРТХМ корпуса је и очекивана зато што и подручје ППО Књажевца, као и РПП Златиборског и Моравичког управног округа имају изражен брдско-планински рељеф.

Именица *пристаниште* (жута линија) у целом корпусу појављује се 144 пута, међутим, како се може видети на графикону, ова именица није употребљавана у ППО Књажевца, ППППН слива акумулације „Телије“ и РПП Златиборског и Моравичког управног округа, једва да је споменута у ППППН слива акумулације „Грлиште“, незнатно више у ППППН Националног парка „Бердап“, а једини просторни план у ком је присуство ове именице изражено је ППППН међународног водног пута *E-80 – Дунав (Паневропски коридор VII)*. Имајући у виду да је овим просторним планом обухваћено цело подручје Подунавља у Републици Србији, тј. читав ток Дунава у држави, потпуно је разумљиво да се именица *пристаниште* користи у свим фазама представљања Плана.

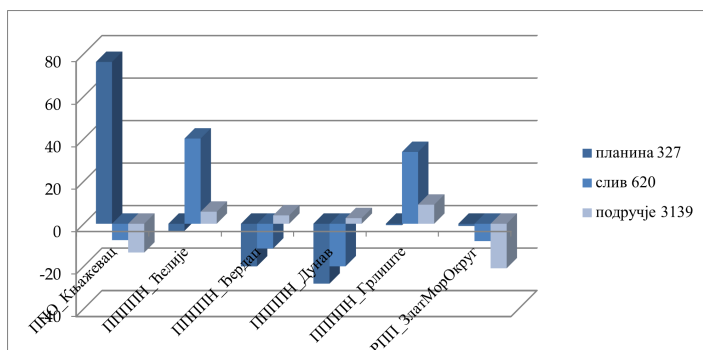
Именица *водоток* (зелена линија) је заступљена у свих шест планова и мада је, нешто чешће њено појављивање у ППО Књажевца и ППППН међународног водног пута *E-80 – Дунав (Паневропски коридор VII)*, употреба ове именице није занемарљива ни у преостала четири плана.

Будући да се цела територија Републике Србије одликује изузетним хидропотенцијалом и развијеним речним системом, оваква учесталост у свих шест текстова корпуса се могла и претпоставити.

Распрострањеност природног и културног наслеђа на територији државе Србије може се потврдити и учесталошћу појављивања именице *споменик* (ружичаста линија) на нивоу целог корпуса, али и његових саставних делова. Именица *експлоатација* (плава линија) је још једна од оних именица чије присуство је приметно у свим текстовима РР-ТХМ корпуса. Наравно, на то је утицала чињеница да Србија поседује извесне резерве различитих врста минералних сировина и да је њихова експлоатација заступљена у свим делова Републике.

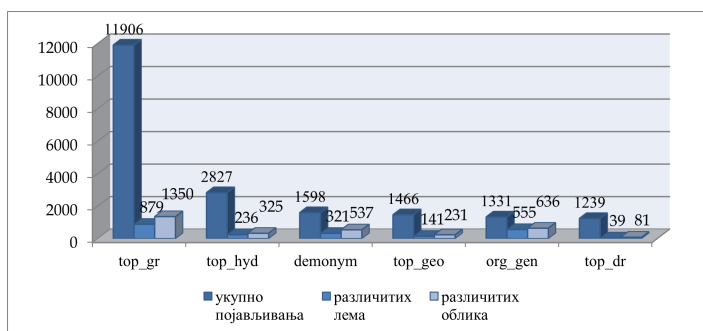
ТХМ алат омогућава још једну значајну врсту анализе која пружа могућност праћења дистрибуције учесталости појављивања одређених речи по партицијама. На слици 6 приказане су *планина*, *слив* и *подручје*. Свака од њих има различит степен заступљености у целом корпусу. Именица *подручје* је са 3.139 појављивања најзаступљенија именица на нивоу читавог корпуса, док су именица *слив* са 620 и *планина* са 327 појављивања знатно мање коришћене у планским документима. Резултати, након обраде у ТХМ окружењу, показују изразито високу фреквентност именице *планина* у *ППО Књажевац*, значајно изнад просека у односу на њену употребу на нивоу целог корпуса, док је њена фреквентност испод просека у преосталих пет просторних планова, а посебно у *ППППН међународног водног пута Е-80 – Дунав (Паневропски коридор VII)*. Коришћење именице *слив* најизраженије је у *ППППН слива акумулације „Телије“*, а позитиван скор има и у *ППППН слива акумулације „Грмиште“*. У сва четири преостала плана коришћење ове именице је испод просека, а као и код именице *планина*, то је најизраженије у *ППППН међународног водног пута Е-80 – Дунав (Паневропски коридор VII)*. Именица *подручје* негативну заступљеност, у односу на цео корпус, има у *РПП Златиборског и Моравичког управног округа* и *ППО Књажевац*. У осталим партицијама заступљеност ове именице је изнад просека у односу на цео текст.

Имајући у виду чињеницу да је у ТХМ платформу увезен корпус који је претходно обележен етикетама именованих ентитета, па је у њему, поред утврђивања фреквенције различитих врста речи, омогућена и анализа којом се утврђује и степен заступљености именованих ентитета, коришћењем опције *Index* и постављањем различитих упита као што су */region[top_gr]*, */region[top_hyd]*, */region[demonym]* и др. Слика 7 приказује шест најфреквентнијих именованих ентитета



Слика 6. Специфичност употребе именица у PPTXM корпусу.

у PPTXM корпуса. Поред укупне фреквентности приказан је и број различитих лема и различитих облика за сваки именовани ентитет. Убедљиво најзаступљенији тип именованог ентитета је *top_gr*, којим су обележена градска насеља и друга насељена места.



Слика 7. Најзаступљенији именовани ентитети у PPTXM корпусу.

Упитом `/region[org_rel]` утврђено је да се у целом корпусу појављују 33 различита облика именованих ентитета *org_rel* којим су обележени религијски објекти и да је за њих евидентирано 29 различитих лема са укупно 56 појављивања. Најфреквентнији су *Манастир Милешева* (6), *Црква Св. Петра* (4), *Црква Св. Николе*, *Св. Тројице*, *Св. Ђорђа* и *Манастир Рача* са три појављивања, затим осам религијски објеката који се у корпусу појављују два пута и преостали

именовани ентитети *org_rel* који се јављају само једном. Међутим, у ТХМ окружењу омогућено је постављање и сложенијих упита. Тако се упитом `/region[org_rel][0,2]/region[top_gr]` утврђује фреквентност именованих ентитета *org_rel* после чијег назива следи и именовани ентитет *top_gr*. Као резултат се добија потпуно другачији редослед у броју појављивања религијских објеката, као што је и приказано на слици 8. У овом случају, као најфреквентније су излистане *Црква Св. Илије у Брекову*, *Светог Николе* у Брезови и *Светог Саве* у Савинцу.

Query 		
<code>/region[org_rel][0,2]/region[top_gr]</code>		
word	Frequency	
crkva Sv . Ilije u Brekovu	2	
crkva Svetog Nikole u Brezovi	2	
crkva Svetog Save u Savincu	2	
crkva Sv . Bogorodice u Donjoj Kamenici	1	
crkva Sv . Nikole - Stari Slankamen	1	
crkva Sv . Preobraženja , Pančevo	1	
crkva Svete Bogorodice u Donjoj Kamenici	1	
crkva Svete Bogorodice u Prilipcu	1	
crkva Svete Trojice u Bistrici	1	
crkva Svetog Đorđa u Godoviku	1	
crkva Svetog Đorđa u Knjaževcu	1	
crkva Svetog Đorđa u Užicu	1	
manastira Mileševo (Prijepolje	1	
manastir Vojlovica , Pančevo	1	

Слика 8. Учесталост појављивања религијских објеката на основу упита `/region[org_rel][0,2]/region[top_gr]`.

7. Закључак

У раду су приказане разноврсне лингвистичке и статистичке анализе, коришћењем ТХМ алата, на текстовима просторних планова у којима су претходно обележени именовани ентитети коришћењем *SrpNER*-а, система за препознавање именованих ентитета у српском језику. Указано је на заједничке карактеристике текстова планске документације, као и на извесне специфичности појединачних просторних планова, који су

као део доменског корпуса просторног планирања уврштени у српски корпус *СрпКор2021*, који тренутно садржи више од 600 милиона речи.

Просторни планови, који представљају главни фокус овог рада, чине специфичну врсту докумената карактеристичних за област просторног планирања. Како би била омогућена детаљнија истраживања ових текстова, креиран је корпус РРТХМ који обухвата шест докумената којима су обрађени различити региони Републике Србије: Златиборски и Моравички управни округ, сливови акумулација „Грлиште“ и „Телије“, Национални парк Ђердап, Међународни пловни пут Е-80 – Дунав (Паневропски коридор VII) и општина Књажевац.

Анализом је обухваћено преко 400.000 речи, чија се заступљеност креће од 11,18% у *ППППН слива акумулације „Грлиште“* до 21,42% у *ППППН међународног водног пута Е-80 – Дунав (Паневропски коридор VII)*. Корпус садржи више од пола милиона токена, од којих трећину чине именице.

Треба напоменути да, сем приказаних анализа, ТХМ окружење пружа и друге могућности и да њихова употреба, као и употреба представљених текстометријских приступа зависе од унапред постављених захтева као и од потреба и склоности истраживача. Оно што је неоспорно је да ТХМ платформа омогућава приказ особености корпуса у целини, као и његових саставних делова. Из тог разлога је примена ових метода веома значајна за разнолике квантитативне анализе и утврђивање карактеристичних обележја текстова. Неки од представљених резултата, као што су заступљеност конкретних имена у целом корпусу и учесталост њиховог појављивања у појединачним плановима, могли су се и претпоставити на основу подручја која су планом обухваћена. Међутим, за добијене резултате о заступљености заменица по партицијама корпуса не може се дати неко логичко објашњење само на основу сазнања о обухвату планова, већ је за њихово разумевање неопходно спровести додатна истраживања и анализе.

Захвалност

Ово истраживање је вршено за потребе израде докторске дисертације пријављене на Филолошком факултету Универзитета у Београду под називом „Развој библиотечких и језичких ресурса за организовање и проналажење информација о просторном планирању“. Алати коришћени током истраживања развијани су (SrpNER) или су прилагођени за

различита лингвистичка истраживања за српски језик (ТХМ) у оквиру Друштва за језичке ресурсе и технологије. Посебну захвалност ауторка дугује професоркама др Цветани Крстев и др Ранки Станковић без чијег личног ангажовања, истраживања представљена овим радом не би била могућа.

Литература

- Bjekić, Jovana, Ljiljana Lazarević, Milica Erić, Elena Stojimirović, and Teodora Đokić. 2012. "Razvoj srpske verzije rečnika za automatsku analizu teksta (LIWCser)." *Psihološka istraživanja* XV (1): 85–110. <https://doi.org/10.5937/PsIstra1201085B>.
- Francis, W Nelson. 1975. "Problems of Assembling, Describing, and Computerizing Corpora." *Research Techniques and Prospects. Papers in South-west English*, (1), 15–38.
- Heiden, Serge. 2010. "The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme." In *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*, 389–398. Tohoku University, Sendai, Japan: Institute of Digital Enhancement of Cognitive Processing, Waseda University. <https://aclanthology.org/Y10-1044>.
- Jaćimović, Jelena. 2019. "Textometric methods and the TXM platform for corpus analysis and visual presentation." *Infotheca - Journal for Digital Humanities* 19 (1): 30–54. <https://doi.org/10.18485/infotheca.2019.19.1.2>.
- Krstev, Cvetana, Ivan Obradović, Miloš Utvić, and Duško Vitas. 2014. "A system for named entity recognition based on local grammars." *Journal of Logic and Computation* 24 (2): 473–489. <https://doi.org/10.1093/logcom/exs079>.
- MacMurray, Erin, and Marguerite. Leenhardt. 2011. "Textometry and Information Discovery: A New Approach to Mining Textual Data on the Web." In *Proceedings of the 2011 International Conference on Artificial Intelligence, ICAI 2011*, edited by H.R. Arabnia, D. Fuente, E.B. Kozrenko, and J.O. Olivas, II:605–611. Las Vegas: Worldcomp'11.

- Maurel, Denis, Nathalie Friburger, and Iris Eshkol-Taravella. 2014. "Enrichment of Renaissance Texts with Proper Names." *Infotheca* 15 (1): 29a–41a. <https://infoteka.bg.ac.rs/index.php/en/archives/2014/1/infoteka-15-1-2014-30-41>.
- McEnery, Tony, Richar Xiao, and Yukio Tono. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge. <https://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xCBLS/CBLS.htm>.
- Mehl, Matthias R., and Alastair J. Gill. 2010. "Automatic text analysis." In *Advanced methods for conducting online behavioral research*, edited by S. D. Gosling and J. A. Johnson, 109–127. Washington: American Psychological Association. <https://doi.org/10.1037/12076-008>.
- Pincemin, Bénédicte, Serge Heiden, and Matthieu Decorde. 2020. "Textometry on Audiovisual Corpora Experiments with TXM software." In *JADT 2020 : 15es Journées internationales d'Analyse statistique des Données Textuelles*. http://lexicometrica.univ-paris3.fr/jadt/JADT2020/jadt2020_pdf/PINCEMIN_HEIDEN_DECORDE_JADT2020.pdf.
- Popović, Ljubomir, and Duško Vitas. 2003. "Konspekt za izgradnju referentnog korpusa srpskog standardnog jezika." In *Naučni sastanak slavista u Vukove dane*. Beograd, Novi Sad: MSC.
- Trtovac, Aleksandra, Vasilije Milnović, and Cvetana Krstev. 2021. "The Serbian Part of the ELTeC - from the Empty List to the 100 Novels Collection." *Infotheca - Journal for Digital Humanities* 21 (2): 7–25. <https://doi.org/10.18485/infotheca.2021.21.2.1>.
- Utvić, Miloš. 2013. "Izgradnja referentnog korpusa savremenog srpskog jezika." PhD diss., Univerzitet u Beogradu, Filološki fakulte. <https://nardus.mpn.gov.rs/handle/123456789/4091>.
- Vitas, Duško. 2019. "Food as Text." *Infotheca - Journal For Digital Humanities* 19 (2): 139–161. <https://doi.org/10.18485/infotheca.2019.19.2.7>.
- Vitas, Duško. 2022. "From Onions to Champagne – Food and Drink in the SrpELTeC Corpus." *Infotheca - Journal for Digital Humanities* 21 (2): 88–118. <https://doi.org/10.18485/infotheca.2021.21.2.5>.
- Васиљевић, Небојша М. 2014. "Аутоматска обрада правних текстова на српском језику." PhD diss., Универзитет у Београду, Филолошки факултет. <https://nardus.mpn.gov.rs/handle/123456789/4091>.

- Витас, Душко, and Цветана Крстев. 2016. "Оглед из гастрономатике [Experiments in gastronomatics]." In *Теме језикословне у србистици кроз дијахронију и синхронију [Linguistic topics in Serbian through diachrony and synchrony]*, edited by Јасмина Дражић, Исидора Бјелаковић, and Дејан Средојевић, 1–10. Novi Sad: Филозофски факултет.
- Витас, Душко М. 2018. "Храна из нежељене поште : (анатомија језика брзе хране [Spam Food: (Anatomy of Fast Food Language)])." In *Српски језик и његови ресурси: теорија, опис и примене Научни састанак слависта у Вукове дане*, edited by Божо Ђорић and Александар Милановић, 21–35. Београд: Међународни славистички центар, Филолошки факултет, Универзитет у Београду. <https://doi.org/10.18485/msc.2018.47.3.ch2>.
- Обрадовић, Иван, Александра Томашевић, Ранка Станковић, and Лазић Биљана. 2017. "Увођење доменских и семантичких маркера за област рударства у српске електронске речнике." In *Научни састанак слависта у Вукове дане - Српски језик и његови ресурси: теорија, опис и примене*, edited by Рајна Драгићевић and Александар Милановић, 147–158. Београд: Међународни славистички центар на Филолошком факултету. <https://doi.org/10.18485/msc.2017.46.3.ch10>.
- Службени гласник РС. 52/2021. *Закон о планирању и изградњи*. Бр, 72/2009, 81/2009 - испр., 64/2010 - одлука УС, 24/2011, 121/2012, 42/2013 - одлука УС, 50/2013 - одлука УС, 98/2013 - одлука УС, 132/2014, 145/2014, 83/2018, 31/2019, 37/2019 - др. закон, 9/2020 и 52/2021), 52/2021.
- Службени гласник РС. 32/2019. *Правилнику о садржини, начину и поступку израде докумената просторног и урбанистичког планирања*. Бр. 32/2019), 32/2019.
- Трговац, Александра. 2017. "Аутоматизација библиотека у Србији - историјски преглед." *Библиотекар*, no. 2, 101–114. <https://bds.rs/wp-content/uploads/2018/01/bibliotekar-2-2017-7.pdf>.

Дигитални извори о историји ваздухопловства

УДК 004.738.5:027]:656.7(091)

Љиљана В. Мацура
ljiljana.macura@gmail.com
*Народна библиотека Србије
Београд, Србија*

САЖЕТАК: У раду се описују дигитални ресурси историје ваздухопловства и истражује да ли постоји дигитална библиотека историје ваздухопловства. Предмет истраживања представљају историја ваздухопловства у општем смислу и историја ваздухопловства Србије. Истраживања спроведена од априла 2019. до новембра 2021. године показала су да не постоји јавно доступна дигитална библиотека историје ваздухопловства неке државе или нације, већ да постоје ресурси, односно, веб-странице на којима се могу наћи различити дигитализовани артефакти историје ваздухопловства.

КЉУЧНЕ РЕЧИ: историја ваздухопловства, дигиталне библиотеке, културна баштина, метаподаци.

РАД ПРИМЉЕН: 23. март 2022.

РАД ПРИХВАЋЕН: 4. април 2020.

1. Увод

Фокус овог рада јесу мрежни извори којима је основна тежња неговање историје ваздухопловства.¹ Током рада нису пронађена слична истраживања, ни теоријски радови из области дигитализације историје ваздухопловства, али је пронађен велики број мрежних извора који су потпуно или делимично, стручно или тематски опредељени за историју

1. Текст је настао у току израде докторске дисертације „Услови за стварање и иницијална грађа дигиталне библиотеке историје ваздухопловства Србије“, пријављене 2019. године на Филолошком факултету Универзитета у Београду.

ваздухопловства. Док једни приказују историју ваздухопловства у свету, други се баве историјом домаћег ваздухопловства или историјом ваздухопловства одређеног региона. Рад је подједнако намењен стручњацима из области библиотекарства и информационих наука, колико и из ваздухопловства. Тема може бити занимљива и ширем аудиторијуму, стручњацима из саобраћајне делатности у општем смислу, ентузијастима и публици која има афинитете према ваздухопловству, а циљне групе читалаца из области ваздухопловства јесу особе које су у фази образовања, предавачи и ваздухопловци по занимању.

Потребу постојања претпостављене дигиталне библиотеке историје ваздухопловства потврђује мноштво различитих дигиталних извора који садрже једно заједничко, а то је неговање традиције ваздухопловства. Осим организације ваздухопловних манифестација, неговању поменуте традиције доприноси и дигитализација различитих докумената. Дигитализују се текстуални документи, звучни и видео-записи, различита визуелна грађа, а може се претпоставити и да се чувају на исти начин и рачунарски програми за потребе ваздухопловства, рецимо, софтвери произвођача авиона и авио-компанија за симулаторе лета или авио-компанија за ваздухопловну комерцијалну делатност. У зависности од циљева претпостављене дигиталне библиотеке и циљне групе њених корисника, бирају се врста и обим грађе, као и врста услуга које би та библиотека пружала. У појединим случајевима у пракси, у већој мери су дигитализоване публикације које садрже ваздухопловно-техничке информације, док се у другим више дигитализује грађа као што су фотографије, цртежи и белешке које нагињу друштвеној историји ваздухопловства. Постоје дигитални извори из историје ваздухопловства које искључиво дигитализују фотографије, док неки други у великој мери дигитализују звучне и видео-записе. Постоје и дигитални извори општег типа чије колекције из ваздухопловства чине плакати, екслибриси, разгледнице са мотивима из ваздухопловства, а на светској мрежи се могу пронаћи и дигитализовани каталогски листићи² – библиографско-каталогски записи садржајно или тематски везани за ваздухопловство. У изворима општег типа често се проналазе информације из ваздухопловства, али је успех у претраживању много већи уколико је реч о ретроспективној конверзији каталога.

Нажалост, до сада нису пронађени теоријски радови о дигитализацији историје ваздухопловства, односно о дигиталним библиотекама историје ваздухопловства или о збиркама из историје

2. Реч је о ретроспективној конверзији каталога.

ваздухопловства. Ипак, пратећи елементе IFLA/UNESCO манифеста за дигиталне библиотеке³ може се указати на значај дигиталних извора или претпостављене дигиталне библиотеке историје ваздухопловства. Помажући у организацији и „дисеминацији информација“ дигиталне библиотеке имају у друштву „стратешку улогу“. Тако, дигитални извори историје ваздухопловства чувају и представљају грађу која је културно и научно наслеђе. Претпостављена дигитална библиотека историје ваздухопловства може представљати на изванредан начин и систематизовану историју ваздухопловства и организацију знања из те области. Као „збирка дигиталних објеката потврђеног квалитета“ и „услуге неопходне за проналажење и коришћење“ дигиталних извора, претпостављена дигитална библиотека омогућава систематизовану информацију на „брз и економски приступачан начин“. Информације намењене истраживачима, предавачима и особама у процесу образовања, као и широј јавности (особама ван ваздухопловне струке), које стављају на тај начин ваздухопловну делатност у одређени друштвени контекст, у случају наше друштвене стварности освешћују значај и упоредивост ваздухопловства Србије на глобалном нивоу.

2. Дигитални извори историје ваздухопловства – ток истраживања

У дигиталним медијима постоји мноштво извора који се баве историјом ваздухопловства на различите начине. Истраживање је спроведено преко претраживача Google, различитом динамиком, од априла 2019. године до новембра 2021. године. Обављено је упитима који садрже следеће термине: AERO*, AERONAUTIC, AIR FORCE, AIR TRAFFIC, AIR TRANSPORT, AIR*, AVIATION HISTORY, AVIATION HISTORY AND DIGITAL LIBRARY, AVIATION MUSEUM, AVIATION, CIVIL AVIATION AUTHORITY, CIVIL AVIATION, DIGITAL LIBRARY, FLIGHT, FLYING, HISTORY, LIBRARY, MUSEUM и TRADITION.

Истраживање је започето упитима AVIATION HISTORY, те AVIATION HISTORY AND DIGITAL LIBRARY и пружило је мноштво несистематизованих резултата. Увођење термина AERO*, AERONAUTIC, AIR FORCE, AIR TRAFFIC, AIR TRANSPORT, AIR*, AVIATION MUSEUM, AVIATION, CIVIL AVIATION AUTHORITY, CIVIL AVIATION, DIGITAL LIBRARY, FLIGHT, FLYING, HISTORY, LIBRARY, MUSEUM и TRADITION

3. У пасусу се сви наводи односе на поменути текст (IFLA/UNESCO 2012)

пружило је, такође, мноштво резултата. На крају, истраживање је завршено увођењем параметра државе. Назив државе на енглеском језику комбинован је са терминима AIR FORCE, CIVIL AVIATION AUTHORITY, AVIATION MUSEUM и DIGITAL LIBRARY.

Коришћење последње од наведених методологија пружило је најбоље резултате. Ипак, претраживање по том принципу није могло да се примени у потпуности, на све државе, односно, територије на које се односило. Уколико није било другог начина да се добије одговор, претрага је обављана по параметру државе на различитим мрежним странама (*Global Security*, *Flickr*, *Flock Aero*)⁴ које су се јављале у одзиву приликом неког од претходних покушаја (истраживања). На крају је добијен списак организација, установа или друштава који на различите начине приказују историју ваздухопловства у електронском облику.

3. Резултати истраживања

Већ половином деведесетих година XX века било је веб-страница које су наставиле дотадашњу традицију неговања историје ваздухопловства. У електронском облику најчешће су се налазили текстуални записи и фотографије, али су се с временом у садржаје укључивали и записи са других медија. Данас је на светској мрежи представљена историја ваздухопловства која је изворно чувана на дводимензионалним и тродимензионалним медијима, звучним и видео-записима, тј. у виду слике, текста, звука и видеа. Историју ваздухопловства у електронској форми представљају и појединачне ставке на веб-страницама музеја или библиотека, као и веб-странице креиране искључиво ради презентовања историје ваздухопловства.

Дигитални извори историје ваздухопловства представљени у овом раду јесу поједини примери мрежних извора из историје ваздухопловства. Представљени су: називом, територијалним обухватом, односно, историјом домаћег, регионалног или ваздухопловства у свету, начинима на које се врши претрага, начинима на који су представљени резултати претраге, коментаром у погледу функционалности странице, условима за регистрацију и могућностима које пружа регистрација на страници уколико је понуђена или неопходна.

4. О страницама *Global Security* и *Flock Aero* биће више речи у даљем тексту. Извор *Flickr* неће бити описиван у раду јер није усмерен на ваздухопловство него је општег типа.

Веб-странице које следе приказане су у азбучном редоследу према држави порекла,⁵ а резултати претраге – дигитални објекти описани су метаподацима који у погледу опсега варирају (Тртовац 2017).

3.1 Велика Британија

Heritage Collection потиче са веб-странице Аеронаутичког друштва Велике Британије (Royal Aeronautical Society 2021). Садржи 14 тематских колекција и седам биографија особа које представљају пионире теорије и праксе британског ваздухопловства. Истраживаче и друге посетиоце стране *Heritage Collection* упућује како да користе ресурсе поштујући ауторска права. Посебан акценат стављен је на објављивање фотографија, цртежа и других дигитализованих садржаја са странице што није спорно уколико се позива на извор, тј. на страну *Heritage Collections*. Претраживање је једноставно, уносом упита у предвиђено поље. На приступној страни приказани су најтраженији објекти (Top Terms), систематизовани по терминима (Taxonomy Terms), односно, метаподацима и бројеви ставки које су њима обележене. Другим речима, приказане су збирке фотографија, цртежа и бележница, модели авиона, конструкције авиона и једрилица, погонске јединице; збирке које се односе на особе и науке од значаја за историју пионирског доба ваздухопловства. Сваки изабрани објекат садржи (осим дигиталног документа) метаподатке који се односе на сâм дигитални докуменат (Item Details) и на збирку (Collection Details) којој припада. То су следећи метаподаци: опис (Description), време (Date), а често и различити термини (Taxonomy Terms) – за најмању јединицу грађе дигиталне библиотеке (дигитални објекат), тј. назив колекције (Collection), опис (Description) и термини (Collection Terms) – за колекцију, при чему су метаподаци за колекцију увек исти (описују исту колекцију). Страница је прегледна и функционална. На главној страни Друштва постоји могућност регистрације.

Aeroflight (Aeroflight, n.a.) је страница регистрована у јуну 1996. године, а од октобра 2019. године функционише и као клауд (енгл. cloud, cloud computing) – *Aeroflight Cloud* (Aeroflight Cloud, n.a.), представљајући простор за прикупљање и дисеминацију информација из ваздухопловства – актуелних информација и информација из историје ваздухопловства, које се односе на ратно, спортско и

5. Азбучни редослед односи се на стране земље. Извори у Србији дати су након прегледа иностраних извора.

цивилно ваздухопловство, музеје, установе и организације, власнике ваздухоплова и сл. За потребе странице, на основу краудсорсинг активности, ствара се база података о монографским публикацијама, примарно на енглеском језику. Реч је о публикацијама које имају ISBN и које су објављене до 2000. године, без обзира на изворни облик. База је креирана у програму Microsoft Excel, састоји се од три листа: са насловом, дефиницијама атрибута и вредностима атрибута. Примећује се велика међусобна сличност атрибута метаподатака који описују дигиталне документе на страници *Aeroflight* са онима у библиографским базама које користе неки од MARC формата. Осим базе монографских, постоји и база серијских публикација. Резултати претраживања представљају чланке, листе и везе ка другим страницама. На *Aeroflight*-у се могу наћи и садржаји из историје нашег домаћег ваздухопловства и то под одредницом Југославија и Србија. Страница је оптерећена разнородним садржајима у оквиру тематике, али је њена функционалност ипак сачувана. На овој страници постоји могућност регистрације.

Flock Aero (FltCtrl Ltd, [n.a.](#)) је берза авио-служби у свету. За овај рад занимљива је њена страница о цивилним ваздухопловним властима, и то у смислу трагова о постојању ваздухопловства у појединим државама или територијама (таб Countries на (FltCtrl Ltd, [n.a.](#))) у случајевима када претрага по метаподацима (упитним терминима као што су AVIATION HISTORY AND DIGITAL LIBRARY или само AVIATION HISTORY) није пружала резултате.

Страница *Air Team Images* регистрована је 2003. године (AirTeamImages, [n.a.](#)). Њен покретач је уједно и реномирани фотограф, фокусиран на ваздухопловство. *Air Team Images* представља онлајн библиотеку фотографија летелица, која је постављена да посетиоцу странице или комитенту пружи што квалитетнију „авио-фотографију“ (“About Us” на (AirTeamImages, [n.a.](#))). Иако је акценат на летелицама, ова страница садржи фотографије које откривају све аспекте ваздухопловства. Осим фотографија, на мрежној страни могу се пронаћи и видео-записи, као и текстови. На приступној страни понуђен је слајд шоу сачињен од педесет тек пристиглих фотографија, као и место на коме се налази листа и везе ка ставкама које су посетиоци стране најчешће тражили (Popular Photo Searches). Претраживање се обавља по метаподацима једноставним уносом упита у поље Search, а може се обавити и одабиром атрибута метаподатака, из понуђених поља, ако се одабере опција Popular Photo Searches. У том случају, понуђен је велики број

атрибута, а неки од њих су: произвођач ваздухоплова, оператор, држава, тема фотографије, оријентација фотографије, фотограф, сортирање, ID фотографије. Резултати упита су најчешће фотографије, обележене ID бројем, а постоји и шири опис уколико се отвори фотографија.⁶ У том случају, не само што се фотографија боље види већ се добије детаљан опис у виду (атрибута и вредности) бројних метаподатака: произвођач, тип и модел ваздухоплова, ознака оператера, место или држава, датум када је фотографија снимљена, ID фотографије и анотација. Испод сваког дигиталног документа, у овом случају фотографије, налази се веза – адреса (image link). Овај мрежни ресурс садржи и грађу из историје нашег ваздухопловства, тј. ваздухопловства Југославије и Србије. Страница је прегледна и функционална. На страници постоји могућност регистрације.

3.2 Канада

Avialogs (Avialogs, [n.a.](#)) мрежна заједница и библиотека, регистрована је 2010. године да би се сачувале публикације из области ваздухопловства, а најширој публици омогући онлајн приступ. За потребе ове библиотеке дигитализоване су монографске и серијске публикације из области ваздухопловне технике. Функционише у виду неколико збирки које се односе на ваздухоплове, погонске групе, моторе, опрему, делове, конструкцију, приручнике и серијске публикације. За сада, претраживање се врши по метаподацима. Најављена нова верзија *Avialogs*-а, која за сада не предвиђа и претраживање по тексту – свим речима публикације већ потпуни препис (унапређење софтвера),⁷ цртеже и шеме, практично неограничено преузимање за све кориснике, било да су они правна или физичка лица, односно установе ваздухопловства, музеји или колекционари и месечни план. У погледу монографских публикација, дигитални документи су описани метаподацима који најчешће имају само следеће атрибуте: годину издавања, број страница, језик публикације, врста публикације и назив извора са веб-странице, на којој је објекат објављен. Уколико је реч о објекту из колекције Регистар ваздухоплова документи су описани метаподацима са више атрибута: назив серије, ознака серије, назив фабрике у којој је ваздухоплов произведен, назив типа мотора, као и називом извора (странице на којој је објекат објављен). Дигитални

6. Све фотографије представљају интерактивна места.

7. Примедба ауторке.

објекти (публикације) у потпуности су доступни регистрованим посетиоцима. Регистровани корисници (посетиоци странице) могу да приступе целим публикацијама, док остали могу да погледају првих десет страница публикације коју изаберу. Регистровани корисници имају избор и да бесплатно читају публикације или уз неку од понуђених чланарина стекну могућност да уреде страницу, преузму је или одштапају. Пре корака „учлањивања“, односно, претплате за коришћење, сваки посетилац има омогућен приступ свим дигиталним објектима, Корисници нису обавезни да се учлане док год немају потребу за одређеном публикацијом; могу без обавезе да погледају које публикације овај извор нуди). Дигитализоване су и три публикације на нашем језику о авионима послератне југословенске фабрике авиона Соко. Страница је прегледна и функционална.

3.3 Немачка

Страна *Flugzeugbilder* (Flugzeugbilder, **n.a.**) регистрована је 2001. године. Приказује фотографије ваздухоплова – аматерске фотографије и фотографије које су направили професионални фотографи. На приступној страни, која је на немачком језику, пружа се могућност прилагођавања странице говорницима енглеског, у будућности и француског језика. Претрага се обавља по метаподацима који имају следеће атрибуте:

- авио-компанија;
- аеродром;
- имејл;
- тип авиона;
- Id фотографије;
- фотограф;
- коментар;
- датум снимања;
- регистарска ознака авиона.

По страници се може погледати 10, 15, 25 или 50 фотографија. Претрага може да почне, односно, да се заврши одређеним термином, да се обави дословно унесеним термином или тако што посетилац странице унесе кључне речи. Претрага може бити једноставна и комбинована. Резултати претраге дају листу објеката. То су фотографије авиона са вредностима атрибута: Id фотографије, број прегледа, датум

снимања, фотограф, оператор, тип авиона и његова регистарска ознака, сортирани према броју или датуму прегледа, у растућем или опадајућем поретку. Аеродром се такође јавља међу атрибутима и описан је трословном и четворословном ознаком аеродрома; локацијом и називом аеродрома; државом. Ако се одабере одређена фотографија добије се већи приказ са истим метаподацима. На *Flugzeugbilder* постоје подаци о ваздухопловству Србије.⁸ Страница садржи оптималан број метаподатака; прегледна је и функционална. Посетилац странице може поставити и своје фотографије, а упутство о поступку налази се одмах на приступној страни. Регистрација на страници није понуђена.

3.4 Сједињене Америчке Државе

The Aviation History Online Museum (Aviation History Online Museum 1996) представља извор из историје ваздухопловства који нуди текстуалне, визуелне и аудио-визуелне записе. Овај мрежни извор је покренут 1996. године, у едукативне сврхе. Претходница ове странице је *Aviations Models*, страница посвећена авио-моделарству, покренута претходне, 1995. године. На приступној страни *Aviation History Online Museum* налази се садржај странице груписан по подручјима: ваздухоплови, мотори, теорија, ваздухопловци, моделарство, први светски рат, пионирско доба, видео-записи, везе ка другим страницама и укратко о страници. Свако подручје је описано са неколико метаподатака. Ако је реч о моторима ваздухоплова, ту су три атрибута: назив произвођача, тип мотора и држава производње. Уколико је реч о ваздухопловима, јављају се атрибути који се односе на произвођача, тип ваздухоплова, држава производње, атрибути текст и боја; последња два имају вредности „да“ и „не“. Страница је прегледна, обилује метаподацима којима су описани ваздухопловци, пионирско доба и теорија, док су остала подручја скромније описана. Страница *The Aviation History Online Museum* представља оптималан извор за историју ваздухопловства и може се рећи да је функционална. Поље за претрагу не постоји. Регистрација није предвиђена.

Digital Collections of the WWII Museum (The National WWII Museum, n.a.) је веб-страница Музеја Другог светског рата (*WWII Mu-*

8. Упит којим се проналазе подаци о ваздухопловству Србије (приступљено 17. 1. 2022).

seum) који је као установа основан 2003. године у Њу Орлеансу.⁹ Овај мрежни извор садржи фотографије и сведочења (учесника Другог светског рата) у форми видео-записа интервјуа. Претрага може бити и једноставна и врши се уносом упита у поље за претраживање на приступној страни. Претраживање олакшавају метаподаци сортирани по критеријуму службе у војсци (енгл. branch of service) и зони дејствовања (енгл. theater of operations).¹⁰ Иначе, већ са приступне странице, преко претраживача долази се до многобројних метаподатака који олакшавају претраживање. Сортирани су по служби у војсци, зони дејствовања и терминима (предметним одредницама) из речника који се односе на одређену тематику. Термини из речника су бројни, а сâм речник се састоји од хипертекстуалних веза, тј. термина и броја објеката на које се термин односи.¹¹ Термин може бити реч, а може бити и синтагма. Коришћење термина олакшава формирање упита. На пример, на упитни термин FRIENDSHIPS добију се два поготка/објекта: видео *Dabney Montgomery* (део *Early Life*) и видео *Donald Wilson (The Naval Battle of Guadalcanal)*. Сваки објекат садржи документ и метаподатке. У првом наведеном случају, документ је видео-запис; вредности атрибута су: Dabney Montgomery, Early Life; урађена је опсежна анотација; описни термини су Personal/family relationships, Prewar life, Experience with death, Great Depression, Race relations, Ethnic/racial identity, Education, Reaction to Pearl Harbor, Enlistment и Friendships. *Digital Collections of the WWII Museum* садржи и записе који се односе на историју Југословенског ваздухопловства.¹² Страница је прегледна, добро организована и као таква, врло корисна. Постоји и могућност

9. Потекао је из установе *The D-Day museum* која је основана 2000. године и била намењена успомени на америчке ветеране, односно, учесницима Битке за Нормандију. Опширније о настанку обеју установа (приступљено 13. 1. 2022).

10. Синтагме *US Army Air Forces*, *US Navy*, *US Army*, *US Marine Corps* и *Civilian US Merchant Marine* представљају атрибуте метаподатка који се односе на службу, а синтагме *China Burma India Theater (CBI)*, *Pacific Theater of Operations (PTO)*, *Atlantic Theater of Operations (ATO)*, *European Theater of Operations (ETO)*, *Mediterranean Theater of Operations (MTO)* и *Home Front* представљају атрибуте метаподатка који се односе на зону дејствовања. Доступно на (приступљено 11. 1. 2022).

11. Страна за прелиставање (приступљено 7. 2. 2022)

12. *Eben Smith – Final mission and capture*, *Luke McLaurine – Loss of aircraft and capture* (приступљено 13. 1. 2022)

регистрације која је, судећи по вредностима из понуђене листе, намењена грађанству Сједињених Америчких Држава.¹³

Flight Journal (*Flight Journal*, n.a.), часопис из области ваздухопловства покренут је у Рицфилду деведесетих година двадесетог века у класичном облику (ISSN 1095-1075), а почетком наредне деценије добија и своје онлајн издање. Историју ваздухопловства представља из перспективе пилота: “we put readers in the cockpit”. Основни метаподаци странице су: војно ваздухопловство, цивилно ваздухопловство, новости и приче (различити догађаји и теме; рестаурација, личности) и директоријум музеја ваздухопловства Сједињених Америчких Држава, у абecedном редоследу ставки, од А до W (Alabama – Wyoming). На веб-страници публикације могу се пронаћи и ставке из историје ваздухопловства Србије, приказане у облику чланка.¹⁴ Публикација нуди могућност претплате и регистрације.

Global Security (*Global Security*, n.a.) је континуирани онлајн извор, тј. серијска публикација (ISSN 2769-8947) регистрована 2000. године. Покрива безбедносна питања (на планети и ванпланетарном простору), односно, питања војне и обавештајне природе. *Global Security* објављује вести које се ажурирају сваког сата.¹⁵ Претраживање на страници може да се обави још са приступне стране, одабиром државе, односно, територије или уносом упита у поље за претраживање. Један од метаподатака за државе представља и ваздухопловство. Основни метаподаци на *Global Security* су: копнена војска, ратни бродови, авиони, пројектили, ласерске технологије, обавештајне службе, свемир; ратна подручја, паравојне снаге. Ова страница није релевантна за историју ваздухопловства, али је била једини извор о постојању ваздухопловства у случају појединих држава или територија.

Збирка *Aviation History Collection – Selections from the George Hardie Collection* Универзитета у граду Висконсин (UWM Libraries Digital Collections, n.a.) дигитализована је 2006. године. Реч је о десет наслова из поменутог збирке, тј. дигитализованим страницама публикација које се односе на ваздухопловство, у распону од 1865. до 1930. године.¹⁶ Иначе, дигитализација се у овој установи обавља од 2001. године¹⁷ чиме

13. Страна за регистрацију (приступљено 13. 1. 2022)

14. *On This Day in Aviation History* (приступљено 17. 1. 2022)

15. *Company History* (приступљено 18. 1. 2022)

16. Претраживање терминима (приступљено 11. 1. 2022)

17. О дигиталним колекцијама библиотека Универзитета у Висконсину (приступљено 18. 1. 2022)

је омогућен и удаљени приступ библиотечко-информационој грађи. До 2022. године дигитализована је грађа која обухвата преко седамдесет збирки. На страници постоји основна и напредна претрага, а резултати претраге описани су следећим атрибутима: насловом, аутором, годином, насловом публикације, анотацијом, предметним одредницама, врстом садржаја, местом издавања и називом издавача, медијумом, белешкама, опсегом публикације, језиком, поклонодавцем, годином дигитализације, форматом, као и другим елементима. Свака страница дигитализоване публикације описана је и сопственим насловом, годином, насловом публикације, ауторским правима и називом збирке. Нажалост, на овој лепо организованом и прегледној страници нема података и из историје нашег ваздухопловства.¹⁸ На страници није понуђена могућност регистрације.

Историјско друштво ваздухопловства Латинске Америке (*The Latin American Aviation Historical Society* или LAAHS) основано је 1997. године, а регистровано у Мајамију као непрофитна организација (*The Latin American Aviation Historical Society, n.a.*). Приказује различите текстуалне и визуелне форме: чланке, колумне, осврте, форуме на Фејсбуку у вези са историјом ваздухопловства Латинске Америке и представља једино средство које Друштво користи за објављивање, чување и дисеминацију резултата истраживања својих чланова.¹⁹ На приступној страни се, осим самих прилога, налазе и подаци о Друштву, упутства за ауторе, донаторе и посетиоце стране, као и поља за претрагу. Претрага може бити основна и обавља се уношењем упита у поље предвиђено за ову врсту претраге. Осим основне претраге, постоји и претрага која се обавља према критеријуму категорије прилога (чланак, колумна, приказ). Прилозима у форми текста и слике, придружене су и етикете. Иначе, ваздухопловство Латинске Америке је подељено по територијалном принципу. У оквиру приступне тачке за чланке (*Categories, Articles*) постоји подела на *Caribbean Aviation, Central American Aviation, General Latin America Aviation, Mexican Aviation* и *South American Aviation*. На страници постоји и архива прилога која покрива период 2018–2021. године, а претраживање се обавља одабиром месеца у одређеној години. Форум на страници је затвореног типа и односи се само на чланове групе (на друштвеној мрежи Фејсбук). Регистрација на овој страници није предвиђена.

18. Упит SERBIAN AND AVIATION није вратио ниједан погодак, као ни YUGOSLAVIA AND AVIATION.

19. **About LAAHS** (приступљено 19. 1. 2022)

Корпорација *Historynet*, LLC, основана 2015. године,²⁰ као највећи издавач часописа из области историје, објављује и часопис *Aviation history* (HistoryNet, [n.a.](#)). Страница пружа, поред осталог, информације на дневном нивоу и фото галерију. Претраживање на страници може бити основно, уносом теме, догађаја, периода или места у поље за претраживање. На упит SERBIA AND AVIATION HISTORY један од одзива био је и прилог под називом *F-117A: The Black Jet*.²¹ Прилозима нису придружени метаподаци, односно, етикете. Метаподаци за историју ваздухопловства на *Historynet*-у су следећи: *aviation history magazine*, *aviation history*, *Mag. Aviation History Aircraft*, *Mag. Aviation History Featured*, *Mag. Aviation History Hero*, *Mag. Aviation History Modeling*, *Mag. Aviation History Mystery* и *Mag. Aviation History Personalities*. Регистрација на страници није предвиђена, али се нуди могућност претплате.

3.5 Холандија

Aviation News (Aviation News, [n.a.](#)) је веб-страница, покренута 2006. године у Холандији, која пружа најновије вести из ваздухопловства у општем смислу, а од ње потичу стране *Aviation Museum database*, *Airshow reports* и незванична веб-страница *Paris Airshow*.²² Поменуће странице представљају онлајн ресурсе усмерене на актуелности и историју ваздухопловства. За овај рад релевантна је страница *Aviation Museum*. Покрива територије Африке, Азије, Европе, Латинске Америке, Блиског истока, Северне Америке и Океаније. Постоји и мапа странице – листа са хипертекстуалним везама ка свакој држави која је релевантна за ову страницу. Свака од држава може да буде представљена и новом листом (музеја или установе ваздухопловства и везама уколико постоји веб-страница музеја или установе). На упит SERBIA одзив су „Музеј ваздухопловства“ у Београду, „Соко Јастреб“ и „Миг-21“ – два војна авиона који се налазе у дворишту Ваздухопловне академије у Београду. За све три *Aviation News* веб-странице претрага је једноставна и обавља се на исти начин. Уноси се упит у поље за претрагу и добија се одговор. Одзив је представљен у форми текста и слике, а у већини случајева везом ка другим страницама и етикетама. Велики број прилога нарушава

20. *Historynet* (приступљено 18. 1. 2022)

21. *F-117A: The Black Jet* (приступљено 14. 2. 2022)

22. *Aviation museum*, *Airshow reports*, *Paris Airshow* / (приступљено 13. 1. 2022.)

прегледност и функционалност странице. На страницама није понуђена регистрација, али је понуђена претплата на билтен и контакт уколико посетилац има потребу да се обрати информационим захтевом.

3.6 Мрежни извори у Србији

Дигитална Народна библиотека Србије²³ је онлајн ресурс Народне библиотеке Србије која је са дигитализацијом почела 2004. године (Butigan-Vučaj 2003). Претраживање се врши на основу метаподатака и на основу речи из текста. Напредно претраживање може бити детаљно и оно се обавља додељивањем вредности атрибутима метаподатака, а то су наслов, аутор, тема, идентификатор, а могу се додатно унети и назив издавача, место издавања, подаци о години или распону година које су релевантне за претрагу, одабрати језик на којем је објављена публикација, одабрати и врсту библиотечко-информационе грађе (монографске публикације, серијске публикације, чланци или други саставни делови публикације, звучни снимци, картографска грађа, нотна грађа, сликовна грађа, рукописна књига и рукописна грађа) или унети вредност атрибута у „име и презиме власника колекције“. Коначно, претраживање је могуће обавити и на основу речи из текста што је као могућност уведено накнадно и представља много боље решење које доприноси квалитету дигиталне библиотеке. Дигитална НБС садржи дигитализоване публикације и некњижну библиотечко-информациону грађу из области ваздухопловства. На Дигиталној НБС могу се пронаћи и фотографије на којима се може видети, рецимо, најављивање предстојећег аеро-митинга. Документу је придружен скроман опис, са свега два атрибута. То су наслов и идентификациони број записа у библиографско-каталогској бази CO-BISS.SR или COBISS.SR-ID, односно, вредности атрибута. Вредности атрибута могу да буду, на пример: Београд : Теразије : Terazié и 95356428.²⁴ ID 95356428 упућује на библиографски запис који пружа шири опис коришћењем следећих атрибута: упоредни наслов, језик, врста грађе, година, издавање и производња, физички опис, напомене, предметне одреднице, број УДК, линк и сажетак. Дигитализованим

23. Реч је о „старој“ Дигиталној НБС (Narodna biblioteka Srbije, [n.a.](#)).

24. На Дигиталној НБС могу се наћи фотографије које упућују на ваздухопловство наше војске. Налазе се на адресама: [Балкански рат 1912-1913. 1/25](#), [Балкански рат 1912-1913. 1/27](#), [Балкански рат 1912-1913. 1/28](#), [Istorija jugoslovenskog vazduhoplovstva](#) (приступљено 15. 1. 2022)

монографским публикацијама придружено је неколико атрибута: наслов, предмет, аутор, издавач (штампар), датум, језик и COBISS.SR-ID. Опширнији приказ метаподатака садржи и следеће атрибуте: превод наслова, опсег, формат и просторну одредницу. Документу је придружен и атрибут: Колекција, са вредношћу: Књиге, што значи да је реч о колекцији дигитализованих монографских публикација. Домаће дигитализоване новине представљају могуће изворе о историји ваздухопловства. У претраживим новинама, рецимо, у *Политици* (Inјас 2006)²⁵ може се наћи значајан број чланака из ваздухопловне тематике. Тако су пропраћени важни догађаји за ваздухопловство. За потребе рада биће наведени само неки од њих: пад авиона Мерћеп-Русјан трећег дана Божића на Калемегдану, у Београду 1910. године,²⁶ први регистровани судар авиона у цивилном ваздухопловству, који се десио на релацији Париз–Лондон почетком априла 1922. године, прва домаћа ваздухопловна изложба, одржана у Београду 1925. године, као и прва међународна ваздухопловна изложба у Београду 1938. године. Извршена претрага у оквиру овог ресурса указује на домаће дигитализоване новине као могући извор. Дигитална НБС је прегледна и функционална. Регистрација није предвиђена.

У стручној периодици из области друштвених наука, културе и технике постоје радови који се баве историјом ваздухопловства. Конзорцијум библиотека Србије за обједињену набавку (КоБСОН) (Kosanović 2003) представља одличан извор научних радова из области ваздухопловства. КоБСОН може да пружи велики одзив, али и малу прецизност. Претраживање по кључним речима, преко сервиса за обједињену претрагу садржаја EBSCO Discovery Service, обављено је на следећи начин. Следећи упити пружили су велики одзив/поготке (1.277.719, 1.261.374, 111.588, 1.181) у виду чланака из серијских публикација и монографских публикација:

- AVIATION HISTORY,
- AVIATION AND HISTORY,
- AVIATION HISTORY,
- AVIATION HISTORY са одабиром: Full-Text (online), Peer Reviewed.

Применом одговарајуће стратегије могуће је добити прецизније податке. Иако је постављен временски интервал 1900–2021, сужавањем

25. *Politika* (приступљено 15. 1. 2022)

26. Година је по јулијанском календару

претраге (на научне часописе и применом предметне одреднице AVIATION HISTORY), добијен је много мањи број погодака (117) и повећана релевантност.

Радови који се баве историјом домаћег ваздухопловства неретко се могу наћи у часописима као што су *Историја 20. века*: часопис Института за савремену историју²⁷ и *Наслеђе*²⁸ који су на DOAJ (*Directory of Open Access Journals*) листи. Од аутора радова из историје домаћег ваздухопловства у *Историји 20. века* истиче се Бојан Б. Димитријевић,²⁹ а у *Наслеђу*, са неколико чланка из области архитектуре, у којима су теме зграде установа ваздухопловства и ваздухопловне индустрије, не смеју се изоставити ауторке Биљана Мишић, Хајна Туцић и Саша Михајлов.³⁰ Један од последњих радова из историје домаћег ваздухопловства, који се налази на DOI листи представља чланак Илије Кукобата „Почеци ваздушног саобраћаја у послератној Југославији 1945–1947“.³¹

4. Могући модел дигиталне библиотеке о историји ваздухопловства

Према досадашњим искуствима са мрежним странама које презентују историју ваздухопловства, те *Еуропеаном* и *Дигиталном НБС*, претпостављена дигитална библиотека историје домаћег ваздухопловства могла би да буде скуп добрих страна свих поменутих

27. *Istorija 20. veka (Cobiss)*; Часопис је онлајн јавно доступан и описан следећим метаподацима: Institut za Savremenu istoriju, History (General) and history of Europe: History of Eastern Europe, 20th century, balcan, contemporary history, serbia, yugoslavia, world war 2 (приступљено 19. 2. 2022)

28. *Nasleđe (Cobiss)*; Часопис је онлајн јавно доступан и описан следећим метаподацима: Zavod za zaštitu spomenika kulture grad Beograda, Fine arts: Architecture, architecture, belgrade, conservation, cultural monuments, heritage (приступљено 19. 2. 2022).

29. У часопису *Историја 20. века* јавља се као аутор и коаутор чланака, приказа, осврта и других прилога; активан у уредништву часописа *Istorija 20. veka* (приступљено 20. 2. 2022). Димитријевић се такође јавља као аутор, коаутор и рецензент монографских публикација *Bojan Dimitrijević (Cobiss)* (приступљено 20. 2. 2022).

30. *Biljana Mišić, Haјna Tucć, Saša Mihajlov* – radovi u časopisu *Nasleđe*.

31. *Pija Kukobat* - rad u časopisu *Istorija 20. veka* Vol. 38, no. 2/2020, pp. 173 – 186.

веб-извора. Пре свега, требало би да постоји служба за комуникацију са корисницима, тј. посетиоцима странице, у сврхе стручних консултација и унапређења квалитета других услуга. Даље, по угледу на *Digital Collections of the WWII Museum* претпостављена дигитална библиотека могла би да постане зачетник усмене историје домаћег ваздухопловства. Претрага би требало да се обавља по метаподацима и по свим речима у документима типа текста. Имала би одлике: *Avialogs*-а у погледу концизности; *Europeane* у погледу обухвата различитих форми артефаката културне баштине; *Дигиталне НБС* у погледу усмерености на чување националне баштине; *Digital Collections of the WWII Museum* у погледу усмене историје и организације претраживања, метаподатака и постојања својеврсног термилошког речника; *Air Team Images* у погледу свих аспеката фотографије у ваздухопловству.

Ова дигитална библиотека би могла да функционише и као виртуелни институт за историју домаћег ваздухопловства.

Имајући у виду да би претпостављена дигитална библиотека могла да постане значајан извор истраживачима на пољу историје домаћег ваздухопловства, од значаја би био и велики број елемената описа – метаподатака.³²

Неизоставан наслов у претпостављеној дигиталној библиотеци историје ваздухопловства Србије била би већ поменута монографска публикација *Историја југословенског ваздухопловства* (Микић 1933), ка којој би могла да буде упућена веза, као и мемоари истог аутора (Микић 2000). Од серијских публикација добро би било да се дигитализује *Ваздухопловни гласник*³³ са своја два додатка. Претпостављена дигитална библиотека би везама (линковима) упућивала на друге мрежне локације објеката који су у вези са историјом домаћег ваздухопловства и сарађивала са организацијама које дигитализују артефакте историје домаћег ваздухопловства, као и са власницима приватних збирки и породичних колекција. Посебно би добро дошла

32. Пожељни би били следећи атрибути: власник (колекције или појединачног физичког предмета), назив колекције, наслов, превод наслова на енглески, језик, id објекта, аутор, просторна одредница, временска одредница, опсег, опис, стручна класификација, предметна класификација, анотација, медиј, формат, датум дигитализације, а по могућности и следећи атрибути: тип и верзија ваздухоплова, аеродром (летелиште), установа ваздухопловства, догађај (дешавање, манифестација), особа која је пружила допринос (сарадник, претходник, настављач, члан породице).

33. *Vazduhoplovni glasnik (Cobiss)*

сарадња са Министарством одбране РС у погледу заједничких активности Војног архива и *Музеја ваздухопловства* у Београду, на плану дигитализације и промоције архивске и музејске грађе која је од великог значаја за историју домаћег ваздухопловства (Масица 2017). Тимски рад и сарадња се у случају дигиталних библиотека подразумевају, па се тиме успоставља и својеврсна виртуелна заједница поштовалаца историје домаћег ваздухопловства.

5. Закључак

Дигитални, тј. мрежни извори приказују историју ваздухопловства на различите начине. На понуду веб-странице у великој мери утичу различите циљне групе посетилаца: стручна јавност, шира јавност, те они који су у процесу образовања. Постоје веб-странице које су креирали сами стручњаци из праксе ваздухопловства, ентузијастички и заљубљеници у ваздухопловство, као појединци, групе или друштва. Такође, постоје и академски извори, односно, веб-странице универзитета и установа културе који садрже грађу из историје ваздухопловства. Веб-странице које приказују историју ваздухопловства чине то у образовне сврхе, у рекреативне сврхе или своје садржаје представљају као научне информације. Први електронски извори који се баве историјом ваздухопловства настали су већ половином деведесетих година XX века, а организације и удружења не престају ни до данашњих дана да постављају на мрежу садржаје из историје ваздухопловства. Сваки од поменутих дигиталних извора представља историју ваздухопловства на свој начин, али међу њима нема извора који по својим елементима представља дигиталну библиотеку историје ваздухопловства. У том смислу, од наведених извора у тексту, ипак се издваја *Avialogs*. Иако се декларирало као мрежна библиотека и заједница, то је најближи извор идеји дигиталне библиотеке историје ваздухопловства, али ни он ипак не нуди широк спектар услуга, већ је усмерен на дигитализацију публикација, приступ дигиталним колекцијама и евентуално, њихову репродукцију у некомерцијалне сврхе (штампање или преузимање). Нова верзија странице могла би да понуди претраживање по целом тексту публикације, детаљнији опис укључивањем већег броја метаподатака – најбоље по угледу на неки од постојећих стандардних библиографских описа, као и сервис за консултације из области историје технике ваздухопловства. На тај начин унапређен, *Avialogs* би испунио сврху једне дигиталне библиотеке, а онима који приступају, било да су из

струке ваздухопловства или ученици, студенти, истраживачи, предавачи или лаици, омогућио би спектар различитих информационих услуга.

Литература

- Butigan-Vučaj, Tamara. 2003. "Digitalna Narodna biblioteka Srbije (DNBS)." *Pregled Nacionalnog centra za digitalizaciju*, no. 1, 40–42.
- IFLA/UNESCO. 2012. *Manifesto for Digital Libraries – Манифест за дигиталне библиотеке*. Приступљено 3. 4. 2022. <https://cdn.ifla.org/wp-content/uploads/files/assets/digital-libraries/documents/ifla-unesco-digital-libraries-manifesto-sr.pdf>.
- Injac, Vesna. 2006. "Narodna biblioteka Srbije u projektu "Evropska biblioteka" (TEL–The European Library)." *Glasnik Narodne biblioteke Srbije*, no. 1, 37–45.
- Kosanović, Biljana. 2003. "Konzorcijum biblioteka Srbije za objedinjenu nabavku: da li dovoljno koristimo elektronske izvore informacija." *Infoteka* 4 (2): 107–113.
- Macura, Ljiljana. 2017. "Zbirke vazduhoplovnih institucija u očuvanju kulturne baštine: Muzej vazduhoplovstva - Beograd." In *Programi zaštite, promocije i približavanja kulturne baštine široj javnosti : zbornik radova = Programs of protection, promotion and convergence of cultural heritage to general public : conference proceedings*. Sarajevo: Asocijacija informacijskih stručnjaka, bibliotekara, arhivista i muzeologa = Association of Information Professionals – Librarians, Archivists / Museologists.
- Микић, Сава Ј. 1933. *Историја југословенског ваздухопловства: са 615 слика у тексту од којих две Њ. В. Краља и Престолонаследника Петра у бојама*. Стр. V-VII: Предговор / Мил. Ј. Николајевић. Штампарија Драг. Грегорића.
- Микић, Сава Ј. 2000. *Недовршени мемоари*. Стр. 9-10: Предговор / Александра Вранеш. Стручна књига.
- Трговац, Александра. 2017. *Проналажење информација у дигиталним библиотекама*. Vol. 4. Универзитетска библиотека "Светозар Марковић". ISBN: 978-86-7301-103-5.

Извори са веба

- Aeroflight. n.a. *Aeroflight – Home page*. www.aeroflight.co.uk. Уредник John Hayles; приступљено 11. 1. 2022.
- Aeroflight Cloud. n.a. *Aeroflight Cloud – Home page*. https://docs.google.com/presentation/d/10wZReN2_38bvTBAFLYJiCpgSFItPYD0Y4UvCnWLBako/edit#slide=id.g9d40c95f98_0_25. Уредник John Hayles; приступљено 11. 1. 2022.
- AirTeamImages. n.a. *Aviation Image Library*. www.airteamimages.com. Приступљено 6. 2. 2022.
- Avialogs. n.a. *Aviation Library*. <https://www.avialogs.com/>. Креирао Benoit de Mulder; приступљено 11. 1. 2022.
- Aviation History Online Museum. 1996. *Aviation History Online Museum – Home page*. <http://www.aviation-history.com>. Креатор веб-сајта Larry Dwyer; приступљено 18. 1. 2022.
- Aviation News. n.a. *International aviation news, airshow reports, aircraft facts, worlds largest aviation museum dabase*. <https://aviationnews.eu/>. Приступљено 13. 1. 2022.
- Flight Journal. n.a. *Flight Journal (online)*. <https://www.flightjournal.com/>. Приступљено 17. 1. 2022.
- FltCtrl Ltd. n.a. *Flock aero – Home page*. <https://flock.aero/>. Приступљено 17. 1. 2022.
- Flugzeugbilder. n.a. *Aviation photography at its best*. <https://www.flugzeugbilder.de/v3/index.php>. Приступљено 17. 1. 2022.
- Global Security. n.a. *Global Security (online)*. <https://www.globalsecurity.org/index.html>. Приступљено 18. 1. 2022.
- HistoryNet. n.a. *Aviation History*. <https://www.historynet.com/magazine/aviation-history/>. Приступљено 18. 1. 2022.
- Narodna biblioteka Srbije. n.a. *Digitalna Narodna biblioteka Srbije*. <https://digitalna.nb.rs/>. Приступљено 15. 1. 2022.
- Royal Aeronautical Society. 2021. *Heritage Collection*. <https://aerosocietyheritage.com/>. Приступљено 20. 1. 2022.

The Latin American Aviation Historical Society. n.a. *The Latin American Aviation Historical Society*. <https://www.laahs.com/>. Приступљено 19. 1. 2022.

The National WWII Museum. n.a. *Digital Collections of the WWII Museum*. <https://www.ww2online.org/>. Приступљено 13. 1. 2022.

UWM Libraries Digital Collections. n.a. *Aviation History Collection – Selections from the George Hardie Collection*. https://uwm.edu/lib-collections/aviation_history/. Приступљено 11. 1. 2022.

Европеанин модел података: кључ за сарадњу са Европеаном

УДК 025.3:004.9

САЖЕТАК: Европеанин модел података (Europeana Data Model – EDM) креиран је како би се метаподаци о разнородним дигиталним објектима могли пресликати на јединствени модел података гарантујући интероперабилност. Овај модел заснива се на принципима отворених повезаних података и на семантичком вебу. Рад се бави применом овог модела у Народној библиотеци Србије кроз директну сарадњу са Европеаном у својству достављача података, али и у својству националног агрегатора за Европеану.

КЉУЧНЕ РЕЧИ: EDM, дигитална библиотека, Европеана, метаподаци, агрегатор метаподатака.

РАД ПРИМЉЕН: 21. март 2022.

РАД ПРИХВАЋЕН: 07. јуни 2022.

Татјана Домазет

tatjana.domazet@nb.rs

Наташа Петровић

natasa.petrovic@nb.rs

Тамара Бутиган Вучај

tamara@nb.rs

Народна библиотека Србије

Београд, Србија

1. Увод

Европеана је најважнија иницијатива и највећа инвестиција Европске комисије на подручју дигитализације културног наслеђа. Покренута је 2008. године са мисијом да трансформише свет културом, али и да ојача европски идентитет код грађана Европе кроз заједничко, дигитализовано културно наслеђе. Језгро Европеане представља платформа¹ која пружа приступ до преко 51 милион дигиталних објеката из више од 3.500 библиотека, музеја, архива и галерија Европе (март 2022. године). Екосистем који чине три међусобно повезане експертске организације: Фондација Европеана, Форум агрегатора и Асоцијација Мрежа Европеане назива се Иницијатива Европеана. Фондација Европеана, са седиштем у Националној и краљевској библиотеци Холандије, са око 60 запослених, независна је и непрофитна

1. Europeana, „Welcome to Europeana“

организација која развија платформу и сарађује са другим дигиталним иницијативама које се баве културним наслеђем. Форум агрегатора Европеане чини 40 акредитованих агрегатора метаподатака, и то 13 доменских и тематских и 27 националних и регионалних. Асоцијација Мрежа Европеане је отворена и демократска заједница стручњака за дигитализацију културног наслеђа, а броји преко 3.000 стручњака из институција културног наслеђа, ИТ-а, образовних и истраживачких институција. Мисија Европеане гласи: „Европеана подржава сектор културног наслеђа у спровођењу дигиталне трансформације. Развијамо експертизу, алате и политике ради прихватања дигиталне промене и охрабривања партнерстава која подстичу иновације. Олакшавамо људима да користе културно наслеђе у образовању, истраживању, стварању и рекреацији. Наш рад доприноси отвореном, сазнајном и креативном друштву.“ Визија Европеане гласи: „Европеана жели да види како дигитализација покреће сектор културног наслеђа и како култура покреће Европу, дајући јој отпорну и растућу економију, повећану запосленост, више благостања и осећај европског идентитета.“

Како Европеана није конципирана као репозиторијум дигиталних садржаја већ као агрегатор метаподатака о дигиталним садржајима, Европеанин модел података (EDM) кичмени је стуб њене платформе, као схема на којој је заснована база података у њеној позадини. Овај модел обезбеђује интероперабилност метаподатака из различитих институција и представља најмањи заједнички садржалац за све схеме метаподатака које иначе користе различите установе културе у описивању својих дигиталних објеката, на пример METS у библиотекама, LIDO у музејима или EAD у архивима.

Народна библиотека Србије сарађује са Европеаном од њеног настанка као достављач метаподатака, у почетку преко Европске библиотеке (TEL), која је имала улогу агрегатора за све националне библиотеке Европе, до гашења у децембру 2016. Од 2015. године, Народна библиотека Србије ради на успостављању националног агрегатора, а исте године је потписан и уговор о размени података са Европеаном (Data Exchange Agreement – DEA). Агрегаторска инфраструктура за Европеану (Агрегатор за Европеану) развијена је у партнерству са словеначком фирмом Семантика (Бутиган-Вучај 2019) и функционише у оквиру Одељења за развој дигиталне библиотеке и микрофилмовање од марта 2018. године. Исте године, Народна библиотека Србије постаје чланица Форума агрегатора јер је испунила све услове да постане акредитовани агрегатор Европеане и њен поуздани

партнер. Агрегатор за Европеану у НБС препознат је као достављач квалитетних података, а достигао је и високе стандарде у опису садржаја, мапирању метаподатака, додељивању лиценци за коришћење дигиталних објеката, приступачности, вишејезичности и коришћењу контролисаних речника, и уопште као извор знања, савета и подршке дигитализацији. Уз све ово то је једини агрегатор из земље која није чланица Европске уније. У Народној библиотеци Србије се од 2014. године примењује Европеанин модел података (EDM) за припрему скупова података за објављивање на платформи и то кроз следеће европске пројекте: Европеана збирке: 1914–1918, Збирке југоисточне Европе у Европеани (CSEEE), Процват писмености, Миграције кроз науку и уметност и Заједничка култура у Европеани. TEL је био агрегатор метаподатака за прва два пројекта, док је за преостала три агрегатор у Народној библиотеци Србије већ био оперативан. Кроз пројекат Заједничка култура у Европеани,² у коме су учествовала 23 агрегатора из 21 земље Европе, највише је учињено на пољу стандардизације метаподатака и генерално агрегаторског окружења, а Народна библиотека Србије је максимално унапредила квалитет метаподатака својих дигиталних објеката за све скупове података објављене на платформи Европеане.

2. Европеанин модел података (EDM)

Еuropeана представља разноврсну грађу из различитих институција које чувају културно наслеђе из целе Европе. Те институције ослањају се на различите стандарде метаподатака, а они у великој мери зависе од врсте грађе. Сви ти ресурси изузетно су значајни, али услед њихове разноликости, као и тежње да се што више информација сачува, уз што мање преклапања података, установљена је потреба за успостављањем заједничког стандарда метаподатака који би могао да се прилагоди грађи из библиотека, музеја, архива и институција које чувају аудиовизуелни материјал. Услед тога, почело се размишљати о увођењу заједничког модела података, на основу кога подаци могу да се повезују и умрежавају, што олакшава претрагу грађе по одређеним критеријумима.

2. Europeana, “Europeana Common Culture”, преузето 18. 3. 2022

Европеанина схема метаподатака³ (Europeana Semantic Elements – ESE) представља први резултат рада на овој идеји, а њеним даљим развојем настао је и Европеанин модел података⁴ (Europeana Data Model – EDM) (Дрере 2019; Климпел and Ојлер 2019). Предности новог формата, у односу на претходни, у томе су што EDM⁵ омогућава интегрисање семантичких поља и опис хијерархијских структура. Метаподаци могу да се повежу са дигиталним репродукцијама грађе на коју се односе. EDM формат оставља и могућност уноса слободног текста ради што квалитетнијег описа грађе. Осим тога, EDM омогућава унос вредности које се односе на физички објекат, али и на његову дигиталну копију.

Како EDM обједињује грађу из различитих институција културе, он је боље решење од појединачних стандарда који се односе на конкретне типове грађе, а пребацивањем метаподатака у овај формат могу да се пресликају сви метаподаци, па не долази до њиховог осиромашења. EDM се заснива на отвореном оквиру који обезбеђује учешће у семантичком вебу. Омогућава повезивање података различитих институција који се односе на лица, места и предметне области, што поспешује међусобно обогаћивање и употпуњавање података. Овај приступ олакшава корисницима откривање нових ресурса о теми која их занима. Тако се увећава број посета објектима из различитих институција.

EDM формат се заснива на стандардима OAI ORE (Open Archives Initiative Object Reuse & Exchange),⁶ SKOS (Simple Knowledge Organization System),⁷ DC (Dublin Core)⁸ и CIDOC-CRM.⁹ Ослања се и на различите нормативне базе података и речнике. На тај начин, успостављају се нове везе између објеката и превазилазе језичке баријере. Помоћу EDM формата, подаци се прикупљају, повезују и обогаћују, а како се овај стандард заснива на отвореним подацима и семантичком вебу, културно наслеђе из институција које сарађују са Европеаном допире до већег броја корисника. Семантички веб пружа шири контекст за метаподатке и

3. Europeana, „Europeana Semantic Elements Documentation“, преузето 17. 3. 2022

4. Europeana, „Europeana Data Model“, преузето 17. 3. 2022

5. Europeana, „Europeana Data Model – Mapping Guidelines v2.4“, преузето 17. 3. 2022

6. „Open Archive Initiative - Object Exchange and Reuse“

7. „SKOS Simple Knowledge System Namespace Document“

8. „DCMI Metadata Terms“

9. „Classes & Properties Declarations of CIDOC-CRM“

омогућава представљање сложених односа између различитих објеката, као и различитих делова једног објекта, а уједно подстиче поновну употребу података и ослањање на спољне ресурсе, доступне у виду повезаних података. Принципи семантичког веба имплементирани су у виду RDF¹⁰ (Resource Description Framework) стандарда.

У оквиру EDM формата постоје три основне класе. Њима се представљају три различите целине једног записа. На физички објекат културног наслеђа односи се класа <edm:ProvidedCHO>. Ту су подаци о аутору, издавачу, години издања, формату, језику публикације, јединственом идентификатору објекта, као и предметне и временске одреднице. Класа <ore:Aggregation> обухвата метаподатке за дати објекат, као и податке о институцији у којој се чува објекат и институцији која је доставила податке. Ту се налази линк ка дигиталном објекту и линк ка одговарајућој лиценци, на основу које се корисници информишу о начинима на које могу да користе грађу. Класа <edm:WebResource> односи се на дигитални приказ објекта. Поред основних класа, користе се и додатне класе за представљање ентитета који су у вези са објектом <edm:Agent>, <edm:Place>, <edm:TimeSpan>, <skos:Concept> итд. У оквиру класа постоје својства у којима су дескриптивни елементи. Класе и својства дефинишу се на основу RDFS (Resource Description Framework Schema)¹¹ схема и OWL елемената (Web Ontology Language).¹²

Народна библиотека Србије за грађу подигнуту у Дигиталну НБС користи COMARC/В стандард метаподатака у склопу платформе CO-BISS (слика 1). Подаци се одатле експортују и конвертују у EDM формат у коме се достављају Европеани. У циљу конверзије, развијени су бројни интерни алати за прочишћавање, проверу, обогаћивање и конверзију метаподатака. Они омогућавају пребацивање садржаја неопходних поља из comarc.xml записа у одређена поља у EDM-у. Након те трансформације, добијени EDM запис подиже се на Агрегатор за Европеану, који врши додатну трансформацију, након које је запис спреман за преузимање и објављивање на сајту Европеане.

У оквиру етикете <ore:aggregates> налазе се подаци који се односе на један дигитални објекат (књига, један број периодичне публикације или фотографија). У оквиру класе <edm:ProvidedCHO> у rdf:about налази се линк ка дигиталном објекту на платформи НБС за представљање

10. World WideWeb Consortium „RDF Primer“

11. „RDF Schema 1.1“

12. „owl:Ontology“

```
<core:aggregates>
  <edm:providedCHO rdf:about="http://www.digitalna.nb.rs/vb/NBS/Tematske_kolekcije/industrialjsko_nasledje/borski_rudnici/Krvaro_plato/Q4_85">
    <dc:contributor xml:lang="sr">Jovan Popovic (1905-1952)</dc:contributor>
    <dc:contributor xml:lang="sr">Jovan Bonomoh (1905-1952)</dc:contributor>
    <dc:contributor rdf:resource="http://www.wikidata.org/wiki/Q6926683" xml:lang="sr"/>
    <dc:creator xml:lang="sr">Borče Andrejević-Kun (1904-1964)</dc:creator>
    <dc:creator xml:lang="sr">Popbe Anđeljević-Kun (1904-1964)</dc:creator>
    <dc:creator rdf:resource="http://www.wikidata.org/wiki/Q1833347" xml:lang="sr"/>
    <dc:description xml:lang="sr">Tiraž 250.</dc:description>
    <dc:description xml:lang="sr">Str. 1-11: Umjetnik i njegovo delo / J. Popovic.</dc:description>
    <dc:format xml:lang="sr">12 str., 28 listova s tablama</dc:format>
    <dc:format xml:lang="sr">dvorezi, crno-belo</dc:format>
    <dc:format xml:lang="sr">25 cm</dc:format>
    <dc:identifier>191919516</dc:identifier>
    <dc:language>sr</dc:language>
    <dc:publisher xml:lang="sr">B. Andrejević-Kun</dc:publisher>
    <dc:subject xml:lang="sr">Popbe Anđeljević-Kun</dc:subject>
    <dc:subject xml:lang="sr">Muzejstvo nasleđe</dc:subject>
    <dc:subject xml:lang="en">Industrial Heritage</dc:subject>
    <dc:title xml:lang="sr">Krvaro plato : 28 originalnih dvoreza B. Andrejevića-Kuna</dc:title>
    <dc:terminusD191919516</dc:terminusD191919516>
    <dc:currentLocation rdf:resource="http://www.goonames.org/792680"/>
    <dc:type>TEXT</dc:type>
    <dc:type xml:lang="sr">naslov</dc:type>
  </edm:providedCHO>
</core:aggregates>
<edm:aggregation rdf:about="http://www.digitalna.nb.rs/vb/NBS/Tematske_kolekcije/industrialjsko_nasledje/borski_rudnici/Krvaro_plato/Q4_85">
  <edm:provider>National Library of Serbia</edm:provider>
  <edm:dataProvider>National Library of Serbia</edm:dataProvider>
  <edm:rights rdf:resource="http://creativecommons.org/licenses/by-nc-sa/4.0/">
  <edm:isShownAt rdf:resource="http://www.digitalna.nb.rs/vb/NBS/Tematske_kolekcije/industrialjsko_nasledje/borski_rudnici/Krvaro_plato/Q4_85">
  <edm:isShownBy rdf:resource="http://www.digitalna.nb.rs/attach/NBS/Tematske_kolekcije/industrialjsko_nasledje/borski_rudnici/Krvaro_plato/Q4_85/output.pdf">
  <edm:object rdf:resource="http://www.digitalna.nb.rs/vb/NBS/Tematske_kolekcije/industrialjsko_nasledje/borski_rudnici/Krvaro_plato/Q4_85?pageIndex=thumb"/>
</edm:aggregation>
</core:aggregates>
```

Слика 1. Пример једног записа из Дигиталне НБС након конверзије из CO-MARC/V формата у EDM

дигитализоване грађе¹³ (Европеана не чува грађу из различитих институција културе, већ само представља метаподатке и упућује на портале институција путем линкова). Поље <dc:creator> односи се на аутора. Након интегрисања нормативне базе података имена аутора CONOR¹⁴ у систем COBISS-a, уведена су два поља <dc:creator> (ћирилица и латиница или етимолошки написано име страних аутора), а доступне су и године рођења и смрти, уколико су познате. Осим тога, у једно од тих поља уноси се и линк ка имену аутора у бази података Wikidata.¹⁵ За секундарно ауторство се, по истом принципу, креирају поља <dc:contributor>. Поље <dc:format> такође је поновљиво и садржи информације о физичком примерку грађе (димензије, број страна и сл.), као и поље <dc:description> у које се уноси податак о тиражу и одређене напомене из описа грађе. Поље <dc:subject> односи се на предметне одреднице које могу да се уносе на српском и енглеском језику (језик означавамо у оквиру атрибута xml:lang), у циљу ефикаснијег повезивања података. Поље <dc:identifier> представља јединствени идентификатор објекта (у случају података НБС то је COBISS ID). У склопу атрибута xml:lang уноси се језик метаподатака (што је обавезно код текстуалног садржаја у оквиру етикета), док се поље <dc:language> односи на

13. Народна библиотека Србије, „Дигитална НБС“

14. Izum, „CONOR нормативна база“

15. „Wikidata“

језик публикације (поновљиво поље код вишејезичних публикација). Користе се двословни језички кодови према стандарду ISO 639-1.¹⁶ Поље <dc:publisher> односи се на издавача, док <dc:title> садржи наслов објекта. Поље <dcterms:issued> представља годину издавања, а <edm:currentLocation> је место где се чува физички примерак грађе. Ту се додаје локација из географске базе података Geonames.¹⁷ У <edm:type> се уносе дефинисане вредности “TEXT” за текстуалну или “IMAGE” за сликовну грађу, а у <dc:type> врста грађе.

Класа <ore:Aggregation> садржи линкове ка страници на којој се налази дигитални објекат. Линкови ка самом објекту уносе се у следеће етикете <ore:Aggregation>, <edm:aggregatedCHO> као и <edm:isShownAt>. Поље <edm:isShownBy> садржи линк ка PDF за текстуалну грађу, док се за сликовну грађу уноси линк ка увећаном приказу слике у прегледачу. У поље <edm:object> ставља се линк ка умањеном приказу објекта (thumbnail). Поред ових линкова који воде на дигитални објекат, у ову класу се у оквиру поља <edm:rights> уноси линк за одговарајућу лиценцу Креативне заједнице.¹⁸ Поље <edm:provider> је агрегатор који доставља податке Европеани, а <edm:dataProvider> је библиотека чија је грађа и која доставља метаподатке агрегатору.

Тако формиран запис у EDM формату поставља се на Агрегатор за Европеану који врши додатну трансформацију, након које је запис спреман за објављивање на Европеани (слика 2).

3. Оквир за објављивање података на Европеани

Европеана је прописала стандарде за опис метаподатака и садржаја. Њихов квалитет одређује начин на који ће грађа бити приказана и промовисана на порталу Европеане, као и могућност њеног преузимања. Када су метаподаци¹⁹ у питању, постоје три критеријума за њихов опис и означени су словима А, В и С.²⁰ Ниво А је најнижи, док је ниво С највиши. Европеанин принцип је „што више дате више ћете добити“,²¹

16. Wikipedia, „List of ISO 639-1 codes“

17. „GeoNames“

18. „Creative Commons“

19. Europeana, „Europeana Publishing Framework: Metadata“, преузето 15. 3. 2022

20. Tier A, Tier B, Tier C.

21. Europeana, „The more you give, the more you get“, преузето 18. 3. 2022

```

▼ <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2022-02-14T10:50:18Z</responseDate>
  <request metadataPrefix="edm" set="ECC_kun1_EDU_0kt_20_IP" verb="ListRecords">http://aggregator.nb.rs/oai/OAIHandler</request>
  ▼ <ListRecords>
    ▼ <records>
      ▼ <header>
        <identifier>2354823</identifier>
        <datestamp>2020-10-19T00:00:00Z</datestamp>
      </header>
      ▼ <metadata>
        ▼ <rdf:RDF xmlns:ore="http://www.openarchives.org/ore/terms/" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:rdfs="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:edm="http://www.europeana.eu/schemas/edm/" xmlns:ns13="http://www.w3.org/ns/odrl/2/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:ns12="http://xmlns.com/foaf/0.1/" xmlns:ns11="http://rdfs.org/sioc/services#"
xmlns:ns10="http://www.dcu.ch/metadata/ontologies/ebcore/ebucore#" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:ns16="http://www.europeana.eu/schemas/edm/enrichment/"
xmlns:ns15="http://usefulinc.com/ns/dpns#" xmlns:ns14="http://creativecommons.org/ns#" xmlns:dc="http://purl.org/dc/elements/1.1/">
          ▼ <skos:Concept rdf:about="/direct/2354833">
            <skos:prefLabel xml:lang="sr">Внујерство Кнежеве</skos:prefLabel>
            </skos:Concept>
          ▼ <edm:Agent rdf:about="/direct/2354828">
            <skos:prefLabel xml:lang="en">Jovan Popović</skos:prefLabel>
            <skos:note xml:lang="en">Latin script</skos:note>
            <rdar2:dateOfBirth>1905</rdar2:dateOfBirth>
            <rdar2:dateOfDeath>1952</rdar2:dateOfDeath>
            </edm:Agent>
          ▼ <skos:Concept rdf:about="/direct/2354826">
            <skos:prefLabel xml:lang="sr">Нопље Јањеђевић-Кун</skos:prefLabel>
            </skos:Concept>
          ▼ <edm:Agent rdf:about="/direct/2354829">
            <skos:prefLabel xml:lang="en">Нопље Јањеђевић-Кун</skos:prefLabel>
            <skos:note xml:lang="en">Cyrillic script</skos:note>
            <rdar2:dateOfBirth>1904</rdar2:dateOfBirth>
            <rdar2:dateOfDeath>1964</rdar2:dateOfDeath>
            </edm:Agent>
          ▼ <edm:Agent rdf:about="/direct/2354832">
            <skos:prefLabel xml:lang="sr">Јован Поповић</skos:prefLabel>
            <skos:note xml:lang="en">Cyrillic script</skos:note>
            <rdar2:dateOfBirth>1905</rdar2:dateOfBirth>
            <rdar2:dateOfDeath>1952</rdar2:dateOfDeath>
            </edm:Agent>
          ▼ <edm:Agent rdf:about="/direct/2354827">
            <skos:prefLabel xml:lang="sr">Јорђе Андрејевић-Кун</skos:prefLabel>
            <skos:note xml:lang="en">Latin script</skos:note>
            <rdar2:dateOfBirth>1904</rdar2:dateOfBirth>
            <rdar2:dateOfDeath>1964</rdar2:dateOfDeath>
            </edm:Agent>
          ▼ <edm:Agent rdf:about="/direct/2354838">
            <skos:prefLabel xml:lang="sr">http://www.wikidata.org/wiki/Q1283347</skos:prefLabel>
            </edm:Agent>
          ▼ <skos:Concept rdf:about="/direct/2354825">
            <skos:prefLabel xml:lang="sr">војска</skos:prefLabel>
            </skos:Concept>
          ▼ <edm:Agent rdf:about="/direct/2354824">
            <skos:prefLabel xml:lang="sr">http://www.wikidata.org/wiki/Q6296683</skos:prefLabel>
            </edm:Agent>
          ▼ <skos:Concept rdf:about="/direct/2354831">
            <skos:prefLabel xml:lang="en">Industrial Heritage</skos:prefLabel>
            </skos:Concept>
          ▼ <edm:ProvidedCHO rdf:about="URI:RS:NAE:d27476c7-f752-4a15-901b-e7e802b5f7eecho">
            <dc:identifier>URI:RS:NAE:d27476c7-f752-4a15-901b-e7e802b5f7eecho</dc:identifier>
            <dc:identifier>170197516</dc:identifier>
            <dc:language>sr</dc:language>
            <dc:terms:issued>1937</dc:terms:issued>
            <dc:currentLocation rdf:resource="http://sus.geonames.org/792680"/>
            <dc:title xml:lang="sr">Крваво злато : 28 оригиналних дрвенеа Ђ. Андрејевића-Куне</dc:title>
            <dc:description xml:lang="sr">Прва зба. </dc:description>
            <dc:description xml:lang="sr">Стр. 1-11: Уметник и његово дело / Ж. Поповић.</dc:description>
            <dc:publisher xml:lang="sr">Ђ. Андрејевић-Кун</dc:publisher>
            <dc:format xml:lang="sr">32 стр., 28 листова с таблама</dc:format>
            <dc:format xml:lang="sr">дрвенеи, црно-бело</dc:format>
            <dc:format xml:lang="sr">25 cm</dc:format>
            <dc:subject rdf:resource="/direct/2354833"/>
            <dc:contributor rdf:resource="/direct/2354828"/>
            <dc:creator rdf:resource="/direct/2354829"/>
            <dc:contributor rdf:resource="/direct/2354832"/>
            <dc:creator rdf:resource="/direct/2354827"/>
            <dc:contributor rdf:resource="/direct/2354830"/>
            <dc:type rdf:resource="/direct/2354825"/>
            <dc:contributor rdf:resource="/direct/2354824"/>
            <dc:subject rdf:resource="/direct/2354831"/>
            <edm:type>PAGE</edm:type>
          </edm:ProvidedCHO>
          ▼ <edm:Aggregation rdf:about="URI:RS:NAE:d27476c7-f752-4a15-901b-e7e802b5f7eecho">
            <edm:aggregatedCHO rdf:resource="http://www.digitalna.nb.rs/nb/NBS/Tematske_kolekcije/Industrijsko_nasledje/borski_rudnici/Krvaavo_zlato/GH_85"/>
            <edm:dataProvider>National Library of Serbia</edm:dataProvider>
            <edm:isShownAt rdf:resource="http://www.digitalna.nb.rs/nb/NBS/Tematske_kolekcije/Industrijsko_nasledje/borski_rudnici/Krvaavo_zlato/GH_85"/>
            <edm:isShownBy rdf:resource="http://www.digitalna.nb.rs/attach/NBS/Tematske_kolekcije/Industrijsko_nasledje/borski_rudnici/Krvaavo_zlato/GH_85/output.pdf"/>
            <edm:object rdf:resource="http://www.digitalna.nb.rs/nb/NBS/Tematske_kolekcije/Industrijsko_nasledje/borski_rudnici/Krvaavo_zlato/GH_85/pageindex-thumb"/>
            <edm:provider>National Library of Serbia</edm:provider>
            <edm:rights rdf:resource="http://creativecommons.org/licenses/by-nc-sa/4.0/">
          </edm:Aggregation>
        </rdf:RDF>
      </metadata>
    </records>
  </ListRecords>
</OAI-PMH>

```

Слика 2. Пример истог записа након трансформације коју врши Агрегатор за Европеану

тачније, што су богатији, разноврснији и потпунији метаподаци, то ће резултат у претрази бити бољи, тј. тражени објекат ће бити понуђен као резултат претраге. Ниво А означава да ће садржај бити приказан на страници Европеане, ниво В, поред приказа на страници Европеане, омогућава и коришћење у различитим тематским колекцијама, као и употребу у различитим рекламним садржајима (нпр. објаве на друштвеним мрежама). Ниво С је највиши и, самим тим, има највише метаподатака. Он, поред свега наведеног за претходне нивое, може да се користи и у различитим пројектима, апликацијама и сервисима. Метаподаци које је НБС досад доставила Европеани припадају углавном нивоу В.

Када је садржај²² у питању, важи исти принцип: „што више дате више ћете добити“. Што је већи квалитет, то је више могућности да шири круг корисника види садржај, али, исто тако, и да га поново употреби. Разликују се четири нивоа квалитета садржаја.²³ Као и код метаподатака, најнижи ниво омогућава само да објекат буде пронађен на Европеани, док највиши ниво омогућава поновну употребу у комерцијалне сврхе, али и на различитим платформама и апликацијама. Да би садржај био нивоа 4 мора да има лиценцу која омогућава слободну употребу, тј. јавни приступ (PDM, CC-0, CC-BY, CC-BY-SA) (Касадевал and Бутиган Вучај 2020). Подаци које је до сада НБС доставила су углавном нивоа 3 или 4 (зависно од врсте лиценце). Још један услов који утиче на ниво квалитета садржаја је и присуство PDF-а за текстуалну грађу (НБС је усвојила као начин рада обавезан рашчитан PDF за текстуалну грађу). Народна библиотека Србије се, такође, придржава правила да скенирани материјал буде у JPG формату резолуције 0,95 mpx тј. минималне величине 1.200x800 пиксела²⁴ (што је неопходан услов за нивое 3 и 4).

Како Европеана прикупља податке из установа културе које се баве наслеђем из читаве Европе, да би их објединила у јединствену платформу, она настоји да ауторско право не стоји на путу дигитализацији и креирању збирки које су доступне јавности. У области ауторских права кључни документ је „Оквир за додељивање лиценци“²⁵ чији елементи стандардизују и усклађују информације и праксе у вези

22. Europeana, „Europeana Publishing Framework: Content“, преузето 16. 3. 2022

23. Tier 1, Tier 2, Tier 3, Tier 4.

24. Europeana, „Europeana publishing guide“, преузето 16. 3. 2022

25. Europeana, „Europeana Licensing Framework“, преузето 16. 3. 2022

са ауторским правима и чине да однос између крајњих корисника и установа које достављају податке буде јасан и транспарентан. Европеана користи лиценцу Креативне заједнице, али и лиценце конзорцијума Rights Statements,²⁶ чији је суоснивач.

4. Унапређење агрегаторске структуре

Кроз пројекат „Заједничка култура у Европани“, који је трајао од почетка 2019. до краја 2020, рађено је на унапређењу агрегаторске инфраструктуре. Циљ пројекта био је, као што је већ споменуто, да се установе заједничке препоруке и стандарди, обогате метаподаци и омогући што слободнији приступ садржају. Европеана се залаже да све публикације буду у слободном приступу. НБС је ажурирала метаподатке за 15 збирки, укључених у Европеану кроз раније пројекте, и послала осам нових збирки. Да би метаподаци могли да се ажурирају и допуне, сама агрегаторска структура требало је да буде програмски допуњена. Технички партнер НБС у успостављању Агрегатора за Европеану је словеначка фирма Семантика. За потребе пројекта Агрегатор је допуњен просторним и временским одредницама, метаподацима о ауторству у којима се приказују година рођења и смрти, линком ка бази података Wikidata, као и податком о писму на коме је наведен аутор. Овај последњи податак важан је због специфичности српског језика који користи два писма.

На Слици 3 приказано је како изгледају допуњени метаподаци о аутору публикације и писцу предговора. У оквиру етикета <dc:creator> и <dc:contributor>, поред имена, додате су године рођења и смрти, а у оквиру атрибута xml:lang је вредност sr која означава српски језик. Линк ка бази података Wikidata је записан као вредност атрибута rdf:resource.

```
<dc:contributor xml:lang="sr">Jovan Popović (1905-1952)</dc:contributor>
<dc:contributor xml:lang="sr">Јован Поповић (1905-1952)</dc:contributor>
<dc:contributor rdf:resource="http://www.wikidata.org/wiki/Q6296683" xml:lang="sr"/>
<dc:creator xml:lang="sr">Đorđe Andrejević-Kun (1904-1964)</dc:creator>
<dc:creator xml:lang="sr">Ђорђе Андрејевић-Кун (1904-1964)</dc:creator>
<dc:creator rdf:resource="http://www.wikidata.org/wiki/Q1283347" xml:lang="sr"/>
```

Слика 3. Обогађивање метаподатака у примеру са српским ауторима

26. „Rights Statements“

На Слици 4 боље је приказана разлика између вредности атрибута `xml:lang`. У овом примеру за једног аутора постоје по три етикете `<dc:creator>` у којима је приказано име аутора на немачком језику (`xml:lang="de"`), ћирилици и латиници. Уколико аутор не постоји у бази података Wikidata, тај податак се изоставља (код Лудвига Ерминија не постоји линк ка бази података Wikidata)

```
<dc:creator xml:lang="de">Ludwig Erminy</dc:creator>
<dc:creator xml:lang="sr">Ludvig Ermini</dc:creator>
<dc:creator xml:lang="sr">Лудвиг Ермини</dc:creator>
<dc:creator xml:lang="de">Franz Wolf (1795-1859)</dc:creator>
<dc:creator xml:lang="sr">Franc Volf (1795-1859)</dc:creator>
<dc:creator xml:lang="sr">Франц Волф (1795-1859)</dc:creator>
<dc:creator rdf:resource="http://www.wikidata.org/wiki/Q20772339" xml:lang="sr"/>
```

Слика 4. Обогаћивање метаподатака у примеру са страним ауторима

Ова два примера су део EDM.xml записа. Када се направљени EDM подигне на агрегатор добије се мало другачији документ. Трансформација која постоји у самом агрегатору свако поље пребацује у одређени еквивалент и као резултат имамо запис приказан на Слици 5.

Подаци о години рођења и смрти аутора трансформишу се у етикете `<rdsGr2:dateOfBirth>` и `<rdsGr2:dateOfDeath>` у оквиру `<edm:Agent>` који представља аутора. Уколико не постоје подаци о рођењу и смрти, ове етикете ће бити изостављене. Линк ка бази података Wikidata наведен је у оквиру етикете `<skos:prefLabel>`, док је у `<skos:note>` обележено да ли је у питању ћирилично или латинично писмо, а програм сам препознаје које писмо је у питању.

Линк са агрегатора шаље се Европеани и као крајњи резултат²⁷ добијамо следећи испис метаподатака (Слика 6):

Још једна промена која је направљена на агрегатору је приказ временских и просторних одредница. У EDM су додате етикете `<dcterms:spatial>` и `<dcterms:temporal>` и оне трансформацијом у агрегатору постају `<edm:Place>` и `<edm:TimeSpan>`.

Пример просторне одреднице и временске одреднице у EDM-у дат је на Слици 7.

Како на агрегатору изгледа овај део записа приказано је на Слици 8.

27. „Krvavo zlato: 28 originalnih drvoreza Ђ. Andrejevića-Kuna“

```

▼<edm:Agent rdf:about="/direct/2354828">
  <skos:prefLabel xml:lang="sr">Jovan Popović</skos:prefLabel>
  <skos:note xml:lang="en">Latin script</skos:note>
  <rdaGr2:dateOfBirth>1905</rdaGr2:dateOfBirth>
  <rdaGr2:dateOfDeath>1952</rdaGr2:dateOfDeath>
</edm:Agent>
▼<skos:Concept rdf:about="/direct/2354826">
  <skos:prefLabel xml:lang="sr">Ђорђе Андрејевић-Кун</skos:prefLabel>
</skos:Concept>
▼<edm:Agent rdf:about="/direct/2354829">
  <skos:prefLabel xml:lang="sr">Ђорђе Андрејевић-Кун</skos:prefLabel>
  <skos:note xml:lang="en">Cyrillic script</skos:note>
  <rdaGr2:dateOfBirth>1904</rdaGr2:dateOfBirth>
  <rdaGr2:dateOfDeath>1964</rdaGr2:dateOfDeath>
</edm:Agent>
▼<edm:Agent rdf:about="/direct/2354832">
  <skos:prefLabel xml:lang="sr">Јован Поповић</skos:prefLabel>
  <skos:note xml:lang="en">Cyrillic script</skos:note>
  <rdaGr2:dateOfBirth>1905</rdaGr2:dateOfBirth>
  <rdaGr2:dateOfDeath>1952</rdaGr2:dateOfDeath>
</edm:Agent>
▼<edm:Agent rdf:about="/direct/2354827">
  <skos:prefLabel xml:lang="sr">Ђорђе Андрејевић-Кун</skos:prefLabel>
  <skos:note xml:lang="en">Latin script</skos:note>
  <rdaGr2:dateOfBirth>1904</rdaGr2:dateOfBirth>
  <rdaGr2:dateOfDeath>1964</rdaGr2:dateOfDeath>
</edm:Agent>
▼<edm:Agent rdf:about="/direct/2354830">
  <skos:prefLabel xml:lang="sr">http://www.wikidata.org/wiki/Q1283347</skos:prefLabel>
</edm:Agent>
▼<skos:Concept rdf:about="/direct/2354825">
  <skos:prefLabel xml:lang="sr">жигра</skos:prefLabel>
</skos:Concept>
▼<edm:Agent rdf:about="/direct/2354824">
  <skos:prefLabel xml:lang="sr">http://www.wikidata.org/wiki/Q6296683</skos:prefLabel>
</edm:Agent>

```

Слика 5. Део EDM-а са агрегатора који се односи на српске ауторе

Најновије унапређење на Агрегатору за Европеану у Народној библиотеци Србије односи се на дигиталне објекте који за приказ користе технологију IIIF.²⁸ Ова технологија подразумева отворени стандард за онлајн испоручивање висококвалитетних дигиталних објеката. Под овим акронимом подразумева се и отворена међународна заједница за развој и имплементацију ове технологије, уз подршку конзорцијума водећих институција културе, као и академских институција. Захтев је стигао из Библиотеке „Милутин Бојић“, а овом интервенцијом на агрегатору биће подржани скупови података из свих библиотека које користе IIIF. EDM спецификација за IIIF захтева четири корака:

1. да се достави IIIF Resource као <edm:WebResource>, што у пракси значи да се наведе идентификатор за WebResource који служи за “преглед” објекта и да се идентификује тип везе између IIIF Resource-а за објекат и самог објекта;
2. да се означи да је <edm:WebResource> усклађен са IIIF-ом, што у пракси значи да се повеже WebResource са извором типа

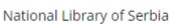
28. “IIIF”

Krivo zlato : 28 originalnih drvoreza Đ. Andrejevića-Kuna

Tiraž 250.

[Read more](#)

COLLECTIONS YOU MIGHT LIKE

 20th century  National Library of Serbia

Good to know All metadata

Providing institution	National Library of Serbia 
Contributors	Јован Поповић ; http://www.wikidata.org/wiki/Q6296683 ; Jovan Popović
Creator	http://www.wikidata.org/wiki/Q1283347 ; Ђорђе Андрејевић-Кун ; Đorđe Andrejević-Kun
Publisher	Đ. Andrejević-Kun
Subject	Ђорђе Андрејевић-Кун ; Индустијско наслеђе ; Industrial Heritage
Type of object	књига

Слика 6. Приказ метаподатака једног објекта на Европеани

```
<dcterms:issued>1921</dcterms:issued>
<dcterms:spatial xml:lang="sr">Југославија</dcterms:spatial>
<dcterms:temporal xml:lang="sr">1921</dcterms:temporal>
<edm:currentLocation rdf:resource="http://sws.geonames.org/792680/" />
```

Слика 7. Пример просторне одреднице и временске одреднице у EDM

```
▼<edm:Place rdf:about="/direct/2352735">
  <skos:prefLabel xml:lang="sr">Југославија</skos:prefLabel>
</edm:Place>
▼<edm:TimeSpan rdf:about="/direct/2352732">
  <skos:prefLabel xml:lang="sr">1921</skos:prefLabel>
</edm:TimeSpan>
```

Слика 8. Део записа са слике 7 на агрегатору

- <svcs:Service>, као и да се наведе да WebService одговара профилу IIIF;
- да се назначи ниво имплементације IIIF-а путем етикете <doap:implements>;
 - да се омогући приступ манифесту IIIF директно или индиректно преко <dcterms:isReferencedBy>.

Сегмент EDM-а за конкретан дигитални објекат Библиотеке „Милутин Бојић“, а у питању је руком писана *Бележница* Милутина Бојића,²⁹ који Агрегатор за Европеану у НБС треба да испоручи да би била објављена на Европеани, представљен је на Слици 9.

```
<!--
  edm:WebResource class - contains all information about the
  digital version of the CHO only-->
<edm:WebResource
rdf:about="https://milutinbojic.digitalna.rs/servlet/ImageTileRenderer?doc_id=0f5e3e1-56af-416f-9d98-79f165e07d4c/00000001/ruko
pis/00000021&pg_seq=1&search_doc=&scale=0.2&rotation=0">
  <dc:format>jpg</dc:format>
  <dcterms:isReferencedBy
rdf:resource="https://milutinbojic.digitalna.rs/iiif/api/presentation/2/0f5e3e1-56af-416f-9d98-79f165e07d4c%252F00000001%252Fru
kopis%252F00000021/manifest"/>
  <svcs:has_service
rdf:resource="https://milutinbojic.digitalna.rs/iiif/api/image/2/0f5e3e1-56af-416f-9d98-79f165e07d4c%252F00000001%252Frukopis%252
F00000021%252F0"/>
  </edm:WebResource>
<!--
  svcs:Service class - contains all information about the
  IIIF resource-->
<svcs:Service
rdf:about="https://milutinbojic.digitalna.rs/iiif/api/image/2/0f5e3e1-56af-416f-9d98-79f165e07d4c%252F00000001%252Frukopis%252
F00000021%252F0">
  <dcterms:conformsTo rdf:resource="http://iiif.io/api/image"/>
  <doap:implements rdf:resource="http://iiif.io/api/image/2/level2.json"/>
  </svcs:Service>
```

Слика 9. Сегмент EDM-а за IIIF

5. Закључак

Народна библиотека Србије је, кроз сарадњу са Европеаном и рад на агрегатору, усвојила у свом раду нека правила и стандарде које је Европеана прописала (ЕДМ, минимална величина скенираних објеката, рапчитани PDF, лиценце за коришћење дигиталних докумената) и тако достигла европску праксу у овој области, а што је још важније, доказала

²⁹ Библиотека Милутин Бојић – Дигитални репозиторијум, „Бележница Милутина Бојића“

да то није немогуће и то са малим тимом стручњака. У плану је објављивање нових садржаја на платформи Европеане, али и наставак рада на побољшању квалитета метаподатака, садржаја, као и на даљем развоју агрегатора. Развој агрегатора би се кретао у два правца: ка укључивању нових партнера за достављање података, како би културно наслеђе Србије било што боље представљено у европским оквирима, и наравно ка даљем праћењу правила и стандарда Европеане.

Литература

- Бутиган-Вучај, Тамара. 2019. “Национални агрегатор за Европеану у Србији: мали корак ближе Европској унији.” *Библиозона* 7:15–19.
- Дреге, Евелин. 2019. “Будућност прошлости: Културно наслеђе у дигиталном свету.” *Чап. Од ESE до EDM и даље: на који начин Европеана омогућава приступ објектима културног наслеђа*, edited by Паул Климпел и Елен Ојлер, 87–111. Београд: Народна библиотека Србије.
- Касадевал, Аријадна Матас, and Тамара Бутиган Вучај. 2020. “Европеана кроз објектив ауторских права.” *Гласник НБС* 22:41–56.
- Климпел, Паул, and Елен Ојлер. 2019. *Будућност прошлости: културно наслеђе у дигиталном свету*. Београд: Народна библиотека Србије.

Мултимедијални пројекат *Добродошли у мој најдражи кутак*

УДК 004.55:378.147]:02(497.11)

Валентина Ивановић
valentinaivanovic2@gmail.com

Милица Чекетић
sarangfree@gmail.com

Јована Стојичић
jovanastojic98@gmail.com

Душица Галић
dusicagalic7@gmail.com

Универзитет у Београду
Филолошки факултет
Београд, Србија

САЖЕТАК: Предмет овог рада је представљање процеса израде пројекта у оквиру курса Мултимедијални документи, који се похађа на завршној години основних академских студија на Катедри за библиотекаство и информатику Филолошког факултета у Београду. Пројекат је резултат рада студената генерације 2017/2018. под менторством доц. др Браниславе Шандрих Тодоровић и асистенткиње мср. Милице Иконић Нешић, уз помоћ професорке др Цветане Крстев. „Добродошли у мој најдражи кутак” је била тема пројекта која је за циљ имала представљање места која су драга студентима као веб презентације.

КЉУЧНЕ РЕЧИ: информатика, мултимедијални документи, HTML, CSS, библиотекаство

РАД ПРИМЉЕН: 16. новембар 2021.

РАД ПРИХВАЋЕН: 01. април 2022.

1. Увод

Мултимедија, као обједињавање текста, слике и звука, представља ефикасан начин коришћења савремене технологије у циљу очувања тековина једног друштва. Мултимедија је, заправо, информација представљена или меморисана у виду комбинације текста, графике, звука, анимације и видеа који су обједињени помоћу рачунарских програма. Она кориснику омогућава лакше сналажење и детаљније упознавање са темом о којој се говори, мотивише корисника и одржава његову пажњу, али такође помаже у разјашњавању недоумица (Тртовац 2010).

Као стечиште знања, библиотека мора бити умешна у руковању свим видовима информација које се производе и ићи у стопу са свим облицима развоја технологије и потребама корисника. Самим тим, писменост у оном изворном значењу („The ability to read and write”¹) за библиотекара није довољна, никада није једнозначна и увек мора одражавати производе промена које са собом носи ново доба, а један од тих производа је управо и мултимедија.

Сходно томе, програм Катедре за библиотекарство и информатику у оквиру четири године обухвата многе информатичке предмете, попут Информатике за библиотекаре, Информатичког практикума, Дигиталног текста, База података, Структуре информација, које за циљ имају свеколико информатичко оспособљавање студената за будући рад у информатичком (али и информационом) пољу. Управо курс Мултимедијални документи служи као крајња провера свог знања стеченог на споменутим курсевима, кроз израду једног заједничког пројекта. Важност овог пројекта огледа се и у томе што, поред провере знања, студенти развијају тимски и истраживачки дух, који су важни за сваки вид усавршавања.

На овом пројекту радили су студенти четврте године Катедре за библиотекарство и информатику Филолошког факултета у Београду генерације 2017/2018. и овде су објединили сва занимљива места у свом родном граду која би посетили или показали људима који им долазе у посету. Тема овогодишњег пројекта је „Добродошли у мој најдражи кутак”. Ментори пројекта биле су доц. др Бранислава Шандрих Тодоровић и мср. Милица Иконић Нешић, асистенткиња. У документу су приказана следећа места: Београд, Земун, Ужице, Смедерево, Шабац, Браничево, Голубац, Обреновац и Гроцка. Студенти који су учествовали у пројекту су: Валентина Ивановић, Милица Чекетић, Даница Кричковић, Јована Стојичић, Душица Галић, Сања Игњатовић, Маја Савић, Марија Ћинић и Драгиша Стокић.

2. Инспирација за израду пројекта

„Добродошли у мој најдражи кутак” није иницијално била тема намењена завршној години генерације 2017/2018. Првобитна тема за студенте је била „Знамените личности мога краја”, где би сваки студент одабрао до 10 личности којима се одређено место поноси и би било

1. „Literacy” Cambridge Dictionary (посећено 25. 3. 2022).

потребно детаљније представити о којима би писао. Међутим, убрзо су се појавиле недоумице око опсега теме (да ли је у питању само личност која је ту рођена и ту се прославила, или се ту може сврстати и личност која се ту прославила, иако ту није рођена, или је то може бити и личност која је ту живела након што се прославила, а ту није рођена. . .).

Због тога је професорка Бранислава Шандрих Тодоровић предложила тему конкретнијег концепта, „Добродошли у мој најдражи кутак”, коју су студенти прихватили. Основна замисао реализације ове теме је једноставна: изабрати места која бисмо показали госту који први пут долази код нас. У свом одабиру студенти нису имали ограничења, важност коју су наведена места за студента имала нису морала бити строго укорењена у историјско-културном контексту. Студенти нису морали да се ограниче само на музеје, споменике, позоришта, већ им је дата слобода да представе места која чине неизоставни део њихове свакодневнице и улепшавају је или носе неки другачији вид сентименталне вредности за њих (Каленић пијаца коју је студенткиња посећивала са баком док је била мала, пољана под називом Баре на којој се лети окупљају млади близу села Браничева...). Такође, овај пројекат је пружио могућност студентима да сазнају нешто више о местима која су изабрали, што им можда раније није било познато (Павиљон Великог рефрактора Астрономске опсерваторије набавио је Милан Недељковић на име ратне одштете након II светског рата (Михајлов 2010), али и да се, преко својих колега, упознају са местима за која нису чули или су о њима мало знали. Битно је нагласити да је, сходно околностима изазваних пандемијом одлучено да се ресторани и слични угоститељски објекти ове године избегну.

3. Класификација места

Сваки студент требало је да изабере између 5 и 10 места која су, пре почетка израде саме веб-странице, представљена путем презентација, у оквиру којих је било речи и о основним одликама изабраног места, уз обавезне коментаре који појашњавају шта то место чини посебним за студента.

Професорка Шандрих Тодоровић је у оквиру Google Drive-a поставила табелу у коју су студенти уносили своје предлоге, који су накнадно били класификовани по критеријумима који су договорени са професорком. Класификација је подразумевала 8 категорија и 21 поткатегорију:

1. КУЛТУРНИ ЦЕНТРИ- библиотеке, музеји, позоришта/биоскопи, научне установе, галерије
2. МЕСТА ЗА РЕКРЕАЦИЈУ- спортски центри, игралишта
3. СПОМЕНИЦИ- тврђаве, куле
4. ТРГОВИ
5. ВЕРСКИ ОБЈЕКТИ- цркве, манастири
6. ВОДЕНЕ ОАЗЕ- купалишта, кејеви, реке, језера
7. ЗАНАТИ И ТРГОВИНА- занатске радње, пијаце
8. ЗЕЛЕНЕ ОАЗЕ- паркови, шеталишта, баште, видиковци

Након класификације, уследило је прикупљање материјала за израду презентација и веб-странице.

4. Прикупљање садржаја

За прикупљање грађе за овај пројекат, у периоду од четири месеца, сваки студент је одлазио на место које је одабрао за пројекат. Студенти су се потрудили да сами фотографишу свако место које су посетили, а текстови за одређена места преузимани су из бројних извора, које су обухватале монографије (попут „Шареног света казивања” Јована Ђорђевића или књиге Даринке Лекић „Београд, град за сва времена”), периодику(чланке из дневних новина, попут Политике), усмени извори, веб-странице и сл..Прикупљање текстова подразумевало је одласке у локалне библиотеке, претрагу кућних библиотека, информација на интернету и званичних веб-страница уколико је неко од одабраних места то поседовало.Усмени извори обухватали су локалне приче и легенде везане за конкретно место са којим су студенти одрастали.

При прикупљању садржаја студенти су наилазили на препреке попут радова, реновирања и незгодног радног времена локалитета што их је у неким случајевима спречавало да дођу до потребних фотографија. Уколико студенти нису могли да изађу на крај са наведеним препрекама, проналазили су нове локалитете које би уврстили у пројекат.

Након завршеног прикупљања грађе, студенти су се договорили како да уобличи садржај и како ће он бити представљен на веб-страници.

5. Израда и дизајн веб-странице

Као инспирацију за изглед веб-странице тим је изабрао пчелиње саће. Идеја иза тог избора била је да, иако долазимо из различитих крајева

земље, ипак припадамо једној кошници унутар које сарађујемо, а то је Катедра за библиотекарство и информатику. Цела страница подељена је на три нивоа, највише из практичних разлога и ради ефикасније расподеле задужења. Први ниво странице је Насловна страна (Homepage), она је полазна тачка за даљу навигацију страницом и на њој се налазе опште категорије. Други ниво садржи називе локалитета повезаних са општим категоријама првог нивоа. Трећи ниво представља личне стране студената на којима се налазе локалитети који њима представљају најдраже кутке.

При изради веб-странице коришћене су технологије са којима се студенти сусрећу на другој години студија, у оквиру предмета Дигитални текст 2. HTML² је основни описни језик који је специјално развијен за употребу при изради веб-страница,³ док се CSS⁴ користи за дефинисање изгледа HTML елемената на веб-страници (посећено 8.11.2021).⁵ Одабир HTML-а и CSS-а показао се веома практичним с обзиром на то да су ове технологије студентима добро познате и једноставне су за употребу. Такође, HTML у комбинацији са CSS-ом пружа бројна креативна решења када су у питању пројекти попут овог из Мултимедијалних докумената с обзиром на то да је садржај богати фотографијама, анимацијама, музиком и другим медијима.

5.1 Израда првог степена веб-странице

У раним фазама пројекта тим је одлучио да изглед првог степена буде у облику кошнице са саћима, а за палету су одабране боја меда и боја чоколаде које одају пријатан утисак. За израду Насловне стране и обрасца (template) за остала два степена била је задужена чланица тима Валентина Ивановић. На наслов странице (Слика 1) „Добродошли у мој кутак“ стављен је ефекат неонског знака помоћу CSS-а, а фонт „Sacramento“ је преузет са Google Fonts платформе која пружа бесплатан приступ бројним фонтовима.

2. акроним од Hyper Text Markup Language

3. „HTML introduction“ (посећено 8.11.2021).

4. Cascading Style Sheets

5. „CSS introduction“



Слика 1. Изглед Насловне стране

На свако саће стављена је анимација која, када се курсор корисника нађе изнад саћа (hover), чини да нестане назив опште категорије и открива се слика којом је категорија представљена. Ауторка фотографија је чланица тима, Валентина Ивановић. Поред анимација и слика, саћа имају функцију, кликом на њих воде до другог степена са којим су повезане путем релативних адреса (Слика 2).



Слика 2. Приказ анимације саћа, пре и после

Поред саћа, на Насловној страни се налази анимирано дугме које води до странице „О пројекту“ која припада другом степеноу. Чланови

тима су такође изразили жељу да се на Насловној страни чује музика, те је помоћу HTML-а додат плејер, а изабрана је композиција „Нол” италијанског пијанисте Лудовика Еинаудија.

Још један од ефеката који се може приметити на Насловној страни је суптилно преливање боја позадине, од светлије до тамније нијансе наранџасте. Ефекат је постигнут помоћу CSS својства `animation` унутар етикете `body`.

5.2 Израда другог и трећег степена веб-странице

У раним фазама пројекта одлучено је да ће се за други степен веб-странице користити образац по коме је израђена Насловна страна, ради уједначенијег изгледа, док је за трећи степен одлучено да сваки студент има потпуну уметничку слободу при изради.

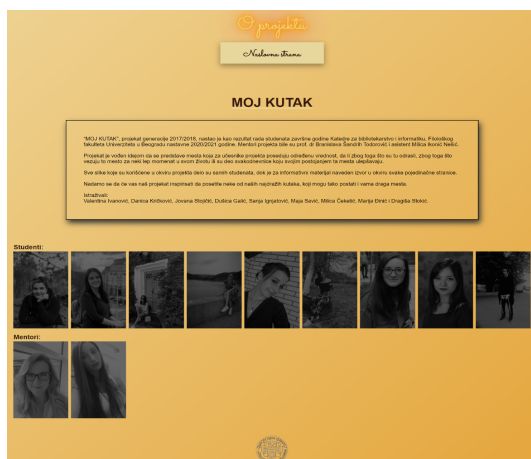
За дизајн другог степена студенти су такође имали дозу слободе, те је сваки члан тима који је био задужен за једну страницу другог степена, одабрао слике и боју којима ће она бити украшена. Свака страница другог степена носи наслов опште категорије са Насловне стране којој припада и саћа, релативне адресе, која воде до одређених локалитета. Како би осигурали да ће кликом на саће корисник доћи до одређеног дела трећег степена који говори о жељеном локалитету, користили су сидра (anchor). Поред „кошнице“, на свакој страни другог степена налазе се три анимирана дугмета која воде до претходне категорије, Насловне стране и следеће категорије (Слика 3).

Посебан део другог степена чини страница „О пројекту“ на којој корисник може наћи основне податке о пројекту, чланове тима, њихове фотографије (повезане са трећим степеном сваког члана тима) (Слика 4) и фотографије менторки доц. др Браниславе Шандрих Тодоровић и асистенткиње Милице Иконић Нешић (њихове фотографије су, такође, повезане са персоналним веб-страницама).

Када је реч о трећем степену, који је представља личне странице сваког од чланова тима, неки су користили образац, а неки су се користили сопственим идејама. Највећи број чланова тима је користио HTML и CSS, док је једна чланица тима, Душица Галић, користила донекле и JavaScript. Сваки трећи степен поседује два обавезна дугмета, а то су дугме за повратак на Насловну страну и дугме за страну која води до података о аутору и коришћене литературе. У оквиру својих страница, чланови тима су представили своје најдраже кутке, најчешће из родних места, што је овом пројекту донело разноврсност и боље

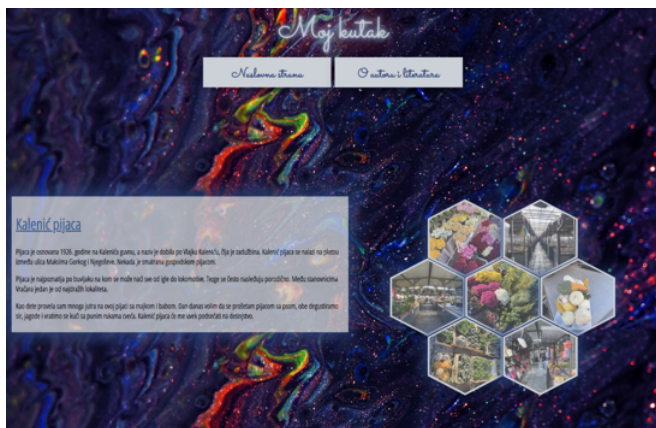


Слика 3. Приказ другог степена



Слика 4. Приказ странице „О пројекту“

упознавање са земљом у којој живимо (Слике 5 и 6) и на тај начин се придружили низу сличних пројеката које су радиле раније генерација за потребе курса Мултимедијални документи, попут пројекта „Ал’ се некад добро јело” генерације 2012/2013, у оквиру ког су студенти сакупљали материјал везан за традиционалну српску кухињу, или пројекта „Игра рокенрол цела Југославија” генерације 2014/2015, која је за циљ имала „да на једном месту прикаже што обухватнију грађу о почецима и развоју рок сцене на просторима бивше Југославије”.⁶



Слика 5. Приказ стране трећег степена

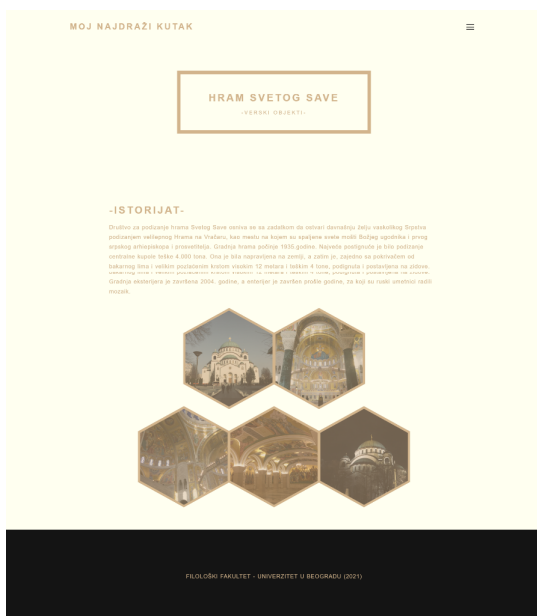
6. Закључак

Захваљујући предмету Мултимедијални документи, студенти су имали прилику да на завршној години основних студија обједине своја знања стечена током претходних година. Лекције које су у току студирања могле деловати небитно и апстрактно, информације које су могле деловати нејасно, уобличене су и у потпуности усвојене у оквиру рада на овом пројекту.

Кроз лична залагања, слогу и марљив рад свих студената, као и помоћ професора, завршни пројекат на тему „Добродошли у мој најдражи кутак“ представља драгоцену искуство стечено током летњег семестра 2020/2021. године. Мотивисани и жељни да на најкреативнији

6. „Игра рокенрол цела Југославија” (посећено 25.3.2022).

начин представе делове наше земље из којих долазе или места која за њих имају неку посебну вредност, студенти су се лако и брзо договорили око свих делова пројекта. Неслагања готово да није ни било, па је сам пројекат завршен у року без већих потешкоћа.



Слика 6. Приказ стране трећег степена

Иако су локалитети представљени у пројекту изабрани јер су студентима посебно драги, те је критеријум за одабир био вођен емоцијама и лепим успосменама, многи од локалитета носе историјски и културни значај за целу заједницу. Имајући ове чињенице у виду, пројекат се може посматрати и као вид туристичког водича који посетиоце странице упућује на интересантне локалитете које би могли сами посетити.

Надамо се да ће вас наш пројекат инспирисати да посетите неке од наших најдражих кутака, који могу тако постати и вама драга места.⁷

7. "О пројекту", Добродошли у мој најдражи кутак (посећено 11.11.2021).

Захвалност

Велику захвалност дугујемо професорки др Цветани Крстев, професорки др Бранислави Шандрих Тодоровић и асистенткињи Милицы Иконић Нешић, за стрпљиво и стручно усмеравање, спремност да увек издвоје време за наша питања и за сво пренето знање, не само у току израде овог пројекта, већ и у току претходне четири године нашег студирања.

Литература

- Михајлов, Саша. 2010. *Астрономска опсерваторија*. Посећено 25.3. 2022. Београд: Завод за заштиту споменика културе града Београда. http://beogradskonasledje.rs/wp-content/uploads/2012/06/astronomska_opservatorija.pdf.
- Тртовац, Александра. 2010. “Мултимедијални документ као предмет-пројекат на Катедри за библиотекарство и информатику Филолошког факултета у Београду.” *Инфотека* 11 (2): 45–54. http://infoteka.bg.ac.rs/pdf/Srp/2010-2/INFOTHECA_XI_2_December2010_45-54.pdf.

”Дигитална трансформација и књижнице у посебним околностима” – приказ конференције

РАД ПРИМЉЕН:
РАД ПРИХВАЋЕН:

07. март 2022.

16. мај 2022.

Оја Кринуловић
okrinulovic@unilib.rs
Вера Петровић
veramilanovic
petrovic@gmail.com

*Универзитетска библиотека
„Светозар Марковић“
Београд, Србија*

О конференцији

Дани специјалних и високошколских књижница се одржавају већ дуги низ година. Термин одржавања ове конференције је крај априла. У почетку је место одржавања била Опатија, приморски градић у Републици Хрватској, међутим, од 2015. године, место одржавања је померено у Ловран, недалеко од Опатије. Ова конференција представља скуп библиотекара специјалних и високошколских библиотека. Организатори су: Хрватско књижничарско друштво (ХКД), Секција за специјалне и високошколске библиотеке, Национална и свеучилишна библиотека (НСК) из Загреба, као и Књижничарско друштво и Свеучилишна библиотека из Ријеке. Циљ је представљање високошколских библиотека које су у саставу факултета и других високих школа, а и специјалних библиотека које су у саставу разних института и других институција чија је делатност углавном везана за научноистраживачке пројекте са Министарством науке, просвете и спорта.

На претходним конференцијама су обрађиване теме везане за концепцију и организацију специјалних библиотека, као извора и места коришћења знања. Поред тога, пажња је посвећена едукацији корисника у образовању, као и континуираном усавршавању библиотечког особља. Анализирани су и библиотечки системи хрватских библиотека, а такође се говорило и о партнерству између библиотека у Хрватској и библиотека у региону.

На конференције су позвани гости из иностранства, колеге из универзитетских библиотека, представници издавача и стручњаци из других сродних струка. Програм је обухватао предавања по позиву, представљање постера на задату тему, радионице и представљање спонзора.

Теме са неких од претходних конференција

На претходним конференцијама разматране су изузетно занимљиве теме везане за високошколске и специјалне библиотеке, а неке од њих су:

- Библиотеке, оснивачи, заједница: дијалог за промене
- Повезивање и сарадња: прилике и изазови
- Библиотеке : где и како даље?
- Успостављање квалитета у специјалним и високошколским библиотекама –библиотеке у болоњском окружењу
- Искорак према новом библиотечком систему Војаџер
- Нове стратегије у специјалним и високошколским библиотекама, промовисање и прикупљање финансијских средстава
- Партнерства у специјалним и високошколским библиотекама

17. дани специјалних и високошколских књижница

Због глобалне пандемије изазване корона вирусом Ковид 19, термин одржавања конференције 2021. године је померен на крај септембра, па су 2021. године одржани 17. дани специјалних и високошколских књижница, али у нешто измењеном облику.

Седамнаести дани специјалних и високошколских књижница су одржани у периоду од 30. септембра до 2. октобра 2021. године у Ловрану, у облику хибридне конференције, што значи да су учесници могли да је прате и уживо и виртуелно. Тема конференције је била "Дигитална трансформација и књижнице у посебним околностима". Излагања су била у оквиру следећих подтема:

1. Ауторско и сродна права – нове околности и дигитално окружење
2. Компетенције библиотекара за рад у дигиталном окружењу
3. Улога и положај библиотека у дигиталном окружењу
4. Простор, услуге и грађа библиотека у дигиталном окружењу

Излагања су била у облику позваних излагања, округлих столова, радионица и постер галерије.

Конференцију је отворио Округли сто под називом "Улога књижничара у пружању потпоре зnanственицима с обзиром на промјене у начинима зnanствене комуникације". У округлом столу учествовала су два позвана излагача: проф. др Ивана Кунда, продекан за науку на Универзитету у Ријеци, професор Правног факултета са радом на тему „Хибридни правни књижничар или књижничар правник: како вам драго у дигиталном окружју” и проф. др. Мирослав Рајтер, управник Уреда за истраживање Универзитета у Загребу са темом "Како нас рангирања рангирају и како ми рангирамо рангирања". На отварању коференције била су и два позвана излагача: проф. др. Татјана Апарац Јелушић, председница Хрватског библиотекарског већа са радом на тему "Нормативни оквир у подручју специјалних и високошколских књижница" и Лоида Гарција-Фебо, председница Америчког библиотекарског друштва 2018–2019. са радом на тему "Libraries and digital transformation: supersonic library services in the 2020s". Конференцију је закључио позвани излагач проф. др. Мирослав Рајтер са радом на тему "Трендови развоја хрватских свеучилишта из перспективе зnanости."

Сви радови су били изузетно занимљиви и корисни, али треба посебно истаћи рад Ренате Петрушић из Народне и свеучилишне књижнице у Загребу, на тему "Дјела која су недоступна на тржишту: ауторско право, дигитализација, кориштење." У раду су представљене промене које доноси нови Закон о ауторском праву и сродним правима, а које се односе на дела која су недоступна на тржишту (енгл. out-of-commerce works). Нови Закон омогућује институцијама које се баве културним наслеђем да могу да дигитализују и користе дела која су недоступна на тржишту. Занимљиво је да је тај закон прихваћен баш у тренутку одржавања Конференције. С обзиром да у Србији, дела која нису доступна на тржишту се могу само фотокопирати, али не и дигитализовати, можда би требало размотрити могућност увођења и у нашој земљи, оваквог закона који би омогућио дигитализацију дела која нису доступна на тржишту.

На Конференцији је представљено 40 радова и 14 постерских излагања. Конференцију је пратило 140 колега, 50 је било физички присутно, а 90 је слушало предавања онлајн. Имали смо нарочито задовољство да сазнамо да је Министарство културе и медија Републике

Хрватске у току трајања Конференције усвојило Стандарде за дигиталне библиотеке.

Приказ београдске радионице *European Language Grid*

РАД ПРИМЉЕН:
РАД ПРИХВАЋЕН:

19. април 2022.

16. мај 2022.

Бојана Башарагин
bojana.basaragin@ivi.ac.rs
*Истраживачко-развојни
институт за вештачку
интелигенцију Србије*

11. марта 2022. године одржан је први у низу дисеминационих догађаја у оквиру ЕУ пројекта European Language Grid (ELG)¹. У питању је била онлајн ELG радионица у организацији Филолошког факултета у Београду и Друштва за језичке ресурсе и технологије ЈеРТеХ. Овај догађај је уједно најавио и скори, успешан завршетак пројекта средином 2022. године.

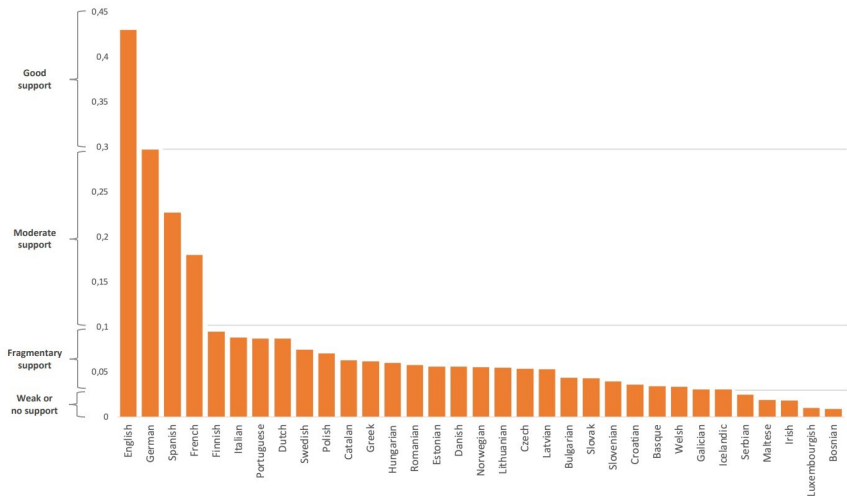
ELG пројекат (2019-2022) настао је с идејом да приближи академску заједницу и индустрију у домену језичких технологија и оснажи позицију Европе у односу на друге континенте у овом погледу. Резултат пројекта је јединствена ELG платформа на којој су сабрани ресурси и алати за обраду европских језика, како би постали видљивији и доступни како академским заједницама, тако и пословним партнерима, невладиним организацијама и државном сектору. ELG као један од својих циљева наводи и помоћ у решавању питања дигиталне језичке угрожености пружањем подршке европским језицима који нису довољно подржани кроз језичке технологије. Према подацима из јануара 2022. године, на платформи постоји преко 10.000 сервиса, алата, скупова података, ресурса и језичких модела за 87 језика, а тај број се увећава из месеца у месец. Након завршетка пројекта планирано је да платформа настави да живи, омогућавајући приступ како некомерцијалним, тако и комерцијалним језичким технологијама.

Радионица је трајала три сата и одржана је онлајн, преко платформе Zoom, а снимак радионице се може наћи на YouTube каналу Друштва ЈеРТеХ². Званични језик радионице био је енглески. Након поздравне речи коју је упутила организаторка, проф. Цветана Крстев, одржана је уводна презентација координатора пројекта проф. Георга Рема. Проф. Рем је направио детаљан преглед пројекта ELG, његових резултата,

1. Пројекат European Language Grid (ELG).

2. YouTube канал друштва ЈеРТеХ.

као и перспектива за даљи раст и развој платформе. Представио је и проблем језичке неједнакости у заступљености језичких технологија за већину европских језика, изузимајући шпански, немачки и француски (Слика 1). Посебно се осврнуо на прилике српског језика, који према извештају пројекта European Language Equality (ELE) за српски језик³ и даље спада у језике са слабом покривеношћу језичким технологијама. На питање шта види као следећи корак у развоју ресурса и алата за српски језик, навео је израду опсежног језичког модела који би се затим могао применити на различите задатке обраде и генерисања српског језика.



Слика 1. Покривеност европских језика језичким технологијама

Затим је уследила практична и детаљна демонстрација употребе платформе, коју је приказала Пени Лабропулу, виши научни сарадник грчког Института за обраду језика и говора. Присуство страних предавача и гостију дало је повода да се за потребе радионице искористи резултат скоро завршеног европског пројекта ELITR⁴. У оквиру овог

3. ELE извештај за српски језик.

4. Пројекат European Live Translator (ELITR).

пројекта развијен је аутоматски систем за превођење и титловање састанака и конференцијских представљања, као и аутоматског хватања белешки. Аутоматски запис свих говора на радионици и њихов превод на српски (или неки други европски) језик у реалном времену могли су се пратити у одвојеном прозору.

Након уводних презентација уследили су примери из домена језичких технологија за српски језик. У духу пројекта, радионица је замишљена тако да обједини примере успешне примене језичких технологија у индустрији и најновија достигнућа у научном домену. Тако су прве две презентације дошле из две велике српске фирме – телекомуникационог оператера Yettel и ИТ групе ComTrade. Ђорђе Хирш, вођа тима за управљање подацима у Yettel Serbia, представио је методу ненадгледане анализе сентимента коју у фирми Yettel примењују на базу гласовних жалби упућених корисничком сервису како би побољшали своје услуге. Душан Јовичић, програмер, приказао је регионално искуство фирме ComTrade у прављењу разговорних агената (четботова).

Примери из индустрије били су праћени представљањем ресурса и алата за српски језик развијених у сарадњи Београдског универзитета (БУ) и Друштва ЈеРТеХ. Најпре је проф. Ранка Станковић говорила о корпусу савременог српског језика СрпКор, који је крајем 2021. године добио нову верзију – СрпКор2021. У односу на претходну верзију, овај корпус је обogaћен новим текстовима из различитих домена, међу којима се истиче корпус литерарних текстова из 19. и с почетка 20. века ELTeC, настао у оквиру COST акције Distant Reading⁵. Овај корпус, састављен од преко 100 књижевних дела обележених врстама речи, уједно је био тема и последњег представљања на радионици. Корпус ELTeC су представиле проф. Цветана Крстев и Милица Иконић Нешић, као потпуно слободан ресурс који се може користити за претраге, анализу и тренирање модела за српски језик. Овај корпус се тренутно може преузети са GitHuba пројекта. СрпКор2021, па тако и ELTeC корпус, могу се претраживати преко система NoSketch, доступног за регистроване кориснике преко веб-странице ЈеРТеХа⁶.

Биљана Рујевић је дала практичну демонстрацију платформе *Leximirka*, направљене за обједињено допуњавање, одржавање и коришћење електронских речника, корпуса, и других лексичких ресурса и алата развијених у сарадњи БУ и ЈеРТеХа. Платформи је, уз неопходну регистрацију, такође могуће приступити са веб-странице ЈеРТеХа.

5. COST акција Distant Reading for European Literary History.

6. Веб-страница друштва ЈеРТеХ.

Михаило Шкорић представио је BEaST, новоразвијени POS тагер за српски језик који представља побољшање у перформансама у односу на тагере у својој основи: *treetagger*, *spaCy* и *stanza*. Овај тагер, који је уједно искоришћен за обележавање корпуса текстова ELTeC, доступан је за преузимање преко GitHuba. Бранислава Шандрих Тодоровић говорила је о веб порталу *NER & Beyond*, развијеном у оквиру COST акције Distant Reading. Портал је направљен да омогући обележавање именованих ентитета различитим доступним алатима, али и мапирање између анотација као производа тих алата. Ова платформа је доступна за употребу преко веб-странице JePTexa.

Радионица ELG била је право освежење. С једне стране, пружио је прилику да се сазна за пројекат ELG, као важан корак ка обједињавању и већој видљивости алата и ресурса за српски језик. С друге стране, учесници су стекли увид и у неке од последњих резултата истраживања у домену језичких технологија за српски језик и њихове примене. Динамична структура у оквиру које су представљана сервиса, ресурса и алата трајала по 15 минута, омогућила је да се за кратко време добије преглед великог броја пројеката, уз довољно информација о сваком од њих. Радујем се следећој прилици да присуствујем овако садржајном и добро организованом догађају, уз уверење да сваки овакав напор доприноси побољшању ситуације језичких технологија за српски језик.

Обука за читање на даљину – *Exploring ELTeC: Use-Cases for Information Extraction and Analysis*

РАД ПРИМЉЕН:
РАД ПРИХВАЋЕН:

28. мај 2022.
8. јун 2022.

Александра Марковић
aleksandra.markovic@isj.sanu.ac.rs
Институт за српски
језик САНУ
Београд, Србија

У Београду је од 22. до 24. марта одржана радионица *Exploring ELTeC: Use-cases for Information Extraction and Analysis* (Истраживање *ELTeC*-а: примери екстракције и анализе информација). Радионица је била посвећена обуци за истраживање *ELTeC*-а (*ELTeC* је скраћеница за *European Literary Text Collection* – Колекција текстова европске књижевности). Ово је била завршна обука у оквиру COST акције (16204) *Distant Reading for European Literary History* (Читање на даљину за европску књижевну историју).

Радионица је била организована на Рударско-геолошком факултету Универзитета у Београду, а организатори су били проф. др Ранка Станковић с Рударско-геолошког факултета (Универзитет у Београду), проф. др Цветана Крстев с Филолошког факултета (Универзитет у Београду),¹ као и мр Јоана Бишук (Joanna Byszuk) са Института за пољски језик Пољске академије наука. Одржана је у хибридном облику, па су учесници могли да бирају да ли ће учествовати уживо или онлајн. Уживо су учествовали полазници из Србије, Словеније, Румуније, Белгије и Литваније, док су онлајн учествовали још и полазници из Аустрије, Британије, Португала и Мађарске. За учешће нису били постављени формални захтеви у смислу образовања, али је, с обзиром на интензитет и обим програма, било препоручено поседовање основних информатичких вештина. Радионица је била

1. Цветана Крстев и Ранка Станковић заслужне су за израду српске колекције текстова. Заинтересовани за српску колекцију романа дигитализованих у оквиру акције *Distant Reading* могу прочитати више о томе у часопису *Инфотека*, бр. 21/2, посвећеном у целости српској колекцији дигитализованих романа, док самој колекцији могу приступити на [github-y](#) акције.

намењена истраживачима из земаља укључених у пројекат *Distant Reading*, заинтересованим за дигитално истраживање књижевности, корпусну и рачунарску лингвистику, теорију књижевности, као и њихову методолошку примену ван граница националних традиција.

Представљени су практични приступи екстраховању информација и анализи текстуалних података, нарочито из корпуса ELTeC развијеног у оквиру COST акције. Међу обрађеним темама били су различити аспекти рада са именованим и географским ентитетима: њихово препознавање (NER, *Named entities recognition*) и екстраховање (NEE, *Named entities extraction*), радом са Вики подацима (Wiki-ELTeC data), повезивање (историјских) података у Nodegoat-у (апликацији посебно израђеној за истраживања у хуманистици),² семантичка анализа помоћу вектора речи и језичких модела, као и поређење корпуса помоћу стилметрије.

Међу предавачима је било познатих имена, учесника у акцији; поменућемо нека од њих: Christof Schöch (професор дигиталне хуманистике на Универзитету у Триру, руководилац COST акције *Distant Reading for European Literary History*; Maciej Eder (директор Института за пољски језик Пољске академије наука); Diana Santos (Универзитет у Ослу), Fotis Yannidis (професор дигиталне хуманистике на Универзитету у Вирцбургу), проф. др Цветана Крстев (проф. информатике на Катедри за библиотекарство и информатику Филолошког факултета УБ, у пензији) и проф. др Ранка Станковић (професор информатике на Катедри за примењену математику и информатику Рударско-геолошког факултета УБ).

Радионица је била организована у девет модула и од учесника се очекивало да похађају сваки од њих.³

1. Кристоф Шех (Christof Schöch) одржао је (онлајн) уводно предавање на тему: *What is ELTeC all about?* (Шта је то ELTeC?). Говорио је о циљевима акције акције *Distant Reading for European Literary History*, с посебним освртом на структуру основног резултата овог пројекта – вишејезичне колекције европских књижевних текстова (ELTeC).
2. Маћеј Едер, Јоана Бишук, Артјомс Шела (Maciej Eder, Joanna Byszuk, Artjoms Šeļa) одржали су предавање (онлајн): *Exploring and comparing ELTeC corpora with stylometry* (Истраживање и поређење корпуса ELTeC помоћу стилметрије).

2. Nodegoat

3. Заинтересовани могу погледати програм радионице.

3. Дијана Сантос (Diana Santos) говорила је на тему (онлајн): *NER exploitation and analysis* (Препознавање именованих ентитета (NER) – експлоатација и анализа).
4. Бенедикт Перак (Benedikt Perak) одржао је предавање (уживо): *ELTeC Data Analysis, Representation of the Geo-Entities and Interlinking with Knowledge bases* (Анализа података из ELTeC-а, представљање географских ентитета и повезивање с базама знања).
5. Фотис Јанидис и Леонард Конле (Fotis Jannidis, Leonnard Konle) говорили су на тему (онлајн): *Semantic analysis using word embeddings and language models* (Семантичка анализа помоћу вектора речи и језичких модела).
6. Ранка Станковић и Милица Иконић Нешић одржале су (уживо) сесију посвећену Вики подацима: Увод у Вики податке; схема Вики-ELTeC (сви метаподаци из заглавља заједно са основним ликовима који се помињу у романима, њиховим односима, локацијама; процедура и кораци за попуњавање Вики-ELTeC података; предефинисана испитивања помоћу SPARQL упита. Вежбе: попуњавање Вики података).
7. Денис Морел, Ерик Лапорт и Цветанав Крстев (Denis Maurel, Eric Laporte) припремили су (онлајн) излагање на тему: *Unitex for processing of literary text: the case of NER automata. Enriching ELTEC texts by Named Entity Recognition using CasSys to parse texts with Unitex graph cascade of finite state transducers in different languages* (Unitex за обраду књижевних текстова: случај аутомата за препознавања именованих ентитета. Обогаћивање текстова из ELTEC-а именованим ентитетима препознатим Unitex-овим алатом CasSys за парсирање текстова на различитим језицима помоћу каскада коначних трансдуктора).
8. Џеси Лабов, Пим ван Бре, Герт Кеселс (Jessie Labov, Pim van Bree, Geert Kessels) имали су (онлајн) излагање на тему: *ELTec in Nodegoat – Introduction to the Nodegoat interface and how it works with this kind of data*, (ELTec на платформи Nodegoat – Увод у интерфејс Nodegoat-а и начин обраде текстуалних података).
9. Пим ван Бре, Герт Кеселс имали су (онлајн) излагање на тему: *Using Nodegoat for working with the ELTeC data* (Употреба Nodegoat-а за рад са подацима из ELTeC-а), у коме су се фокусирали посебно на NER и представљање начина за обогаћивање тих података повезивањем са отвореним изворима података).

Радионица је била интензивна, веома информативна и динамична. Материјали потребни за рад и вежбе били су постављени на платформи GitHub.⁴ Организација и вођење кроз одабране теме били су одлични, теме занимљиве, а вежбе добро осмишљене (чак и за оне који нису довољно информатички обучени, као ауторка овог приказа, која је и сама учествовала на радионици).

4. Материјали са радионице

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 обајављују 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 прештампавати 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 ТИРИЛИЧНИМ 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 угњежености.
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>).